International Journal of Industrial Engineering, 18(3), 109-119, 2011.

A SUPPLIER DISCOVERY FRAMEWORK FOR EFFECTIVE AND EFFICIENT CONFIGURATION OF A SUPPLY CHAIN

Choonghyun Lee[a], Young Hoon Lee[a], Yun Peng[b], Hyunbo Cho[c],

[a]Department of Information and Industrial Engineering, Yonsei University
262 Seongsanno, Seodamun, Seoul 120-749, South Korea
[b]Computer Science and Electrical Engineering, University of Maryland Baltimore County
Baltimore, MD 21250, USA
[c]Department of Industrial and Management Engineering, Pohang University of
Science and Technology San 31 Hyoja Pohang 790-784, South Korea

A supplier registry can play a central role in configuring a global supply chain for service-oriented enterprise integration by providing an open platform for publishing and discovering suppliers distributed over Internet. The availability of correct classification schemes used to organize suppliers based on their capability descriptions is the key to building an effective registry. This paper discusses the clustering-based construction of classification schemes from existing capability descriptions of suppliers and then the registration and retrieval of suppliers based on these classification schemes. This approach is based on the use of similarity measures to discriminate any two suppliers' capability descriptions. The results look promising in that small medium suppliers can expose themselves to the large companies in an efficient manner, while large companies can diversify their supply sources.

Keywords: Supplier discovery, supplier registry, classification and clustering, classification scheme, similarity measure

(Received 7 Aug 2010; Accepted in revised form 25 Jan 2011)

1. INTRODUCTION

Although supply chain operations have been intensively studied in supply chain management, how to discover suppliers automatically necessary to build a supply chain has yet to be developed. Supplier discovery is a collection of activities to find suppliers who provide large companies with services and/or materials for production of final goods. In these days, many large companies form a supply chain to enhance their competitiveness, whose prerequisite is to discover suppliers no matter where they are located in the world. Moreover, large companies continue to diversify supply sources. Once a set of suppliers are discovered (i.e., business partners are found) and then a supply chain is formed, B2B (Business to Business) operations (for example, contracting, price negotiation, payment, etc.) and logistics will be conducted. In the past, a buyer depended largely on several conventional tools, such as Yellow Pages in the phone directory, Internet surfing, and word of mouth. Many of these suppliers are located not only in different countries, but also from at different domains from the buyer's industry sector. These suppliers are all different from each other in terms of capabilities, such as the variety of products, production capacity, costs, quality, delivery reliability, reputation, payment methods, etc. The buyer should be able to evaluate all of these capabilities in selecting the best suppliers. This, of course, forces suppliers to advertise their capabilities in such an effective way that they can be easily exposed to the buyer. We will propose a supplier registry framework through which suppliers publish their capabilities and buyers discover suppliers.

Service-oriented architectures (SOAs) have become the underlying paradigm for application deployment and integration within and across enterprises. SOA will provide a set of well-defined services that can be invoked and used within business domains (Gottschalk et al., 2002). We adopt the concept of the service registry to develop a supplier registry, through which suppliers advertise their manufacturing capabilities and a buyer subscribes to the registry to discover the suppliers appropriate for building a supply chain. A leading service registry specification is UDDI (Universal Description Discovery and Integration) (http://www.uddi.org), which supports service registration and discovery by its categorization (i.e., Yellow Pages) based on standard taxonomies such as NAICS (the North American Industry Classification System) (http://www.census.gov/eos/www/naics/) and UNSPSC (the United Nations Standard Products and Services Code) (http://www.unspsc.org/).

However, this syntactic registry is often inadequate in that 1) discovery is ineffective and inaccurate due to the use of naive classifications according to the industrial domains rather than according to the capabilities the suppliers provide, 2) a pre-defined taxonomy is not suitable for reflecting manufacturing capabilities in emerging or unknown domains, and 3) multiple classification schemes from the viewpoint of different aspects of supplier capabilities are not supported. Therefore, a well-organized classification registry structure is a prerequisite for the registration and discovery of suppliers. We believe that semantics can be used to more accurately express the needs of the buyer and the manufacturing capabilities of the supplier. Semantic matchmaking is the process of finding suppliers by matching the meaning of the needs of the buyer to

ISSN 1943-670X

the meaning of the capabilities of the suppliers.

The objective of this paper is to design a supplier registry that comprehensively supports multiple classification schemes, including both standard classification schemes (i.e., NAICS and UNSPSC) and capability classification schemes. This paper discusses the automatic construction of capability classification schemes, the guided registration of suppliers, and the discovery and retrieval of suppliers. All of these activities are related to artificial intelligence processes, such as similarity computation, clustering, classification, matchmaking, and reasoning. In the remaining sections, we provide the proposed framework of a supplier registry, the construction and manipulation of the classification schemes, and the discussion and analysis of a supplier registry.

2. RELATED WORK

There are two leading registry standards for e-business (or, more specifically for service oriented architecture): UDDI and ebRS (ebXML Registry Service) (Breininger, 2003). UDDI is a platform-independent XML (eXtensible Markup Language)-based registry used by businesses worldwide. It is a core Web Service standard that is closely coupled with SOAP (Simple Object Access Protocol) messaging and WSDL (Web Services Description Language) publication. UDDI consists of three components for service registration: White Pages for identifier and contact information, Yellow Pages for industrial categorization, and Green Pages for technical information about services. Although the UDDI-based Universal Business Registry is no longer in service (http://soa.sys-con.com/node/164624), UDDI is still a central pillar of web service infrastructure. In contrast, ebRS is built on top of the ebXML (e-business XML) infrastructure and provides a set of services that enable both information sharing and business process integration between trading partners. Although ebRS provides more comprehensive and robust functionalities than UDDI (from an e-business integration perspective), it has not flourished as an open registry for web services.

Recent academic research on registries has focused on semantics enrichment for the effective publication and correct discovery of web services. For example, the UDDI specification was redefined such that DAML-S or OWL-S data structures were used to capture the service capabilities and mapped onto the corresponding UDDI structures (Paolucci et al., 2002; Srinivasan et al., 2004; Luo et al., 2006). Pokraev *et al.* (2003) employed DAML-S, but defined the key values using tModel elements. Jang *et al.* (2005) embedded an OWL-based semantic container in the UDDI registry. Instead of using UDDI, Dogac *et al.* (2004) extended the ebXML RIM (Registry Information Model) using OWL-like data structures and assumed OWL semantics for these structures. An attempt was made to extend DAML-S syntax to include semantic service capabilities (Kulvatunyou et al., 2005). All of the approaches listed above concentrated on the description and discovery of services (and their information contents) in a semantics-rich manner.

An important benefit of the registry arises from the use of classification schemes that organize data in a structured manner. Well-known classification schemes include NAICS, UNSPSC, and MDR (Meta Data Registry).

NAICS is used to classify and measure economic activity in North America. Its numbering system is a six-digit code: business sector (2 digits) - subsector (1 digit) - industry group (1 digit) - particular industry (1 digit) - and national industry (1 digit). For instance, NAICS Code 336111 in the 2007 edition indicates “Automobile Manufacturing” where the first two digits (33) designate the “Manufacturing” sector, the second digit (6) the “Transportation Equipment Manufacturing” sub-sector, the third digit (1) the “Motor Vehicle Manufacturing”, the fourth digit (1) the “Automobile and Light Duty Motor Vehicle Manufacturing”, and the last digit (1) the “Automobile Manufacturing”.

UNSPSC is a coding system for classifying both products and services and was designed for vitalizing global e-commerce. The coding system consists of eight digits: segment (2) - family (2) - class (2) - and commodity (2). It is not possible to match NAICS and UNSPSC codes because NAICS is an industry-driven code scheme whereas UNSPSC is a service- or product-driven scheme. For example, there is no UNSPSC code that corresponds exactly to NAICS Code 336111, but the UNSPSC Code 25101503 denotes the “Automobiles or Cars commodity” in the “Passenger Motor Vehicles” class (25101500), the “Motor Vehicles” family (25100000), and the “Commercial and Military and Private Vehicles and their Accessories and Components” segment (25000000).

The ISO (International Organization for Standardization) provides a Metadata Registry (MDR) standard (i.e., ISO/IEC 11179) (http://metadata-standards.org/11179/index.html) that is used to represent the metadata of an organization. The ISO metadata registry consists of a hierarchy of concepts, where each concept is associated with properties. These standard classification schemes cannot exactly describe all of the diverse industries, products, and services, and therefore the discovery of suppliers is inaccurate and inefficient.

3. PROPOSED ARCHITECTURE OF THE SUPPLIER REGISTRY

3.1 Overview of the Supplier Registry

Suppliers advertise their capabilities using product catalogue either via the internet or by hard copies. They expect buyers to manually investigate their catalogues and to place orders. However the catalogue may not include the detailed manufacturing capabilities necessary to produce products and/or services. The objective of the paper is to propose a conceptual framework of supplier discovery for which a registry enables suppliers to publish their manufacturing capabilities and buyers to discover the suppliers that can provide materials and/or services. To this end, the manufacturing capabilities of a supplier should be defined, whose examples are shown in Table 1. The manufacturing capabilities may include the identifier of the supplier (i.e., the supplier name), the classification codes of the industry domain (e.g., the UNSPCS and NAICS classification codes), the list of intellectual properties, production facilities, the list of products, delivery services, pricing mode, etc. These capabilities are classified into several categories, such as general information, financial information, reputation, etc. These capabilities will be represented in an XML format where capability descriptions are stored in a structured manner.

Table 1. Exemplary items of supplier capabilities

General information	Financial information	Reputation level	Core Resource	Core competency	Extra service	Contract
Supplier name Classification of industry (NAICS, UNSPSC) Number of employees Presence of overseas branches	Financial statement (annual sales, revenue, asset, debt ratio, etc.)	Company history List of awards CEO background Certificates Export ratio Main clients	Presence of research center Employee empowerment plans Intellectual properties Production facilities	List of products Product specifications R&D investment amount	Delivery services Security B2B transaction protocols	Service level agreement Pricing model

The supplier registry plays a central role in supply chain configuration because it can manage the capability descriptions of suppliers, and link buyers and suppliers. We envisaged an intelligent registry framework with the following main functions: the construction of a registry using existing list of suppliers; the registration of a new supplier based on its similarity measure to the existing suppliers; and the retrieval of registered suppliers through matchmaking. The conceptual diagram for this framework is depicted in

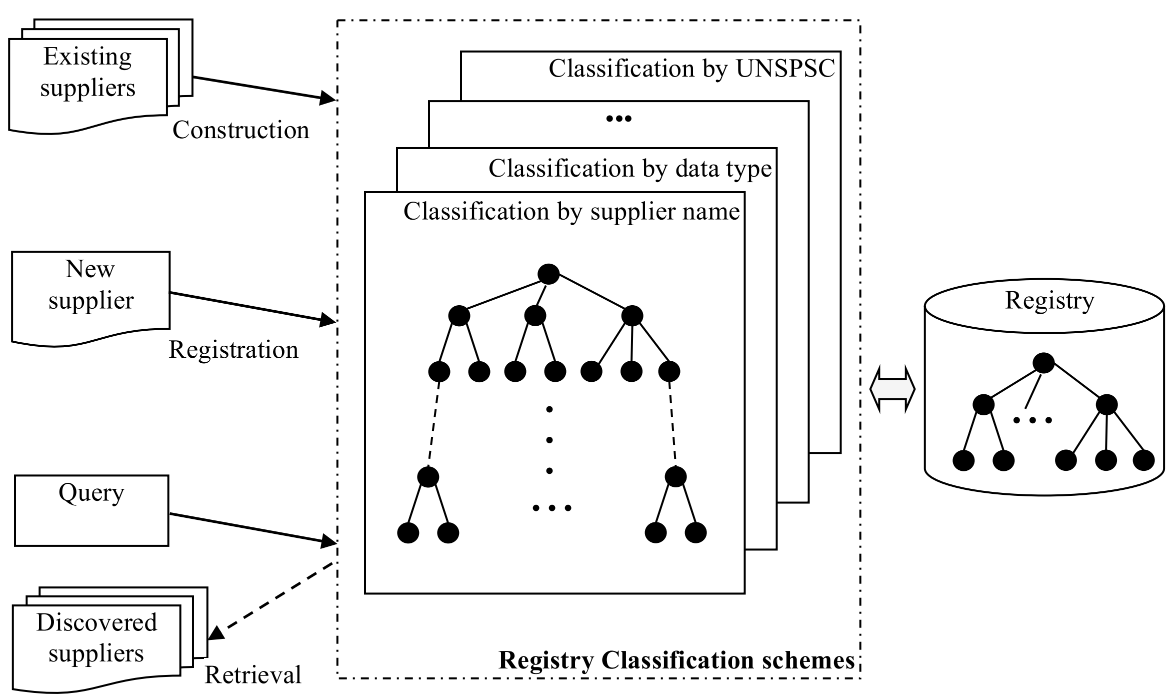

Figure 1. Overview of supplier registration and discovery.

.

The construction module rearranges the existing suppliers according to the classification schemes in such a way that the registration and retrieval of a particular supplier can be performed efficiently and effectively. The constructed schemes should also support various comparison predicates used for matchmaking processes in the registration and retrieval modules.

The registration module finds the location appropriate for registering a new supplier. How to find the location in the registry classification scheme is a classification problem that consists of successive computations of the similarity measures between the new supplier's capability to be registered and the existing suppliers' capabilities. This process makes the supplier retrieval more efficient.

The retrieval module discovers the best matched suppliers for a given query, returns the pointers to the discovered suppliers' capability descriptions, and finally retrieves all appropriate suppliers. An additional process may be required to filter those retrieved suppliers that are most relevant to the query based on quality criteria (i.e., Quality of capability) and functional criteria (Jeong et al, 2007a). This matchmaking process is similar to the classification process used by the registration function.

3.2 Basic Structure of the Supplier Registry

A key prerequisite for building an effective registry is the presence of the acceptable classification schemes that group similar suppliers together and guide the automated registration and retrieval of wanted suppliers. A classification scheme is defined as the descriptive information required for the arrangement or division of objects into clusters based on characteristics that the objects have in common (http://en.wikipedia.org/wiki/Classification_scheme). In the case under consideration, the general term *object* is replaced by the specific term *supplier*. The benefits of using classification schemes are to quickly find a set of relevant suppliers based on their capabilities, to facilitate the detection of duplicate suppliers, and to convey the semantics of capabilities that are not interpreted clearly via capability names and attributes.

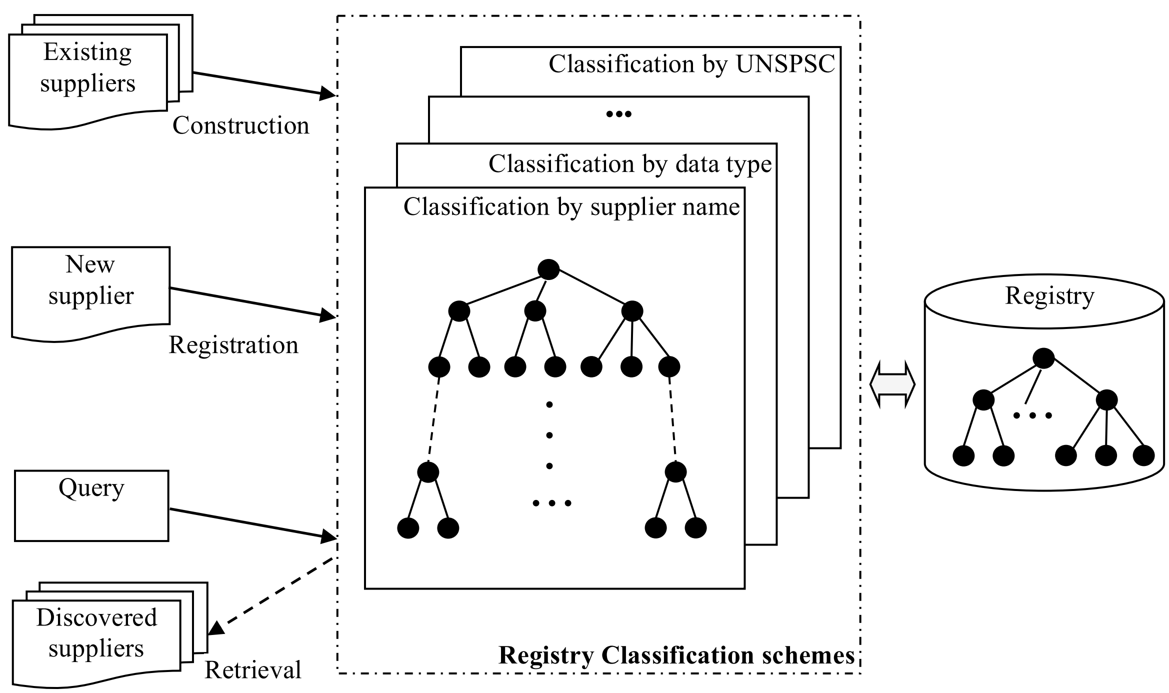

Figure 1. Overview of supplier registration and discovery.

The registry should support multiple classification schemes, including capability descriptions of suppliers, as well as existing schemes, such as NAICS and UNSPSC. A new supplier can then be placed in several classification schemes. The use of multiple schemes not only decreases the chance of missing the relevant suppliers during the search and discovery process, but also enriches the semantics of the new supplier's capabilities in the registration process. Further, it provides practicality and flexibility for supplier discovery in that one can, for example, search suppliers either by using the supplier name only or by using several criteria simultaneously.

Capability-based supplier taxonomy plays an important part in the classification scheme to fulfill complementary needs of the existing schemes that only classify an industry according to the physical products and/or services it provides. It is a meta-model that organizes the suppliers, based on the attributes of capabilities, such as the capability name, its related resources and processes, supplier's reputation, and other tag information such as semantic annotations or keywords (Jeong et al. 2007a).

3.3 Tree Representation of Classification Scheme

A typical classification scheme is structured in a hierarchical form, i.e., a tree-based taxonomy, since a tree structure is an intuitive and practical representation of the classification scheme. Other structure forms may include keywords, thesaurus entries, data models, and ontology taxonomies.

In many classification problems, the CART (Classification and Regression Tree) analysis (Duda et al, 2001) is used for building a decision tree for a number of existing suppliers in a nonparametric manner. In our approach, a top-down hierarchical clustering approach, that is, a divisive method, is employed that successively splits a set of suppliers into finer clusters. The CART decision tree uses a form of binary recursive tree in which an intermediate classification node consists of two subnodes according to the values of a certain attributes of supplier's capability descriptions. Each (leaf) node contains the clusters of pointers to the similar suppliers in terms of their capabilities that may be physically stored in a central repository or distributed over the Internet. By doing this, the discovery and retrieval of relevant suppliers can be easily performed.

Although a parent node may have an arbitrary number of child nodes according to the number of clusters, the most efficient and effective number of child nodes is difficult to determine due to the nature of unsupervised clustering. Balancing the number of suppliers across the cluster is a key issue. Various approaches to the selection of the optimal number of clusters have been discussed in (Hu and Xu, 2004; Lu snd Traore, 2005; Beringer and Hullermeier, 2007). A perfect binary decision tree, in which each leaf node points to a nearly equal number of suppliers, provides a seemingly optimal classification scheme. It is note that the larger number of clusters the classification scheme has, the shorter the classification scheme is. Furthermore, the number of exemplars may affect the efficiency of supplier registration and retrieval.

The criteria for splitting the set of existing suppliers are not easy to determine, due to the high degree of complexity present in the capability descriptions. Since supplier capability is descriptive, the similarity measure between supplier capabilities can be used as a discriminator of the two suppliers. Thus, the suppliers located under the same classification node tend to be similar, but are distant from those under other nodes. This issue is, however, not as simple as it appears because the size and diversity of clusters vary. How to evenly distribute suppliers in the decision tree, with the goal of obtaining an optimal classification scheme for registering and retrieving suppliers, is yet to be solved.

An example of a full binary classification scheme is depicted in **Error! Reference source not found.**. Each node, with the exception of the root node, includes an *exemplar*, which is the medoid supplier of the corresponding cluster. It implies that the average dissimilarity of the medoid supplier to all other suppliers in the cluster is minimal. In addition to the exemplar, every node may also specify a boundary supplier necessary to set the boundary of the cluster.

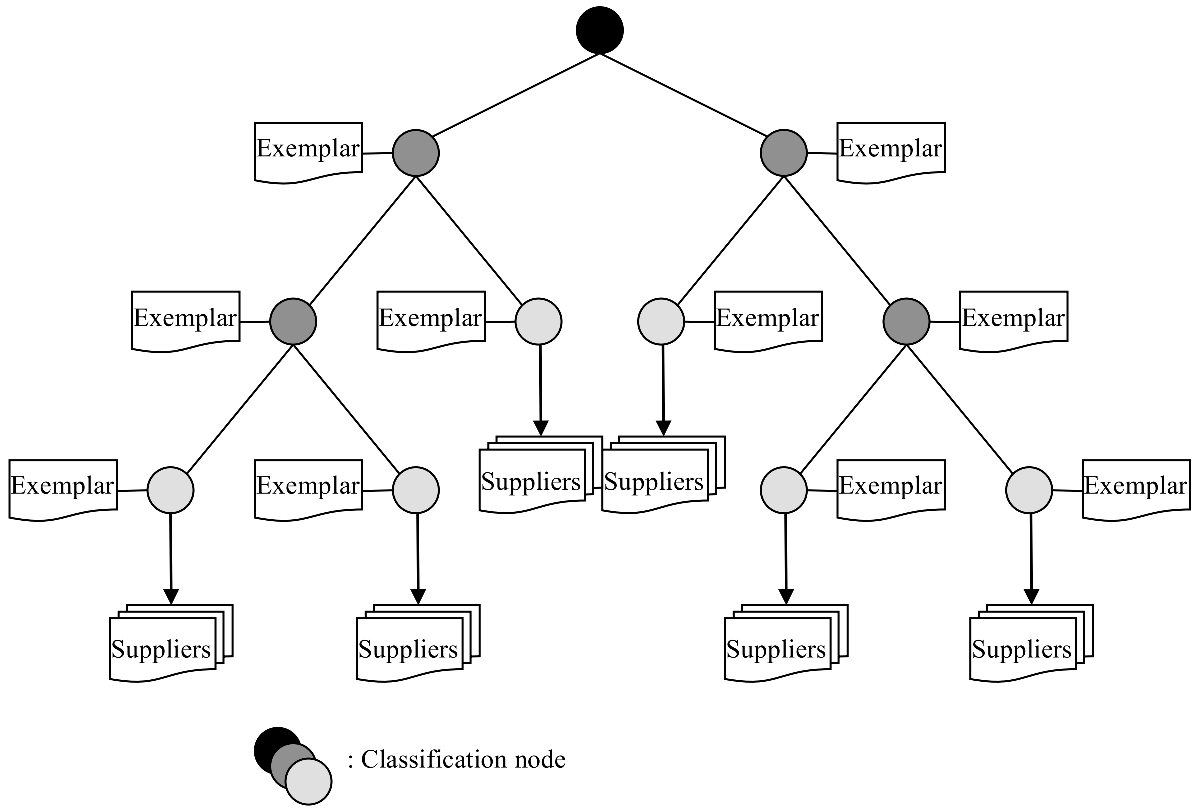

Figure 2. Illustration of a binary classification scheme.

4. CONSTRUCTION OF A CLASSIFICATION SCHEME FOR SUPPLIERS

We address how to construct the classification scheme based on the capability descriptions of suppliers. The construction modules perform the extraction of information from capability descriptions, the computation of similarity measurements, and then the clustering of the suppliers.

4.1 Information Extraction from Capability Descriptions

The most important issue in the construction of the classification scheme is how to compute the similarity measure as a discriminator between suppliers. The pieces of information used to compute the similarity measure are the attribute values and descriptions that describe supplier capabilities. The descriptions are interpreted and then separated into strings (i.e., the supplier name) and XML schema documents (i.e., the production facilities). The extracted information is fed into a parsing processor according to the type and format. The string-type information, which is often a compound word, must pass through a normalization process that separates a compound word into atomic words, which results in morphologically basic forms and removal of unimportant stop words (Jeong, 2006). An abstraction process is needed to transform the XML schema documents into the intrinsic tree structure.

4.2 Similarity Measure between Suppliers

The semantic similarity measure is used to quantify the proximity between suppliers, to classify and register a supplier, and to find the best match between suppliers and the query. The issue of computing the semantic similarity measures used for building a registry is left open in this paper, although we do provide some general guidelines.

Similarity measures between suppliers can be computed based on the capability descriptions, such as strings and XML schema documents. These are non-numeric and non-vectorial data. Vector space models (VSMs) (Salton et al., 1975) can be used for approximating the vectorial representation of string data. Although VSMs use conventional measures for quantitative features, such as the cosine similarity (Xu and Wunsch, 2005), their accuracy depends heavily on the selection of index words. Moreover, our main concern is to process the semantics of single words and structured documents rather than the semantics of plain texts. The semantic similarity of words is often measured via the exploration of synonyms and information contents (Jeong, 2006).

XML documents require more complicated processing. An XML document is an organized collection of interdependent individual terms (which are often compound words). A desirable measure should assess not only the semantics of individual compound terms, but also the relationships between these terms. The similarity of compound words can be measured by solving the stable marriage problem or the assignment problem between normalized terms. A helpful aspect of this process of probing the relationships between terms in XML documents is that terms are structured in a tree, i.e., in a Document Object Model (DOM). Several similarity measures that make use of tree structures include node/path matching (Buttler, 2004; Costa et al., 2004; Amer-Yahia, 2005; Bertino et al., 2004), tree edit distance (TED) (Nierman and Jagadish, 2002; Lian et al., 2004; Dalamagas et al., 2006), extended vector space model (Yang et al., 2005), Fourier transformation (Flesca et al., 2005), and kernel methods (Jeong et al. 2007b).

In conclusion, atomic terms are manipulated using semantic word similarity measures, classification codes are processed according to the corresponding classification scheme, and XML documents are processed using tree-structural similarity measures. Furthermore, an integrated similarity measure that combines multiple attributes can also be created.

4.3 Clustering of Suppliers

The selection of the clustering algorithm relies on the type of input data under consideration. In our problem, the input data is a proximity matrix containing pair wise similarities among suppliers. PAM (Partitioning Around Medoids) and graph theory-based clustering algorithms have this capability of interpreting such a matrix. Another consideration in the selection of clustering algorithm is its ability to handle null values in the proximity matrix. The computation of the similarity measures between suppliers constitutes a large computational burden when many suppliers are involved, resulting in a sparse proximity matrix. Graph theory-based algorithms (e.g., affinity propagation (Frey and Dueck, 2007), Chameleon (Karypis et al, 1999), AMOEBA (Estivill-Castro and Yang, 2000) have this capability.

The construction procedure shown below—recursive top-down partitioning—uses divisive hierarchical clustering (Kaufman and Rousseeuw, 1999), where the divisive task at each cluster is the partitional clustering. The recursive procedure terminates before each cluster has only one supplier.

Step 1: One large cluster consists of all the suppliers. Obtain a proximity matrix describing the pair wise similarity measures between suppliers. Note that the proximity matrix may be sparse (i.e., with null values).

Step 2: Split each cluster assigned to a leaf classification node into an arbitrary number of clusters (e.g., two for a binary decision tree), unless it has met stopping conditions. The stopping criteria may include the number of clusters in a cluster.

Step 3: For each newly constructed cluster, assign its exemplar supplier (and boundary supplier) to the corresponding classification node, and return to Step 2.

In addition to the construction procedure, we envision a self-evolving structure, such that the classification scheme is automatically updated when new suppliers are registered. The update rule is the converse of the above stopping criterion: split a leaf node when the number of suppliers belonging to the node exceeds a bearable limit. This rule prevents a particular leaf node from having dissimilar suppliers. Merging any two classification nodes may also occur. Suppose a classification node is being updated. If its sibling node contains a much smaller number of suppliers, it is desirable to merge the two clusters and then split the resulting cluster to promote efficiency.

5. SUPPLIER REGISTRATION AND RETRIEVAL

Registering and discovering suppliers require the identification of the classification node that is most appropriate to the suppliers to be registered and/or a query about the suppliers to be discovered, respectively.

5.1 Supplier Registration

Supplier registration is the activity that registers suppliers to the correct classification nodes so that they may be effectively discovered. This is a classification problem in which the supplier must be assigned to the node that contains the optimally matched exemplar. There are two methods for classifying suppliers. First, the supplier is compared with all exemplars in the leaf nodes, and the best match is selected. The second method uses a stepwise search (or a routed traversal) from the root node to the leaf node to which the supplier potentially belongs. In both methods, the leaf node containing an exemplar closest to the new supplier is the classification node at which the new supplier is registered. The similarity measures between the supplier and the exemplars are the criteria for classification.

This method may not guarantee to balance clusters at certain parent nodes during registering, we cannot exclude the possibility of unfair splits. As shown in **Error! Reference source not found.**, for example, the unseen supplier X to be registered is closer in Euclidean distance to E_2 than to E_1, despite its belonging to cluster C_1. In such cases, we must investigate the supplier's cluster membership. The use of boundary descriptions is intended to reduce the possibility for such misclassifications, and one (or two or more) step-ahead computations may be necessary. In other words, the decision can be delayed until subordinate clusters are considered. For example, the supplier X will become a member of cluster C_{11}.

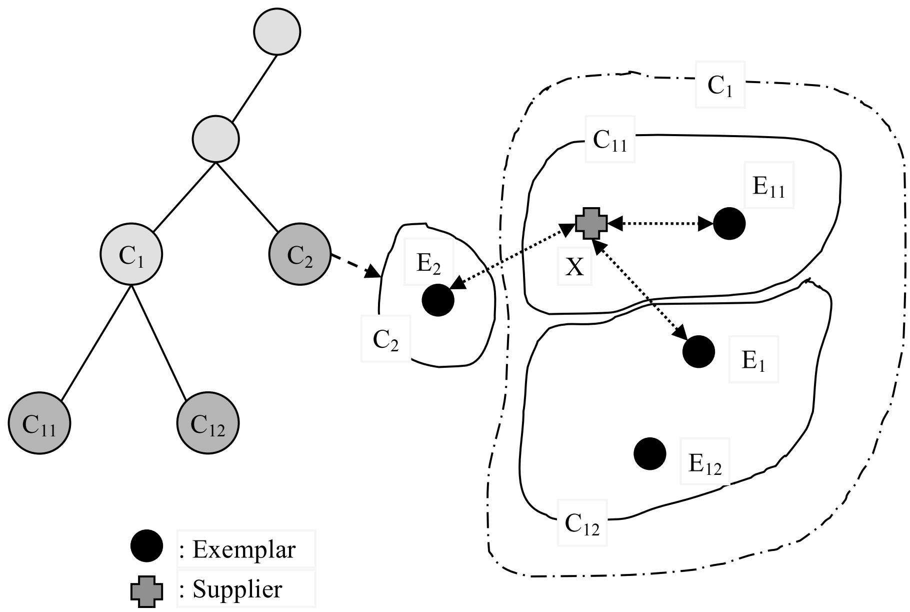

Figure 3. Example of unbalancing between clusters.

5.2 Supplier Discovery and Retrieval

Supplier discovery finds the classification node relevant to a query and then identifies a collection of suppliers, whereas supplier retrieval performs the return of the corresponding suppliers. These processes are treated to be logically interchangeable in this paper. Supplier discovery and retrieval are associated with matchmaking problems in which one finds the classification node (or the set of suppliers in the node) that is most appropriate to the query. The query may contain partial information (e.g., dummy supplier name or incomplete capability descriptions) about the wanted supplier which contains full descriptions of the capabilities.

Given a query, the classification scheme is searched by visiting all the nodes. Searching must be more comprehensive due to the incompleteness of the query. Thus, if an intermediate node best matches the query, all the suppliers under that node (i.e., descriptions belonging to all subordinate leaf nodes) are returned. One difference between retrieval and registration is that a retrieval node is not necessarily a leaf node.

Another issue arises from the use of multiple classification schemes. The various classification schemes constructed according to different capability categories enable to group the sets of different suppliers and place them in different nodes. An ensemble approach that combines multiple classifiers should be constructed either in parallel combination or in cascading combination. A parallel combination chooses the union or intersection of the sets according to the different classification schemes, whereas a cascading combination narrows down the candidate suppliers sequentially.

6. INDUSTRY APPLICATION

The conceptual framework to register and discover suppliers is shown in **Figure 4**. Once the registry is developed and a set of suppliers are registered with their capabilities, the principle of supplier discovery is as follows: (1) A buyer submits to the matchmaker its requirements about the suppliers to be discovered, (2) The matchmaker requests the semantically related terms of buyer's requirements to the domain ontology, (3) In response, expanded keywords are sent to the registry which discovers all the suppliers that match the expanded keywords, (4) Finally, the matchmaker sorts the list of discovered suppliers based on the similarity measurement between buyer's requirements and supplier's capabilities, and then return the ranked list of suppliers to the buyer. The results look promising in that SMEs expose themselves to the large companies in an effective manner, while large companies can diversify their supply sources.

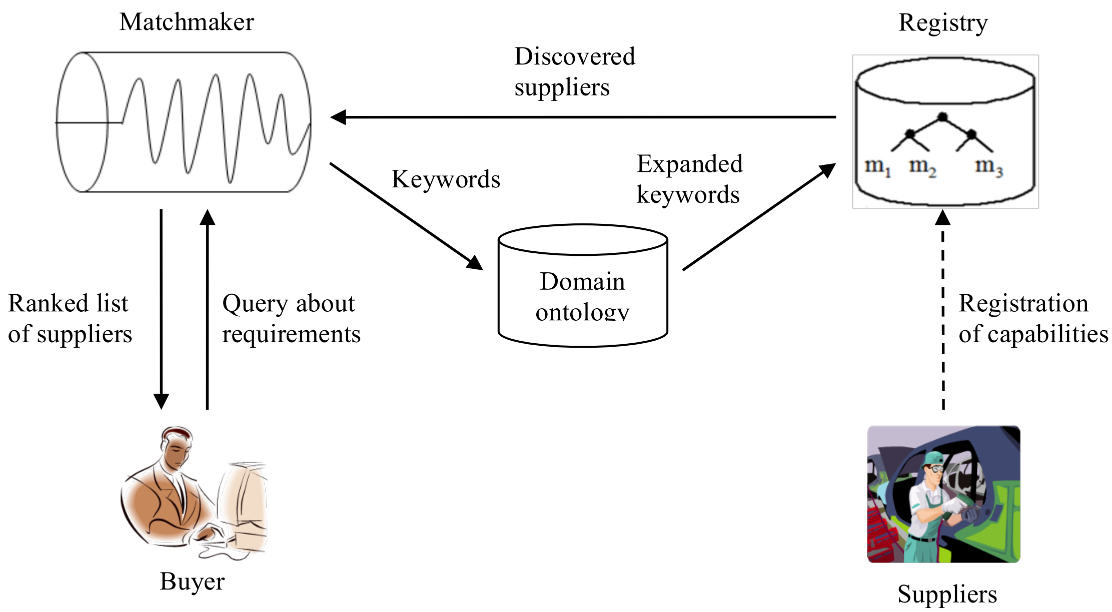

Figure 4. Overview of a supplier discovery framework

7. CONCLUSIONS

The supplier classification schemes can be automatically built using top-down divisive clustering of the supplier list in the registry. The constructed schemes are not only effective in the registration of suppliers to the classification node in which similar suppliers are co-located, but also efficient in the discovery of suppliers relevant to a given query. The conceptual framework can be directly applied to the development of supplier registry through which small medium enterprises can publish their capabilities and large companies can discover suppliers with regards to their capabilities.

The use of multiple classification schemes allows the registry to place a supplier in several classification nodes simultaneously. The registry registers a supplier in conjunction with several similarity measures and cascaded exemplars (from the root node to the leaf node). By this procedure, the registry returns a number of suppliers whose capability descriptions are similar to those of the query. The use of multiple schemes not only decreases the chance of missing relevant suppliers during the search and discovery process, but also enriches the semantics of the capability descriptions during the registration process. Furthermore, it provides practicality and flexibility to search for suppliers. For example, one can search for suppliers by using the supplier name only or by using several criteria in combination.

The use of semantic similarity measures resolves uncertainties regarding the nature of a capability description. This is accomplished by relating a supplier's capability description to semantically and functionally relevant suppliers' capability descriptions and by separating it from irrelevant suppliers. As discussed above, the suppliers that belong to the same classification node are likely to be similar. Therefore, all the suppliers that are relevant to a particular query can be obtained simultaneously.

The supplier registry plays a central role in building a global supply chain for service-oriented enterprise integration and deployment, in that it connects suppliers and their products and/or services to buyers. Although previous studies have extended the standard registry to accommodate the semantic descriptions of web services, this paper proposes how to register and discover suppliers. For this reason, this study has attempted to construct a supplier discovery platform through which both suppliers and buyers have benefits in building a global supply chain.

8. REFERENCES

Amer-Yahia, S., Koudas, N., Marian, A., August 30-September 2, 2005. Structure and content scoring for XML. *Proceedings of the 31st VLDB Conference*. 361–372. Trondheim, Norway.
Beringer, J., Hullermeier, E., July 23-26, 2007. Adaptive optimization of the number of clusters in fuzzy clustering. *Proceedings of IEEE International Conference on Fuzzy Systems*. 1–6. ondon, UK.
Bertino, E., Guerrini, G., Mesiti, M., 2004. A matching algorithm measuring the structural similarity between an XML

document and a DTD and its applications. *Information Systems* 29. 23–46.
Breininger, K., 2003. OASIS/ebXML Registry Services Specification v2.5. Available online via http://www.ebxml.org/specs/ebrs2.pdf, [Accessed in January 2010].
Buttler, D., June 21-24, 2004. A short survey of document structure similarity algorithms. *Proceedings of the 5th International Conference on Internet Computing* (IC 2004). 3–9. Las Vegas, NV.
Costa, G., Manco, G., Ortale, R., tagarelli, A., 2004. A tree-based approach to clustering XML documents by structure. *Knowledge Discovery in Databases: PKDD* 2004 LNAI 3202. 137–148.
Dalamagas, T., Cheng, T., Winkel, K., Sellis, T., 2006. A methodology for clustering XML documents by structures. *Information Systems* 31. 187–228.
Dogac, A., Kabak, Y., Laleci, G., March 28-29, 2004. Enriching ebXML registries with OWL ontologies for efficient service discovery. *Proceedings of the 14th International Workshop on Research Issues on Data Engineering: Web Services for e-Commnerce and e- Government Applications* (RIDE'04). 69–76. Boston, MA.
Duda, R., Hart, P., Stork, D., 2001. *Pattern Classification*, 2nd Edition, John Wiley & Sons, Inc, New York, NY.
Estivill-Castro, V., Yang, J., August 10-12, 2000. AMOEBA: Hierarchical clustering based on spatial proximity using delaunay diagram. *Proceedings of the 9th International Symposium on Spatial Data Handling*. 26–41. Beijing, China.
Flesca, S., Manco, G., Masciari, E., Pontieri, L., Pugliese, A., 2005. Fast detection of XML structural similarity. *IEEE Transactions on Knowledge and Data Engineering* 17. 160–175.
Frey, B., Dueck, D., 2007. Clustering by passing messages between data points. *Science* 315. 972–976.
Gottschalk, K., Graham, S., Kreger, H., Snell, J., 2002. Introduction to web services architecture. *IBM Systems Journal* 41, 170–177.
Hu, X., Xu, L., 2004. Investigation on several model selection criteria for determining the number of cluster. *Neural Information Processing – Letters and Reviews* 4. 1–10.
Jang, J., Jeong, B., Cho, H., Lee, J., August 18-21, 2005. Capability and extension of UDDI framework for semantic enterprise integration. *IFIP 5.7 Advances in Production Management Systems*. Rockville, MD.
Jeong, B., 2006. *Machine Learning-based Semantic Similarity Measures to Assist Discovery and Reuse of Data Exchange XML Schemas*, Ph.D. Dissertation. Department of Industrial and Management Engineering, Pohang University of Science and Technology.
Jeong, B., Cho, H., Kulvatunyou, B., Jones, A., August 13-15, 2007a. A multi-criteria web service composition problem. *Proceedings of the 2007 IEEE International Conference on Information Reuse and Integration* (IRI 2007). 379–384. Las Vegas, NV.
Jeong, B., Lee, D., Cho, H., Kulvatunyou, B., June 26-29, 2007b. A kernel method for measuring structural similarity between XML documents. *Proceedings of the 20th International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems* (IEA/AIE-2007). 572–281. Kyoto, Japan.
Karypis, G., Han, E., Kumar, V., 1999. Chameleon: Hierarchical clustering using dynamic modeling. *Computer* 32. 68–75.
Kaufman, L., Rousseeuw, P., 1999. *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, Inc., New York, NY.
Kokash, N., August 28-29, 2006. A comparison of web service interface similarity measures. *Proceedings of the European Starting AI Researcher Symposium* (STAIRS). 220–231. Riva del Garda, Italy.
Kulvatunyou, B., Cho, H., Son, Y., 2005. A semantic web service framework to support intelligent distributed manufacturing. *International Journal of Knowledge-based Intelligent Engineering Systems* 9. 107–127.
Lian, W., Cheung, D., Mamoulis, N., Yiu, S., 2004. An efficient and scalable algorithm for clustering XML documents by structure. *IEEE Transactions on Knowledge and Data Engineering* 16. 82–96.
Lu, W., Traore, I., 2005. Determining the optimal number of clusters using a new evolutionary algorithm. *Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence* (ICTAI 05). 712–713. Washington, DC.
Luo, J., Montrose, B., Kim, A., Khashnobish, A., Kang, M., June 6-9, 2006. Adding OWL-S support to the existing UDDI infrastructure. *Proceedings of IEEE International Conference on Web Services* (ICWS'06). 153–162. San Diego, CA.
Nierman, A., Jagadish, H., June 6-7, 2002. Evaluating structural similarity in XML documents. *Proceedings of the 5th International Workshop on the Web and Database* (WebDB2002). 61–66. Madison, WI.
O'Brien, L., Bass, L., Merson, P., 2005, Quality attributes and service-oriented architectures. *Technical Report CMU/SEI-2005-TN-014*, Carnegie Mellon University.
Paolucci, M., Kawamura, T., Payne, T., Sycara, K., May 27-28, 2002. Importing the semantic web in UDDI. *Proceedings of the International Workshop on Web Services, E-Business, and the Semantic Web*. 225–236. Toronto, Canada.
Pokraev, S., Koolwaaij, J., Wibbels, M., June 23-26, 2003. Extending UDDI with context-aware features based on semantic service descriptions. *Proceedings of the 2003 International Conference on Web Services* (ICWS'03). 184–190. Las Vegas, NV.
Salton, G., Wong, A., Yang, C., 1975. A vector space model for automatic indexing. *Communications of the ACM* 18. 613–620.
Srinivasan, N., Paolucci, M., Sycara, K., July 6, 2004. Adding OWL-S to UDDI, implementation and throughput.

Proceedings of the 1st International Workshop on Semantic Web Services and Web Process Composition. Sandiego, CA.
Wu, J., Wu, Z., July 11-15, 2005. Similarity-based web service matchmaking. *Proceedings of the 2005 IEEE International Conference on Service Computing* (SCC'05). 287–294. Orlando, FL.
Wu, Z., Palmer, M., June 27-30, 1994. Verb semantics and lexical selection. *Proceedings of the 32nd Annual Meeting of the Associations for Computational Linguistics*. 133–138. Las Cruces, NM.
Xu, R., Wunsch, D., 2005. Survey of clustering algorithms. *IEEE Transactions on Neural Networks* 16. 645–678.
Yang, J., Ceung, W., Chen, X., March 29-April 1, 2005. Learning the kernel matrix for XML document clustering. *Proceedings of the 2005 IEEE International Conference on e-Technology, e- Commerce and e-Service*. 353–358. Hong Kong.
Zeng, L., Benatallah, B., Dumas, M., Kalagnanam, J., Sheng, Q., May 20-24, 2003. Quality driven web services composition. *Proceedings of the 12th International World Wide Web Conference*. 411–421. Budapest, Hungary.

BIOGRAPHICAL SKETCH

Choonghyun Lee is currently working at National IT Industry Promotion Agency, NIPA Korea as a Director of Knowledge & Service Industries Policy Team. He is a Ph.D. candidate in Industrial Engineering from Yonsei University, Seoul, Korea.

Young Hoon Lee is an Associate Professor of Department of Information and Industrial Engineering at Yonsei University, Seoul, Korea. He received his Ph.D. in Industrial Engineering from Columbia University, New York in 1992. Prior to his current position, he worked for Samsung Electronics, Semiconductor Division, Korea. His research interests include production planning and scheduling, operations research application on manufacturing and service industry.

Yun Peng is a Professor in the Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County. His research interests include representing and reasoning with uncertainty in artificial intelligence, machine learning, intelligent agents and agent systems, semantic web, integration and interoperability in e-business and supply chain logistics, neural networks, and artificial life. He has been principle investigator or Co-PI for several large research projects sponsored by NSF, NIST, DARPA, and NSA, and published extensively in these areas.

Hyunbo Cho is a professor of department of industrial and management engineering at the Pohang University of Science and Technology. He received his B.S. and M.S. degrees in Industrial Engineering from Seoul National University in 1986 and 1988, respectively, and his Ph.D. in Industrial Engineering with a specialization in Manufacturing Systems Engineering from Texas A&M University in 1993. He was a recipient of the SME's 1997 Outstanding Young Manufacturing Engineer Award. His areas of expertise include Supply Chain Management, Manufacturing Management and Strategy, and Open Business Model. He is an active member of IIE and SME.