Journal of Theoretical and Applied Information Technology
10th April 2014. Vol. 62 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SEMANTIC SCHEMA MATCHING APPROACHES: A REVIEW

[1]JAFREEN HOSSAIN, [2]NOR FAZLIDA MOHD SANI, [3]LILLY SURIANI AFFENDEY, [4]ISKANDAR ISHAK, [5]KHAIRUL AZHAR KASMIRAN

Faculty of Computer Science and Information Technology, Universiti Putra Malaysia

E-mail: [1]jafreen@gmail.com, [2]fazlida@upm.edu.my, [3]lilly@upm.edu.my, [4]iskandar_i@upm.edu.my, [5]k_azhar@upm.edu.my

ABSTRACT

An extensive review of the existing research work in the field of schema matching uncovers the significance of semantics in this subject. It is beyond doubt that both structural and semantics aspect of schema matching have been the topic of research for many years and there are strong references available for both. However, an in-depth analysis of all the available approaches suggests there are further scopes for improvement in the field of semantic schema matching. Normalization and lexical annotation methods using WordNet have been proposed in several studies, but the level of matching accuracy in those studies have not yet reached a point that can encourage full automation of schema matching in commercial use. This paper lists out several possible future work based on the existing limitations.

Keywords: *Database Integration, Schema Matching, Data Heterogeneity, Semantic Schema Matching, Schema Label Normalization, Stop-Words*

1. INTRODUCTION

The advancement of information and communication technology has opened doors for many data sources to communicate with each other in a semantic web. At the same time it has created data heterogeneity problems in various application domains. Large amount of data is created every day by different sources in different formats. The value of data increases when it can be linked with other data, thus data integration is a major creator of value. So, data integration and data sharing are getting important for many application domains. But at the same time, the semantic integration is getting crucial and complex due to this large scale data and its heterogeneous nature. This heterogeneity can be in terms of data source format, types, representation, or semantic interpretation.

The schema matching problem is considered by many researchers as one of the bottlenecks for semantic integration. It is not a new research area and has received increasing attention since the 1970s [14]. Numerous matching approaches, strategies and algorithms have been developed. Schema matching is the task of identifying semantic correspondences between elements of metadata structures such as database schemas, entity relationship diagrams, and ontologies. It is significant for interoperability and data integration in various applications such as data warehousing, integration of web sources, and ontology alignment in the semantic web. In this review paper, we focus on schema matching in the context of data integration.

Currently, the schema matching process has improved from fully manual to semi-automatic after years of research by numerous researchers. The process is still not fully automated, has shortcomings in lots of areas, and needs improvements that consider the increasing number of data, schema and data sources. Schemas developed for different application domains can be dissimilar in nature, i.e. although the data is semantically related, the structure and syntax of its representation are different.

Automatic or semi-automatic schema matching has to deal with problems arising from the heterogeneity of data sources which can be distinguished into two main types of heterogeneity: structural and semantic heterogeneity [5, 17]. Structural heterogeneity means differences among attribute types, formats, or models whereas semantic heterogeneity means differences in the meaning of schema elements. In this paper, we will mainly focus on semantic heterogeneity and its probable solutions.

Journal of Theoretical and Applied Information Technology
10th April 2014. Vol. 62 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Furthermore, we shall discuss schema normalization approaches and lexical annotation methods which are closely related to the schema matching process. It has been proven that schema normalization approaches improve the lexical relationship and matching accuracy among schema labels. Lexical annotation (i.e. annotation with reference to a lexical resource/dictionary, e.g. WordNet) helps to relate a "meaning" to schema labels. However, the accuracy of semi-automatic lexical annotation methods on real life schemas still suffer from the problem of non-dictionary words such as compound words (CWs), abbreviations and acronyms. Schema normalization approaches can help to resolve this problem and increase the number of similar schema labels.

2. SCHEMA MATCHING

Schema matching has been the focus of research for quite some time. This topic is important in sectors like e-commerce, web technologies, marketing, and the health care sector [29, 9]. Several studies have been conducted to address the schema matching problems.

2.1 Definitions

Definition 1: (Schema) A schema is a set of elements connected by some structure. Examples include SQL schema, XML schema, entity-relationship diagrams, ontology descriptions, interface definitions, or form definitions.

Definition 2: (Schema Matching). Schema matching is a process that takes two heterogeneous schemas (e.g. S1 and S2 in Figure 1) as input and produces as output a set of mappings.

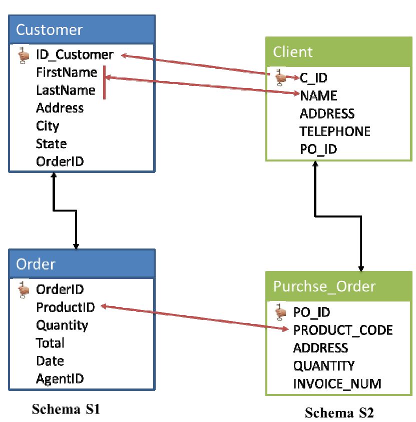

Figure 1: A simple schema matching demonstration

In Figure 1, each mapping indicates that certain elements of the schema S1 are related to certain elements of the schema S2. Mappings may be accomplished by using a set of semantic correspondences (e.g., ProductID = Product_Code) between different schemas.

2.2 Schema Matching Process

Schema matching is a multi-step process. Different researchers have developed different methods for accomplishing the task. Figure 2 shows the general workflow of the COMA schema matching tool [9].

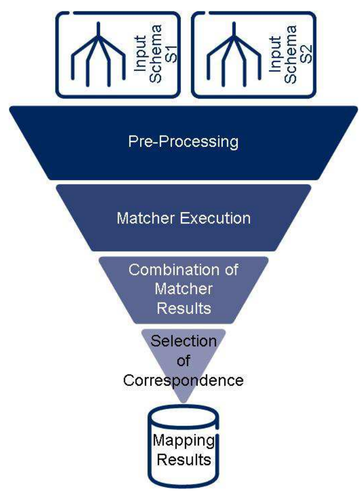

Figure 2: Schema Matching Process

2.3 Schema Matching Application Areas

In the database field, schema matching is usually the first step in generating a program or view definition that maps instances of one schema into instances of another. For example, it arises in object-to-relational mappings, data warehouse loading, data exchange, and mediated schemas for data integration. In knowledge-based applications such as life science applications and the semantic web, it arises in the alignment of ontologies. For example, it may be used to align gene ontologies or anatomical structures. In health care, it may arise in the alignment of patient records and other medical reports. In web applications, it may be used to align product catalogs. In e-commerce, it may be used to align message formats representing business documents such as orders and invoices [4].

Journal of Theoretical and Applied Information Technology
10th April 2014. Vol. 62 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

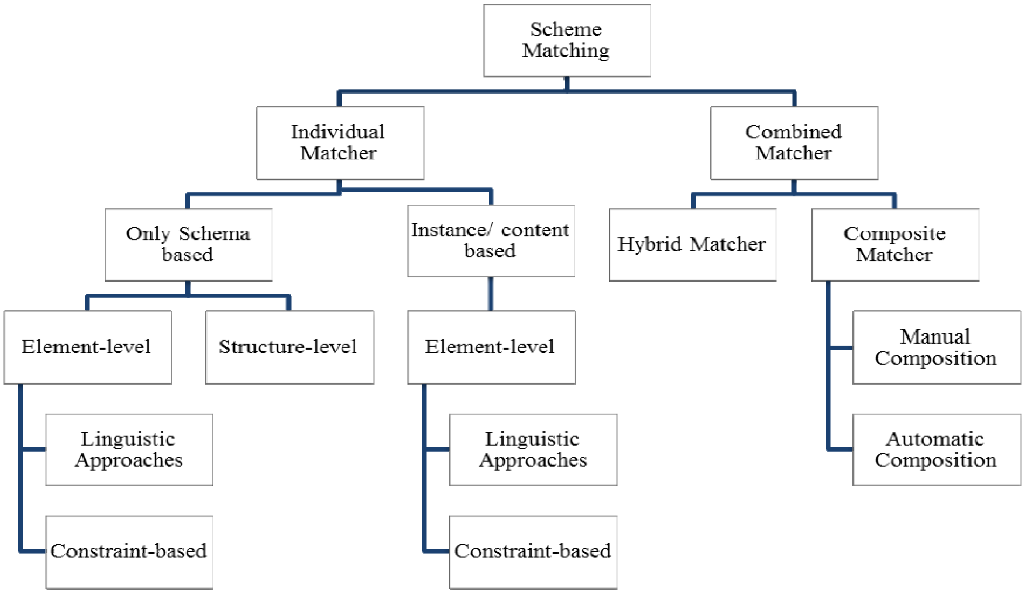

Figure 3: Schema Matching Approaches

2.4 Schema Matching Approaches and Evolution

In 2001, Rahm and Bernstein [29] presented a classification of schema matching approaches which differentiated between schema and instance level, element and structure level, and language and constraint based matching approaches. Figure 3 shows a categorized view of the approaches. Later, many other schema matching approaches have been developed according to the need of specific domains.

Individual matchers: This category includes schema-based and instance-based matchers, element and structural-level matchers, and linguistic and constraint-based matchers. Moreover, the cardinality and the use of external information (like thesauri) are also taken into account.

Individual vs. combinational matcher: A single algorithm is used by an individual matcher to perform the match process. For combinational matchers, two types of combinational matching can be done: (1) hybrid matchers take into account multiple criteria to perform the matching task, and (2) composite matchers run separate match algorithms on two schemas and combine the result.

Different combinational matching approaches have been proposed by different researchers. Cupid, developed by Jayant Madhavan [24], discovered mappings between schema elements based on their names, data types, constraints, and schema structure using a broader set of techniques than past approaches. Some of the innovations used were the integrated use of linguistic and structural matching, context-dependent matching of shared types, and a bias toward leaf structure where much of the schema content resides. Do and Rahm [9] proposed COMA, a Combined Match approach which showed the high value of reuse-oriented strategies, provided better results than previous approaches and compensated for shortcomings of individual matchers. Similar methods were presented by Karasneh et al. [17] which additionally had the flexibility of being domain independent. Bergamaschi et al. [3] proposed MOMIS (Mediator Environment for Multiple Information Sources) which is a framework to perform information extraction and integration from multiple structured and semi-structured heterogeneous data sources.

Schema vs. instance: For schema based approaches, schema-level information is considered such as metadata, element names, data types, and structural properties/models whereas in instance-based approaches data and data content are considered.

Different instance/content based approaches using artificial intelligence and data-mining tools have been developed over time. Kim et al., [20] developed a clustering based schema matching approach which increased the recall rate of method matching by computing more accurate scores which are higher for correctly-matched pairs and lower for incorrectly-matched pairs (one is in a source and another is in a target interface). In

Journal of Theoretical and Applied Information Technology
10th April 2014. Vol. 62 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

addition, it also increased method matching precision without losing correctly matching pairs. Yang, Y. et al. [36] projected an effective content based approach for improved performance results of schema matching. It can either work independently or work together with other schema matching methods.

Element vs. structure: The match action can be compared and matched for single schema elements such as attributes, or the same action can be applied for group of elements that appear together in a structure.

Linguistic vs. constraint-based: The linguistic matching approach considers the name and textual descriptions of schema labels or elements. Different methods including N-gram, EditDistance and SoundEX are used in the linguistic approach [9]. On the other hand, the constraint based approach considers element constraints such as data types, uniqueness, and keys.

Match cardinality: Different matching cardinality (e.g., 1:1, n:1, 1:n, n:m) can be obtained between one or more elements of the first schema with one or more of the second one. Such match relations may in turn be denoted as single or multiple correspondences.

Auxiliary information: Different schema matchers use different auxiliary sources such as, dictionary or thesauri for matching. WordNet is a common external source and used by many systems like MOMIS [3], S-Match [33], Cupid [24] during the schema matching process.

In 2011, Bernstein, P. A., et al. [4] published a revised paper describing different strategies, tools, methods, algorithms, and approaches to perform schema matching that have been used in recent years in different application domains including commercial domains.

Graph matching, usage-based matching, document content similarity, and document link similarity are some newly discussed algorithms. Strategies have been proposed to flexibly combine multiple matching algorithms and to scale to large schema, such as workflow-like strategies, self-tuning match workflows, early search space pruning, partition-based matching, parallel matching, and optimization strategies. Approaches proposed for domain specific schemas include reuse-based matching and holistic matching. Also, different strategies have been incorporated in order to increase user interaction and feedback in the matching process including improved Graphical User Interfaces (GUIs), incremental matching, Top-k matching, Collaborative, wiki-like, and Google-distance [27] user involvement to provide, improve and reuse mappings [4, 10-11].

2.5 Semantic Schema Matching

The meaning/semantics of schema labels plays an important role in the process of determining mappings/matching among various data sources. It is possible to discover semantic correspondences among the elements of different schemas by correctly identifying both the implicit and explicit meaning of schema labels. This identification requires the development of a method for lexical annotation (i.e. finding the meanings of a schema label in a thesaurus or a reference lexical database). Several methods and tools address this problem by using lexical knowledge in different ways.

2.5.1 Different approaches

In order to resolve semantic conflicts and interoperability problems in health care environments, Lee, C. Y., et al. [21] proposed an attribute matching algorithm which does the semantic similarity matching in two steps, first by checking the attribute similarity with domain knowledge and the help of WordNet and secondly by checking word relatedness through overlapped phrases, hypernyms and hyponyms.

Partyka, J., et al. [27] mentioned that semantic heterogeneity among different data sources is still an extensive problem and requires innovative solutions. The traditional N-gram method often fails because it depends mainly on shared instances to discover similarity, which results in an overestimation of semantic matching between independent attributes. They proposed an approach which initially examines the instances of the chosen attributes and computes a similarity value between them, which is known as an entropy-based distribution (EBD). Then they compared the N-gram method and the new TSim method for calculating EBD. They also used K-medoid and Normalized Google Distance for clustering.

Chena, N., et al. [8] stated that the Syntactic schema matching method is often unable to identify possible semantic mapping relationships; for example, element 'abstract' and element 'description' have identical semantics, yet they cannot be identified by the Syntactic method. They proposed the Node Semantic Similarity (NSS) method based on WordNet, conjunctive normal

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

forms and a vector space model. A hybrid algorithm based on label meanings and annotations was designed to compute the relationship between label concepts. The semantic relationship is then translated between nodes into a propositional formula which verifies the validity of this formula to confirm the semantic relationships. The algorithm first calculates the label and node concepts and then computes the conceptual relationship.

Zhao, C. [37] proposed a multilayer schema matching approach: a first layer finds out semantic similarity whereas a second layer introduces functional dependency to formulize structural information of schemas. A third layer proposes a probabilistic factor. Finally, the mapping element pairs with composite and reasonable consideration of each layer's results are selected. The semantic similarity measure initially works on data preprocessing, then it does the lexicographic similarity measure based on WordNet and finally generates the candidate matching sets.

Islam, A. and Inkpen, D. [14] mentioned that in databases, the text similarity used in schema matching to solve semantic heterogeneity is a significant problem in any data sharing system whether it is a data integration system, a distributed database system, a web service, or a one-to-one data management system. They recommended a Semantic Text Similarity (STS) method which discovers the similarity of two texts in terms of semantic and syntactic information (by common-word order). Three similarity functions are considered in order to derive a more general text similarity approach. String similarity and semantic word similarity are considered at the beginning and then an optional common-word order similarity function was introduced to combine syntactic information. Finally, the text similarity is derived by merging string similarity, semantic similarity and common-word order similarity with normalization.

Gillani, S. [12] defined a taxonomy of all possible semantic similarity measures and also proposed an approach that exploits semantic relations stored in the DBpedia dataset while utilizing a hybrid ranking system to dig-out the similarity between nodes of two graphs.

2.5.2 Semantic similarity of non-dictionary words

Measuring similarity of semantics refers to matching the similarity between two schema labels that have the same meaning or related information, but may not be lexicographically similar [23]. This is a key challenge in several computing areas. For example: in data warehouse integration when creating mappings that link mutual components of data warehouse schemas semiautomatically [1-2], or while matching identity when personal information or social identity are used [22], or in the entity resolution field when two given text objects have to be compared [19]. The problem here is that semantic similarity evolves over different time and domains [6]. The traditional approaches for solving such problems have included usage of manually developed taxonomies like WordNet [7]. However, with the emergence of social networks or instant messaging systems [30], a lot of terms (proper nouns, brands, acronyms, new words, and so on) are not included in these kinds of taxonomies; as a result, similarity matching methods that are dependent on these kinds of resources cannot be used in these tasks.

Sorrentino, S., et al. [35] proposed a schema normalization method called NORMS and also described an automatic lexical annotation method called PWSD. NORMS can identify, normalize and annotate the abbreviation and Compound Nouns (CNs) in schema labels with the help of PWSD. PWSD is a probabilistic WSD (Word Sense Disambiguation) algorithm which scores a probability value for every annotation, representing the reliability of the annotation itself [28]. PWSD has five WSD algorithms, each generating a probability allocation based on semantics, and it can be easily extended to the use of other WSD algorithms. It combines the results of each WSD algorithms by using the theory of combination of Dempster-Shafer. Starting from the probabilistic annotations, it is possible to identify relationships among schemas based on probabilistic lexical similarity. The PCT MOMIS component collects the probabilistic lexical relationships and the regular structural relationships, which is extracted from schemas by the description logic tool ODBTools.

Martinez-Gil, J. and Aldana-Montes, J. F. [25] designed and evaluated four algorithmic ways for measuring the semantic similarity amid terms

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Different Methods Of Solving Semantic Similarity

Sl	Author/ Year	Method Discussed	Approach
1	Nastase, V., et al., 2006	Studied the performance of two representations of word meaning in learning noun-modifier semantic relations. One representation is based on lexical resources, in particular WordNet, the other on a corpus. Then they experimented with decision trees, instance-based learning and support vector machines.	Instance based
2	Islam, A. and Inkpen, D, 2008	Semantic Text Similarity (STS) method determines the similarity of two texts by combining string similarity, semantic similarity and common-word order similarity with normalization.	Schema based
3	Lee, C. Y., et al., 2009	An attribute match algorithm which checks the attribute similarity firstly with domain knowledge and the help of WordNet, and secondly by checking word relatedness through overlapped phrases, hypernyms and hyponyms.	Schema based
4	Partyka, J., et al., 2009	Examines the instances of the chosen attributes and calculates a similarity value between them, known as entropy-based distribution (EBD). Then compares N-gram and the new TSim algorithm for calculating EBD. Also uses K-medoid and Normalized Google Distance for clustering.	Instance based
5	Chena, N., et al., 2012	Node semantic similarity (NSS) method based on WordNet, conjunctive normal form and a vector space model. Also a hybrid algorithm based on label meanings and annotations designed to calculate the similarity between label concepts.	Schema based
6	Zhao, C., 2012	A multilayer approach: 1st layer finds semantic similarity by lexicographic similarity measure based on WordNet. 2nd layer introduces functional dependency to formulize structural information of schemas. 3rd layer proposes a probabilistic factor. Finally, the mapping element pairs with composite and reasonable consideration of each layer's result are selected.	Combined approach
7	Gillani, S., 2013	Defined taxonomy of all possible semantic similarity measures; moreover also proposed an approach that exploits semantic relations stored in the DBpedia dataset while utilizing a hybrid ranking system to dig-out the similarity between nodes of the two graphs.	Combined approach
8	Sorrentino, S., et al., 2011	Proposed a schema label normalization method called NORMS including abbreviation expansion and Compound Noun annotation method and also described an automatic lexical annotation method called PWSD.	Schema based

utilizing their associated history search patterns. These algorithmic methods are: a) frequent co-occurrence of terms in search patterns, b) computation of the relationship between search patterns, c) outlier coincidence on search patterns, and d) forecasting comparisons. They have shown experimentally that some of these methods correlate well with respect to human judgment when evaluating general purpose benchmark datasets, and significantly outperform existing methods when evaluating datasets containing terms that do not usually appear in dictionaries.

Nastase, V., et al. [26] compared the performances of WordNet and Corpus in learning noun-modifier semantic relations. Then they tested the results with three methods: i) decision trees, ii) instance-based learning and iii) support vector machines. The corpus based method performed well over the baseline. It had the advantage of functioning with data without word-sense annotations. The WordNet-based method however had higher precision but with the disadvantage of requiring data with word-sense annotation.

Table 1 lists the semantic schema matching approaches discussed in this section.

2.6 Relationship among Schema and Ontology Matching

Ontology describes concepts used for representing knowledge on the web, for example, annotating a picture, specifying a web service interface or expressing the relation between two persons. There are a number of languages for ontologies, both registered and standard-based. OWL (Web Ontology Language) denotes the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ontology W3C standard. OWL is a language for making ontological statements, developed as a follow-on from RDF (Resource Description Framework) and RDFS (RDF Schema).

Similarly with schema matching, ontology matching deals with multiple, distributed, and evolving ontologies. Ontologies can be viewed as schemas for knowledge bases [31]. Therefore, techniques developed for schema matching in the great majority of the cases may be applied in the ontology matching context.

Schema and ontology matching problems are strictly connected even if they present some significant differences. Most of the time, the explicit semantics of database schemas are not available for their data: semantics/meaning of a database schema is generally specified during design time and frequently is not becoming a part of a database specification, therefore it is not available. On the other hand, ontologies are logical systems that follow some formal meaning, that is, ontology definitions can be interpreted as a set of logical axioms. Furthermore, while schema matching is generally executed with the help of methods which tries to find out the semantics or meaning encoded in the schemas, ontology matching systems try to discover knowledge specifically encoded in the ontologies [31].

Regardless of the differences between schema and ontology matching problems, the techniques developed for each of them can be of mutual benefit.

Different researchers are working on ontology matching approaches and several approaches have been emerging. Hlaing [13] proposed a system architecture for schema matching with specific domain ontologies which handle semantic heterogeneity for relational databases. Kavitha, C., et al. [18] identified that interoperability is the main problem when heterogeneous databases are integrated. They proposed an approach which uses a domain specific master ontology for integration of local ontologies created from heterogeneous databases. The major steps involved include Class Name Matching, Property Name Matching using N-grams and synonyms, Property Type Matching and Property Value Classification. Jian, N., et al. [16] developed FalconAO which is an automatic tool for aligning ontologies. There are two matchers integrated in FalconAO: one is a matcher based on linguistic matching for ontologies called LMO; the other is a matcher based on graph matching for ontologies called GMO.

3. DISCUSSION

In the initial sections of this paper, different schema matching approaches, strategies, applications areas and methods by former researchers were discussed. The discussion on the later part of the paper was more focused on semantic schema matching approaches and its significance in the overall process. The discussion shows that in semantic schema matching, it is very important to know the implicit meaning of the schema labels to be matched which is often difficult to accomplish by traditional N-gram methods.

Table 1 lists different methods developed by previous researchers on semantic schema matching approaches. Some of the methods used schema based approaches and other methods used instance based approaches.

Having non-dictionary words in schema labels is one significant recent research topic in this domain. Most of the researchers used auxiliary sources like WordNet to find the meaning of the labels. Although external dictionaries or thesauri like WordNet are rich with wide networks of word meanings and their semantic relationships, they do not cover different domain knowledge with the same kind of detail. Also, many domain-specific non-dictionary words may not be present in them. Some solutions around this limitation have been researched as well, but they are quite limited in scope and further studies are required.

In the latter part of this paper, the relationship between schema and ontology has been discussed, considering the emerging significance of ontology in any semantics study.

4. CONCLUSION

In this paper, we did a review on some previous schema matching approaches, strategies and techniques till recent times. It can be concluded from this review that the implicit meaning or semantics of schema labels plays an important role in the exercise of discovering mappings between different data sources.

Although many strategies were developed to solve this problem including schema normalization approaches [34], there is still room for improvement and future work. Future work may include finding the meaning of domain specific terms, different compound words having

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

prepositional-verbs, conjunctions, digits or stop-words in schema labels. Also more work can be done to improve the number of false positive and false negative relationships. Another relevant future research could possibly be the inclusion of instance-based matching techniques to improve the automatic annotation and relationship discovery processes among schema labels.

5. ACKNOWLEDGEMENT

Special thanks from authors for financial support (Prototype Research Grant Scheme, PRGS) from the Ministry of Education (MoE), Malaysia via Universiti Putra Malaysia. The principal investigator of this research project is Assoc. Prof. Dr. Nor Fazlida Mohd Sani.

REFRENCES:

[1] Banek, M., Vrdoljak, B., Tjoa, A.M., Skocir, Z. (2007). Automating the Schema Matching Process for Heterogeneous Data Warehouses. In DaWaK (pp. 45–54).

[2] Banek, M., Vrdoljak, B., Tjoa, A.M. (2007). Using Ontologies for Measuring Semantic Similarity in Data Warehouse Schema.

[3] Bergamaschi S, Castano S, Vincini M. (1998). MOMIS, An Intelligent System for the Integration of Semi Structured and Structured Data, INTERDATA.

[4] Bernstein P. A., Madhavan J., Rahm E. (2011). Generic Schema Matching, Ten Years Later, Proceedings of the VLDB Endowment, Vol 4, No.11.

[5] Bergamaschi, S., Beneventano,D., Po, L., Sorrentino, S. (2011). Automatic Normalization and Annotation for Discovering Semantic Mappings, Search Computing II, LNCS 6585, pp. 85–100, Springer

[6] Bollegala, D., Honma, T., Matsuo, Y., Ishizuka, M. (2008). Mining for personal name aliases on the web. In WWW (pp. 1107–1108).

[7] Budanitsky, A., Hirst, G. (2006). Evaluating WordNet-based Measures of Lexical Semantic Relatedness. Computational Linguistics, 32(1), 13–47.

[8] Chena, N., Heb, J., Yanga, C., Wanga, C. (2012). A node semantic similarity schema-matching method for multi-version Web Coverage Service retrieval, International Journal of Geographical Information Science.

[9] Do H-H, Rahm E. (2002). COMA - A system for flexible combination of schema matching approaches, Proceedings of the 28th VLDB Conference, Hong Kong, China.

[10] Feng Y., Zhao L., Yang J. (2010). GATuner: Tuning Schema Matching Systems using Genetic Algorithms, IEEE.

[11] Gal, A., Sagi, T., Levy, E., Miklos, Z., (2012). Making Sense of Top-k Matching, IIWeb, Scottsdale, AZ, USA.

[12] Gillani S., Naeem, M., Habibullah, R., Qayyum, A., (2013). Semantic Schema Matching Using Dbpedia, I.J. Intelligent Systems and Application.

[13] Hlaing S. S. (2009). Ontology based Schema Matching and Mapping Approach for Structured Databases, November 24-26, Seoul, Korea, ICIS.

[14] Islam, A., Inkpen, D. (2008). Semantic text similarity using corpus-based word similarity and string similarity, ACM Trans. Knowl. Discov. Data. 2, 2, Article 10.

[15] Islam, A., Inkpen, D. Z., Kiringa, I. (2008). Applications of corpus-based semantic similarity and word segmentation to database schema matching. VLDB J., 17(5):1293–1320.

[16] Jian, N., Hu, W., Cheng, G., Qu. Y. (2010). FalconAO: Aligning Ontologies with Falcon, Department of Computer Science and Engineering Southeast University.

[17] Karasneh Y., Ibrahim H., Othman M., Yaakob R. (2010). Challenges in Matching Heterogeneous Relational Databases Schemas, IKE'10 - 9th International Conference on Information and Knowledge Engineering – USA.

[18] Kavitha, C., Sadasivam, G. Sudha, S., Shenoy, Sangeetha N. (2011). Ontology Based Semantic Integration of Heterogeneous Databases, European Journal of Scientific Research, ISSN 1450-216X, Vol.64 No.1, pp. 115-122

[19] Kopcke, H., Thor, A., Rahm, E. (2010). Evaluation of entity resolution approaches on real-world match problems. PVLDB, 3(1), 484–493.

[20] Kim B., Namkoong H., Lee D., Hyun S. J. (2011). A Clustering Based Schema Matching Scheme for Improving Matching Correctness of Web Service Interfaces, International Conference on Services Computing, IEEE.

[21] Lee, C. Y., Ibrahim H., Othman M., Yaakob R. (2009). Reconciling Semantic Conflicts in Electronic Patient Data Exchange, Proceedings of iiWAS, Kuala Lumpur Malaysia.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[22] Li, J., Alan Wang, G., Chen, H. (2011). Identity matching using personal and social identity features. Information Systems Frontiers, 13(1), 101–113.

[23] Li, Y., Bandar, A., McLean, D. (2003). An approach for Measuring Semantic Similarity between Words Using Multiple Information Sources. IEEE Transactions on Knowledge and Data Engineering, 15(4), 871–882.

[24] Madhavan, J., Bernstein, P. A., Rahm, E. (2001). Generic Schema Matching with Cupid. In Apers, P. M. G., Atzeni, P., Ceri, S., Paraboschi, S., Ramamohanarao, K., and Snodgrass, R. T., editors, Proc. of the 27th International Conference on Very Large Data Bases (VLDB 2001), September 11-14, 2001, Roma, Italy, pages 49–58. Morgan Kaufmann.

[25] Martinez-Gil, J., Aldana-Montes, J. F. (2013). Semantic similarity measurement using historical Google search patterns, Information Systems Frontiers, Springer Link.

[26] Nastase, V., Sokolova, M., Szpakowicz, S. (2006). Learning Noun-Modifier Semantic Relations with Corpus-based and WordNet-based Features, American Association for Artificial Intelligence.

[27] Partyka, J., Khan, L., Thuraisingham, B. (2009). Semantic Schema Matching Without Shared Instances, International Conference on Semantic Computing, IEEE.

[28] Po, L., Sorrentino, S. (2011). Automatic generation of probabilistic relationships for improving schema matching. Information Systems Journal, Special Issue on Semantic Integration of Data, Multimedia, and Services, 36(2):192208.

[29] Rahm, E., Bernstein, P. A. (2001). A survey of approaches to automatic schema matching, The VLDBJournal 10: 334–350.

[30] Retzer, S., Yoong, P., Hooper, V. (2012). Inter-organisational knowledge transfer in social networks: A definition of intermediate ties. Information Systems Frontiers, 14(2), 343–361. Matching Process. In CONTEL (pp. 227–234)

[31] Shvaiko, P., Euzenat, J. (2004). A classification of schema-based matching approaches. In Proceedings of the Meaning Coordination and Negotiation Workshop at ISWC04.

[32] Shvaiko, P., Euzenat, J. (2005). A survey of schema-based matching approaches. 3730:146–171.

[33] Shvaiko, P., Giunchiglia, F., and Yatskevich, M. (2010). Semantic matching with S-Match. Semantic iWeb Information Management: a Model-Based Perspective, XX:183–202.

[34] Sorrentino, S., Bergamaschi, S., Gawinecki, M., and Po, L. (2010). Schema label normalization for improving schema matching. DKE Journal, 69(12):12541273.

[35] Sorrentino, S., Bergamaschi, S., Gawinecki, M. (2011). NORMS: an automatic tool to perform schema label normalization. In Press, Accepted Manuscript (Demo Paper), IEEE International Conference on Data Engineering, ICDE 2011, April 11-16, Hannover.

[36] Yang Y., Chen M., Gao B. (2008). An Effective Content-based Schema Matching Algorithm, International Seminar on Future Information Technology and Management Engineering, IEEE.

[37] Zhao C., Shen D., Kou Y., Nie T., Yu G. (2012). A Multilayer Method of Schema Matching Based on Semantic and Functional Dependencies, Ninth Web Information Systems and Applications Conference (WISA).