Survey: Models and Prototypes of Schema Matching

Edhy Sutanta[1], Retantyo Wardoyo[2], Khabib Mustofa[2], Edi Winarko[2]
[1] Doctoral Program of Computer Science at Department of Computer Sciences & Electronics Instrumentations, Universitas Gadjah Mada, Yogyakarta, Indonesia
[2] Department of Computer Science & Electronics Instrumentation, Universitas Gadjah Mada, Yogyakarta, Indonesia

Article Info

Article history:

Received Dec 25, 2015
Revised Mar 2, 2016
Accepted Mar 16, 2016

Keyword:

Data integration
Heterogeneous database
Information integration
Schema matching model
Schema matching prototype

ABSTRACT

Schema matching is critical problem within many applications to integration of data/information, to achieve interoperability, and other cases caused by schematic heterogeneity. Schema matching evolved from manual way on a specific domain, leading to a new models and methods that are semi-automatic and more general, so it is able to effectively direct the user within generate a mapping among elements of two the schema or ontologies better. This paper is a summary of literature review on models and prototypes on schema matching within the last 25 years to describe the progress of and research chalenge and opportunities on a new models, methods, and/or prototypes.

Corresponding Author:

Edhy Sutanta,
Doctoral Program of Computer Science,
Department of Computer Sciences & Electronics Instrumentations,
Universitas Gadjah Mada,
S2/S3 Building 4th floor, Sekip Utara, Bulaksumur, Yogyakarta, 55281, Indonesia.
Email: edhy_sst@akprind.ac.id, edhy_sst@yahoo.com

1. INTRODUCTION

Schema matching issues that have emerged since the early 1980s is the fundamental problem in many applications for data/information integration. Simply, schema matching is how to construct a mapping between the two elements of the schema or ontologies have in common [1]. Schema matching is an important issue for the integration of information from multiple heterogeneous sources [2]. Schema matching is also important to realize interoperability and implement the integration of data from different applications [3]. Indeed, the schema matching also use in the schema evolution and reuse of software [4]. Schema matching is part of the topic of Enterprise Application Integration (EAI) in particular Enterprise Information Integration (EII), is an integration task at the back end level with the aim to overcome the problems caused by schematic heterogeneity [5]. The meaning schematic heterogeneity is the difference naming in the schema definition, including the type, format, and precision of data [6]. The main process of schema matching is to develop mapping and matching between elements of inter schema [7].

From initial appearance until the end of 2002, the schema matching process most still done manually [8]. And, only some models that have been developed for common domain and according to the different application and schema languages [5]. Manually schema matching has the disadvantage, among others, requires a long time, boring, and not practically if applied in a case that involve many schema [8]. Manually models are also expensive and most likely there was an error, and therefore needed a new method which is a semi-automatic [6]. Schema matching is an exciting research objects to direct effectively to the user to solve the problem of schema matching [3]. The research on schema matching is still open to finding smarter ways to develop models and software, in particular on the combined use of the methods already exist [9].

The paper provides a summary of the results of the study of literature depicting the development of models and prototypes on schema matching over the last 25 years, as well as showing chances of and challenges of research on schema matching. The rest of the paper is organized as follows. Section 2 describes some basic concept of schema matching. Section 3 illustrates some review on the existing schema matching research. Section 4 explains the future direction for schema matching research. Finally, Section 5 concludes the paper.

2. SCHEMA MATCHING CONCEPPT

2.1. Schema Matching Definition

The term of schemas matching has been defined in different ways by the experts, but all of them have similar meanings. According to [10], schemas matching is a similar job with matching, whereas the [6],[11]-[13] defines a schemas matching as a process to find the relationship between elements of the pair schema. The goal of matching schemas is given input two different schemas, and/or additional information, and input schemas mapping rules, then specify mapping result schemas elements both schemas after verified by the user [14]. Schema matching process involves two schemas or ontologies, one serves as source and the other as a target [1].

2.2. Schema Matching Classification

The schema matching process can involve a wide variety of algorithms, eg. to determine the elements to be matched, the transformation mapping, or merging [15]. Based on algorithm is used, the model in matching schema can be classified into several categories.

A classification of schema matching by [14],[16] consists of schema-based vs. instance-based, element vs. structure granularity, linguistic based, constraint-based, matching cardinality, auxiliary information, as well as individual vs. combinational. A classification according to [3] consists of schema based, instance-based, and reuse oriented. A schema matching classification by [17] divides the model based on the level and type of information that explored include elements and structure, and based on the type of information being explored include terminology (involving aspects of linguistics (consists of language-based and linguistic-based) or involve aspects of linguistics (string-based)), based on structural aspects (including the internal aspect (constraint based) and relational (consists of alignment reuse, graph-based, taxonomy based, and repository structure))), and semantics (consists of upper level formal ontology and model based).

In [1], the schema matching model classified by level components that are matched (conceptual and structure), the level of user intervention (manual and automatic), the method used (stand-alone and combined), and the type of components used as the basis for matching (using the schema or schema and instance). According to [18] a schema matching models consists of RSM (relations schemas matcher), ANM (relations attribute name matcher), DTM (data type matcher), CM (constraint matcher), and IDM (instance of data matcher). In different ways, [5] grouping schema matching algorithm into three type, namely linguistic matcher (NTA (name, connected terms, attributes) linguistic matcher, prefix/suffix based matcher for name, and prefix/suffix based matcher for types), vocabular matchers (WordNet-based word matcher for names and NTA (name, connected terms, attributes) related terms similarity), and structural matchers (flooding similarity, WordNet-based ancestor context similarity, string comparison based child context similarity, child context similarity, and the direct ancestor similarity using string comparison). Similar with [14],[16], a schema matching classification is given by [9], there are covering linguistic matching, auxiliary information, instance-based matching, structure-based matching, constraint-based matching, rule based matching, and hybrid matching.

In another reference, [19] classifies the schema matching models into two categories, namely schema-based and instance-based. A schema based consists of element-based and structure-based. Element-based consists of linguistic-based and constraint-based, while the structure based is developed based on constraint-based. Instance-based model are developed based on the element level which consists of linguistic based, constraint-based and learning-based.

According to [9], several other models ever developed including those utilizing additional information, namely graph matching, usage-based matching, document content similarity, and document link similarity. A classification combinations model of schema matching for large schema is covering independent strategy or sequence execution or combination of execution, parallel matching, self-tuning match work flow, early search space pruning, partition-based matching, and schema optimization [9]. Schema matching model in a particular domain, i.e. reuse based matching and holistic matching is a different approach [9]. Some strategies that include interaction and feedback from users in the matching process have also been developed, including GUI (graphical user interface) support, incremental matching, Top-k matching, and collaborative [9]. Another strategy is the use of semantic matching algorithms to extend, for example semantic tagging and conditional tagging [9].

2.3. User Intervention

The main problem in the schema matching is often found naming the schema that is not clear, difficulty found synonymous in naming, or differences in schema definition language [6]. Thus the schema matching model is not likely to produce mapping schema that is 100% accurate accordance with user expected [6]. For these reasons, the schema matching cannot be fully done automatically, usually must be corrected by the user to obtain the correct final results [15].

There are two cases in which the schema matching will experienced failure and require user involvement. First, when the source element schema cannot be matched by any element in the targets schema using the rules used, or second, if the source element schema produces some of the elements which are considered suitable in the targets schema and the system cannot determine the best fit elements automatically [20].

According to [21], generally matching the two schemas requires information that is not always available in the schema and cannot be done automatically, so it requires the involvement of the users to review and determine suggestions on schema matching results. Schema matching process could never be done automatically fully there is a complete semantic matching model for the information systems integration [22]. Another reason that causes the schema matching cannot be done automatically is the naming conflict and levels of abstraction conflict [4].

2.4. Individual vs Combinatorial Matchers

Schema matching models can be developed using the individual or combinational matchers [14],[16]. Individual matcher faster in process of completion, but has the disadvantage is only appropriate in certain cases, so that generally requires more than one matcher which combined [19]. According to [16],[19], the *combinational matchers* can be implemented as composite or hybrid. The term of composite matchers is synonymous with inter-matcher parallelism, while hybrid matcher is synonymous with intra-parallelism [23]. Hybrid models uses multiple criteria simultaneously matching [20],[24],[25], while composite matchers run separately or independent algorithms and combines at the results [26].

Thus, hybrid matcher combines two different methods are processed simultaneous, whereas composite matcher combine two methods that are processed in a sequence that is a method to be implemented after the other method is completed. According to [16] hybrid matcher is to combine more than one method simultaneously to perform matching between the schema elements, and should give better results and improved performance (effectiveness) rather than individual matcher.

3. SCHEMA MATCHING MODELS AND PROTOTYPES

The study found at least 34 models and prototypes on schema matching in 71 scientific publications that are relevant in the last 25 years. Models and prototypes first on schema matching is SEMINT [27], while the latest is COMA 3.0 [28]. Each of models and prototypes may use an input schemas such a relational model (RDF/Relational Database Format), XML model (DTD/Document Type Definition, or XSD/W3C XML Schema), or ontology (OWL/Web Ontology Language). Prototype schemas matching has been evaluated extensively by [3],[8],[29], and the results showed that most schemas matching prototype developed for a limited scope or specific, and some others there were developed specifically by utilizing the ontology.

Refers to [18],[30],[31], the study [4] was to integrate heterogeneous database based on semantic ontology, and tested on academic database that is defined using MS Access and MySQL. Discovery a model for connecting between data sources on the web by [32],[33] is also a research area in which development of a semantic approach to the relationship between entities is required and is valid for widespread availability of ontology.

Furthermore, refers to [28], the study [9] has studied prototypes schema matching that developed in the year 2001-2011 and then compared prototypes (Cupid, COMA ++, ASMOV, Falcon-AO, RiMOM, AGREEMENTMAKER, and OpenII) on aspects of architecture , schema representation, representation schema mapping, information an input and matching algorithms, as well as the execution matching on schema element. According to [9], prototypes using different approaches in the matching algorithm between schema elements, while COMA ++, Falcon-AO, RiMOM, and AgreementMaker is the prototype that combines three methods, ie linguistic, structure, and instance-based. Use of external dictionaries, such as thesaurus are generally used to improve accuracy of matching linguistic.

Limited GUI support has been provided by several prototypes [34],[29], and partially are able to do on two schema ontology matching [1],[35],[36]. Several prototypes were participate in OAEI (Ontology Alignment Evaluation Initiative) increased significantly, but still need to be developed to overcome the problem of schema matching on a wider scope [37]. Sophisticated techniques such as partitioning schema, parallel matching, reuse in mapping and self-tuning capabilities (e.g., dynamic matching options) is supported to a limited extent [28].

Models and prototypes of hybrid schema matching ever developed earlier was CLIO [38]-[42], and research by [43]. While SEMINT [25],[27],[44], LSD [26], the Cupid [14], COMA [45], COMA ++ [3], COMA 3.0 [28],[46], IMAP [47], PROTOPLASM [48]-[51], FALCON-AO [6] and [52], as well as the ASMOV [53] was developed using a combination of methods as composite.

Linguisitic based matching method used on DIKE [54]-[56], MOMIS [57],[58], ONION [59]-[61], ARTEMIS [62], UNIFORM [63], WISE-INTEGRATOR [2],[64],[65], PROMPT [66],[67], RONDO [21], OLA [68], Qom [69],[70], S-MATCH [62],[71],[72], RiMOM [10], AGREEMENT MAKER [1], OPENII [36] and [4]. While DELTA [73],[74], similarity flooding [75], XCLUST [76], and research by [77] implements the structure-based matching method.

Usage constraint in the on schema matching assumes that constraint has a meaning to set a similarity database element, for example, attribute AT1 in table X is defined as a character was same as attribute AT2 in the table Y which is defined as a text [31]. According to [19], the use of constraint-based is part of model group on schema matching which is included in level structure, but not described in more about what properties which explored and included as constraint. Instance-based method is used on TRANSCM [20], Autoplex [78], Automatch [79]-[81], GLUE [82],[83], SCM [84], as well as DUMAS [85].

Auxiliary based matching such a dictionary, WordNet, or Corpus are used on DIKE [54]-[56], MOMIS [57],[58], ONION [59]-[61], ARTEMIS [62], the Cupid [14, 80, 81], COMA [45], XCLUST [76], UNIFORM [63], WISE-INTEGRATOR [2],[49],[64],[65], OLA [50],[68], S-MATCH [51],[62],[71],[72],[77], COMA ++ [3],[6], OPENII [36], as well as the COMA 3.0 [28]. The survey results show Auxiliary based matching such a dictionary, WordNet, or Corpus are used on DIKE [54]-[56], MOMIS [57],[58], ONION [59]-[61], ARTEMIS [62], the Cupid [14],[80],[81], COMA [45], XCLUST [76], UNIFORM [63], WISE-INTEGRATOR [2],[49],[64],[65], OLA [50],[68], S-MATCH [51],[62],[71],[72],[77], COMA ++ [3],[6], OPENII [36], as well as the COMA 3.0 [28]. In detail, the survey results show the comparison of the methods used in the model and prototypes on schema matching shown in Table 1.

Table 1. The comparison of the methods used in the model and prototypes on schema matching

		Input			Using Method [9]							Domain		GUI
Prototype Name	Author /Researcher	**RDF**	**DTD /XSD**	**OWL**	***Linguistic based***	***Structure based***	***Constraint based***	***Instance based***	**Method combination**		**Dictionary/ *WordNet* /Corpus**	**Specific**	**General**	
									Hybrid	**Composite**				
SEMINT	Li & Clifton (1994)	√					√	√		√				
	Li &Clifton (2000)	√					√	√		√				
	Li *et al*. (2000)	√					√	√		√				
DELTA	Benkley *et al*. (1995)	√				√								
	Clifton *et al*. (1996)	√				√								
TRANSCM	Milo & Zohar (1998)		√					√						
DIKE	Palopoli *et al*. (1998)	√			√					√	√			
	Palopoli *et al*. (1999)	√			√					√	√			
	Palopoli *et al*. (2000)	√			√					√	√			
MOMIS	Castano & Antonellis (1999)				√					√	√			
	Bergamaschi *et al*. (2001)				√					√	√			
ONION	Mitra *et al*. (1999)			√	√					√	√			
	Mitra *et al*.			√	√					√	√			

Prototype Name	Author /Researcher	Input			Using Method [9]							Domain		GUI
		RDF	**DTD /XSD**	**OWL**	***Linguistic based***	***Structure based***	***Constraint based***	***Instance based***	**Method combination**		**Dictionary/ *WordNet* /Corpus**	**Specific**	**General**	
									Hybrid	**Composite**				
	(2000)													
	Mitra & Wiederhold (2002)			√	√					√	√			
ARTEMIS	Giunchiglia *et al*. (2005)				√					√	√			
CUPID	Madhavan *et al*. (2001)	√	√		√	√			√		√		√	
LSD	Doan *et al*. (2001)	√					√	√		√		√		
AUTOPLEX	Berlin & Motro (2001)	√						√						
AUTOMATCH	Berlin & Motro (2002)	√						√						
CLIO	Hernández *et al*. (2001)	√	√			√		√	√				√	√
	Naumann *et al*. (2002)	√	√			√		√	√				√	√
	Popa *et al*. (2002)	√	√			√		√	√				√	√
	Haas *et al*. (2005)	√	√			√		√	√				√	√
XCLUST	Lee *et al*. (2002)		√			√				√	√			
COMA	Do dan Rahm (2002)	√	√		√	√				√	√		√	√
GLUE	Doan *et al*. (2002)			√				√						
Similarity-flooding	Melnik *et al*. (2002)	√	√			√							√	
UNIFORM	Palopoli *et al*. (2002)				√						√			
WISE-INTEGRATOR	He *et al*. (2003)		√		√					√	√			
	He & Chang (2003)		√		√					√	√			
	He *et al*. (2004)				√					√	√			
PROMPT	Noy&Musen (2003)			√	√									
	Noy&Musen (2004)			√	√									
RONDO	Melnik *et al*. (2003)	√	√		√								√	
Protoplasm	Bernstein *et al*. (2004)	√	√		√	√				√			√	
IMAP	Dhamankar *et al*. (2004)	√					√	√		√				
OLA	Euzenat *et al*. (2004)			√	√					√	√			
QOM	Ehrig & Staab (2004)			√	√									
	Ehrig & Sure (2004)			√	√									
SCM	Hoshiai *et al*. (2004)			√				√						
S-MATCH	Giunchiglia *et al*. (2004)			√	√					√	√			
	Giunchiglia *et al*. (2005)			√	√					√	√			
DUMAS	Bilke & Naumann (2005)							√						
FALCON-AO	Jian *et al*. (2005)	√		√	√	√		√		√				
COMA++	Do (2005)	√	√	√	√	√		√		√	√		√	√
RiMOM	Li *et al*.			√	√									

Prototype Name	Author /Researcher	Input			Using Method [9]							Domain		GUI
		RDF	**DTD /XSD**	**OWL**	***Linguistic based***	***Structure based***	***Constraint based***	***Instance based***	**Method combination: Hybrid**	**Method combination: Composite**	**Dictionary/ *WordNet* /Corpus**	**Specific**	**General**	
	(2009)													
Agreement-maker	Cruz *et al*. (2009)			√	√								√	√
ASMOV	Jean-Mary *et al*. (2009)			√	√	√				√				√
SYM	Chien & He (2010)	√			√	√			√			√		
OPENII	Seligman *et al*. (2010)	√	√	√	√					√	√			√
COMA 3.0	Rahm, *et al*. (2011)	√	√	√	√	√		√		√	√		√	√
---	Larson *et al*. (1989)						√					√		
---	Hayne & Ram (1990)						√					√		
---	Gotthard *et al*. (1992)						√					√		
---	Spaccapietra & Parent (1992)						√					√		
---	Lerner (2000)						√					√		
---	Mitra *et al*. (2000)						√					√		
---	Castano *et al*. (2001)						√					√		
---	Madhavan *et al*. (2003)						√	√		√				
---	Kang & Naughton (2003)	√				√		√	√	√				
---	Bertino *et al*. (2004)						√					√		
---	Embley *et al*. (2004)			√				√		√	√			
	Xu & Embley (2003)			√				√		√	√			
---	Wang *et al*. (2004)							√						
---	Dragut & Lawrence (2004)	√		√	√	√				√	√		√	√
---	Mork & Bernstein (2004)	√	√	√	√	√				√	√			
---	Lu *et al*. (2005)		√			√				√	√			
---	Tu & Yu (2005)	√		√	√	√		√		√	√		√	√
---	Engmann & Massmann (2007)	√	√		√	√		√		√	√		√	√
---	Kavitha *et al*. (2011)	√		√	√									

Statistically, research, models, and schema matching prototypes that developed in the last 25 years classification based on inputs and the method used (refers to [9] shown on Table 2. According to Table 2, it is known that linguistic-based methods most widely applied in the model on schema matching i.e. 76%, followed by a combination of methods as composite is 73%, auxiliary based i.e. 56%, followed by structure-based and instance-based i.e. 49%, constraint-based was 20%, and the least used was a combination of methods as hybrid i.e. 13% . Utilization of linguistic-based methods are relatively most rapidly, it is growing in line with science development and ontology increasingly widespread availability, nevertheless exploration methods based and instance-based structure is still being conducted by the researchers.

Based on type of input, OWL data model most widely explored, followed by RDF, and then the DTD/XSD. Based on application domain, schema matching models evolve towards more general in the last 10-15 years. It is supported by schema matching the need that increasingly needed in many applications, so that the developed model required to be able use for wider domain.

GUI feature support is also growing on the models and prototypes developed in the last 10-15 years. Of these conditions is possible because endorsement a programming language that makes it easy developing applications based on GUI. According to [16], hybrid matcher combining multiple methods simultaneously to carry out the schema matching process, and supposed the use of hybrid matcher would provide better results and able to provide improvements on performance (effectiveness) rather than individual matcher. On the other hand, the use of hybrid matching is rarely performed by researchers, has found 9 research in the last 25 years.

Table 2. Statistically research, models, and schema matching prototypes in the last 25 years

Period	Number of Research/ Publication	Input			Using Method						
		RDF	DTD/ XSD	OWL	Linguistic Based	Structure Based	Constraint Based	Instance Based	Method Combination		*Auxiliary Based* (Dictionary/ WordNet /Corpus)
									Hybrid	Composite	
1989-1994	5	1	0	0	0	0	5	1	0	1	0
1995-1999	7	4	1	1	4	2	0	1	0	4	4
2000-2004	42	18	12	14	20	11	9	16	5	22	17
2005-2009	13	5	4	8	10	7	0	6	1	9	7
2010-1014	4	4	9	16	20	15	0	11	3	16	12
Total:	**71**	**32**	**26**	**39**	**54**	**35**	**14**	**35**	**9**	**52**	**40**

4. FUTURE RESEARCH DIRECTION

Refers to classification schema matching in [9], hybrid matcher still open to develop on a combination of two or more methods of linguistic-based, structure-based, constraint-based, instance-based, and/or auxiliary based (the use of dictionary/WordNet/Corpus). This suggests a research opportunities on hybrid matcher.

Development of future solutions in the schema matching can also be developed to improve the convergence and resolution on the entity approach, such matching on metadata and instance level to identify the semantic relationships at the an entity or instance [9]. For additional information at the present authors currently developing a model and prototype Hybrid schema matching (name given by [16],[19]) or mixed strategy (according to classification by [23],[28]) by combining constraint-based and instance-based (in a classification by [9],[19]). A hybrid model schema matching developed involves DTM (data type matcher), CM (constraint matcher), and IDM (instance of data matcher) (classified by [18]).

Another problem associated with matching schema is the schema definition differences which caused by differences DBMS. Such a case can be resolved by using intermediary XML language, e.g. relational model database schema could be mapped into XML language using s-XML was developed by [86]. A framework mapping relational database model schema into XML language can be done in a two phases, i.e. mapping a relational schema models to UML class diagrams, and mapping UML class diagram to XML document [87].

5. CONCLUSION

Research on models and prototypes schema matching is still underway and still open to develop a hybrid model that involves two or more different methods which are processed simultaneous. Usage of hybrid matcher would provide a better result and improved performance (effectiveness) than using an individual matcher. New models and prototype can be developed to explore semantic relationships at the entities or instance as the base of mapping the relationship between schema elements. Recent research by the authors, is currently being developed a model and prototyping that combines constraint-based and instance-based, in it involves DTM (data type matcher), CM (constraint matcher), and IDM (instance of data matcher).

REFERENCES

[1] Cruz I. F., *et al.*, "Agreementmaker: Efficient matching for large real-world schemas and ontologies (demo paper)," *International Conference on Very Large Data Base (VLDB). Lyon, France,* vol/issue: 2(2), pp. 1586-1589, 2009. DOI: 10.14778/1687553.1687598.

[2] He B. and Chang K. C. C., "Statistical schema matching across web query interfaces," *The ACM SIGMOD International Conference Management of Data. San Diego, California, USA*, pp. 217-228, 2003. DOI: 10.1145/872757.872784.

[3] Do H. H., "Schema matching and mapping-based data integration," *Ph.D. Thesis,* Interdisciplinary Center for Bioinformatics and Department of Computer Science, University of Leipzig, Leipzig, Germany. URL: *lips.informatik.uni-leipzig.de/files/2006-4.pdf.*

[4] Kavitha C., *et al.*, "Ontology based semantic integration of heterogeneous databases," *European Journal of Scientific Research*, vol/issue: 64(1), pp. 115-122, 2011.

[5] Villanyi B., *et al.*, "A novel framework for the composition of schema matchers," *The 14th WSEAS International Conference on Computers, Latest Trends on Computers. Corfu Island, Greece*, pp. 379-384, 2010. URL: http://dl.acm.org/citation.cfm?id=1981573.1981641.

[6] Kim W. and Seo J., "Classifying schematic and data heterogeneity in multidatabase systems," *IEEE Journal*, vol/issue: vol/issue: 24(12), pp. 12-18, 1991. DOI: 10.1109/2.116884.

[7] Engmann D. and Massmann S., "Instance matching with COMA++," *Datenbank Systeme in Business, Technologie und Web (BTW Workshop): Model Management and Metadata. Aachen, Germany,* pp. 28-37, 2007. URL: http://ceur-ws.org/Vol-814/om2011_Tpaper5.pdf.

[8] Do H. H., *et al.*, "Comparison of schema matching evaluations," *The 2nd International Workshop Web and Databases. In: Lecture Notes In Computer Science (LNCS) 2593. Springer-Verlag, Germany,* pp. 221-237, 2003. URL: http://lips.informatik.uni-leipzig.de/files/2002-28.pdf.

[9] Bernstein P. A., Madhavan J., Rahm E., "Generic schema matching, ten years later," *Very Large Data Bases (VLDB) Endowment Journal*, Seattle, Washington, vol/issue: 4(11), pp. 695-701, 2011. URL: http://www.vldb.org/pvldb/vol4/p695-bernstein_madhavan_rahm.pdf.

[10] Li J., *et al.*, "RiMOM: A dynamic multistrategy ontology alignment framework," *Journal of IEEE Transaction Knowledge Data Engineering,* vol/issue: 21(8), pp. 1218-1232, 2009. DOI: 0.1109/TKDE.2008.202.

[11] Bernstein P., *et al.*, "The Microsoft repository," *The 23rd International Conference Very Large Databases (VLDB). Athens, Greece,* pp. 3-12, 1997. DOI: 10.1.1.50.8527.

[12] Bernstein P. A., "Applying model management to classical metadata problems," *The 1st International Conference Innovative Data Systems Research (CIDR). Asilomar, California, USA,* pp. 209-220, 2003. DOI: 10.1.1.12.2729.

[13] Stabenau A., *et al.*, "An overview of ensemble," *Genome Research Journal,* vol/issue: 14(5), pp. 929-933, 2004. DOI: 10.1101/gr.1857204.

[14] Madhavan J., *et al.*, "Generic schema matching with Cupid," *The 27th International Conference on Very Large Data Bases (VLDB). Roma, Italy,* pp. 49-58, 2001. URL: http://dl.acm.org/citation.cfm?id=645927.672191.

[15] Massmann S., *et al.,* "Evolution of the COMA match system," *The 6th International Workshop on Ontology Matching (OM-2011). Bonn, Germany,* pp. 49-60, 2011. URL: http://ceur-ws.org/Vol-814/om2011_Tpaper5.pdf.

[16] Rahm E. and Bernstein P. A., "A survey of approaches to automatic schema matching," *Very Large Databases (VLDB) Journal*, vol/issue: 10(4), pp. 334-350, 2001. DOI: 10.1007/s007780100057.

[17] Shvaiko P. and Euzenat J., "A survey of schema-based matching approaches," *Journal on Data Semantics,* vol. 4, pp. 146-171, 2005. DOI: 10.1007/11603412_5.

[18] Karasneh Y., *et al.*, "Integrating schemas of heterogeneous relational databases through schema matching," *The 11th International Conference on Information Integration and Web-based Applications and Services (iiWAS). Kuala Lumpur, Malaysia,* pp. 209-216, 2009. DOI: 10.1145/1806338.1806380.

[19] Özsu M. T. and Valduriez P. P., "Principles of distributed database systems," *Pearson Education, Inc., Springer, New York, USA,* 2011. DOI: 10.1007/978-1-4419-8834-8.

[20] Milo T. and Zohar S., "Using schema matching to simplify heterogeneous data translation," *The 24th International Conference on Very Large Data Bases (VLDB). New York, USA,* pp. 122-133, 1998. URL: http://www.vldb.org/conf/1998/p122.pdf.

[21] Melnik S., *et al.*, "RONDO: A programming platform for generic model management," *The ACM-SIGMOD Conference on Management of Data (SIGMOD). San Diego, California, USA,* pp. 193-204, 2003. DOI: 10.1145/872757.872782.

[22] Banek M., *et al.*, "Automated integration of heterogeneous data warehouse schemas," *International Journal of Data Warehousing and Mining (IJDWM),* vol/issue: 4(4), pp. 1-21, 2008. DOI: 10.4018/jdwm.2008100101.

[23] Gross A., *et al*., "On matching large life science ontologies in parallel," *The 7th International Conference Data Integration in the Life Sciences (DILS). Gothenburg, Sweden,* pp. 35-49, 2010. DOI: 10.1007/978-3-642-15120-0_4.

[24] Bergamaschi S., *et al.*, "Semantic integration of semistructured and structured data sources," *SIGMOD Record,* vol/issue: 28(1), pp. 54-59, 1999. DOI: 10.1145/309844.309897.

[25] Li W. S. and Clifton C., "Semint: A tool for identifying attribute correspondences in heterogeneous databases using neural network," *Data and Knowledge Engineering Journal,* vol/issue: 33(1), pp. 49-84, 2000. DOI: 10.1016/S0169-023X(99)00044-0.

[26] Doan A. H., *et al*., "Reconciling schemas of disparate data sources-a machine-learning approach," *The ACM SIGMOD International Conference Management of Data*. Santa Barbara, California, USA, pp. 509-520, 2001. DOI: 10.1145/376284.375731.

[27] Li W. S and Clifton C., "Semantic integration in heterogeneous databases using neural networks," *The 20th International Conference on Very Large Data Bases (VLDB). Santiago de Chile, Chile,* pp. 1-12, 1994. URL: https://www.cerias.purdue.edu/assets/pdf/bibtex_archive/2001-86-report.pdf.

[28] Rahm E., "Towards large-scale schema and ontology matching," in Bellahsene Z, Bonifati A, Rahm E., *Schema matching and mapping, data-centric systems and applications*. Springer. New York, USA, pp. 3-28, 2011. DOI: 10.1007/978-3-642-16518-4_1.

[29] Bellahsene Z., *et al.*, "Schema matching and mapping, data-centric systems and applications," *Springer,* New York, USA, 2011. DOI: 10.1007/978-3-642-16518-4.

[30] Barrasa J., *et al.*, "R2O, An extensible and semantically based database to ontology mapping language," *The 2nd Workshop on Semantic Web and Databases (SWDB2004). Toronto, Canada,* pp. 92-119, 2004. URL: http://www.cs.man.ac.uk/~ocorcho/documents/SWDB2004_BarrasaEtAl.pdf.

[31] Evermann J., "An exploratory study of database integration processes," *IEEE Transactions on Knowledge and Data Engineering Journal*, vol/issue: 20(1), pp. 99-115, 2008. DOI: 10.1109/TKDE.2007.190675.

[32] Bizer C., *et al.*, "Linked data-the story so far," *International Journal of Semantic Web Informations Systems*, vol/issue: 5(3), pp. 1-22, 2009. DOI: 10.4018/jswis.2009081901.

[33] Parundekar R., *et al.*, "Linking and building ontologies of linked data," *The 9th International Semantic Web Conference (ISWC). Shanghai, China,* pp. 598-614, 2010. URL: http://www.isi.edu/integration/papers/parundekar10-iswc.pdf.

[34] Falconer S. M. and Noy N. F., "Interactive techniques to support ontology matching," in Bellahsene Z, Bonifati A Rahm E., *Schema matching and mapping, data-centric systems and applications*, Springer. New York, USA, pp. 29-52, 2011. DOI: 10.1007/978-3-642-16518-4_2.

[35] Aumueller D., *et al.*, "Schema and ontology matching with COMA++," *The SIGMOD (demo paper)*, Baltimore, Maryland, USA, pp. 906-908, 2005. DOI: 10.1145/1066157.1066283.

[36] Seligman L., *et al.*, "OpenII: An open source information integration toolkit," *The ACM SIGMOD International Conference on Management of Data. Indianapolis, Indiana, USA,* pp. 1057-1060, 2010. DOI: 10.1145/1807167.1807285.

[37] Euzenat J., *et al.*, "Results of the ontology alignment evaluation initiative 2010," *The 5th ISWC Workshop on Ontology Matching (OM). Shanghai, China,* 2010. URL: http://disi.unitn.it/~p2p/OM-2010/oaei10_paper0.pdf.

[38] Hernández M. A., *et al.*, "CLIO: A semi-automatic tool for schema mapping (software demonstration)," *The ACM SIGMOD International Conference Management of Data. Santa Barbara, California, USA,* pp. 607, 2001. DOI: 10.1145/376284.375767.

[39] Naumann F., *et al.*, "Attribute classification using feature analysis," *Poster. The 18th International Conference on Data Engineering (ICDE). San Jose, California, USA,* pp. 271, 2002. URL: www.hpi.uni-potsdam.de/fileadmin/hpi/FG_Naumann/publications/ICDE02Poster.pdf.

[40] Popa L., *et al.*, "Mapping XML and relational schemas with CLIO (software demonstration)," *The International Conference on Data Enginnering (ICDE)*. San Jose, California, USA, pp. 498-499, 2002. URL: http://disi.unitn.it/~velgias/docs/PopaHVMNH02.pdf.

[41] Haas L. M., *et al.*, "CLIO grows up: from research prototype to industrial tool," *The ACM SIGMOD International Conference Management of Data. Baltimore, Maryland, USA,* pp. 805-810, 2005. DOI: 10.1145/1066157.1066252.

[42] Kang J. and Naughton J., "On schema matching with opaque column names and data values," *The ACM SIGMOD International Conference Management of Data. San Diego, California, USA*, pp. 205-216, 2003. DOI: 10.1145/872757.872783.

[43] Chien B. C. and He S. Y., "A hybrid approach for automatic schema matching," *The 9th International Conference on Machine Learning and Cybernetics. Qingdao, China,* pp. 2881-2886, 2010. DOI: 10.1109/ICMLC.2010.5580776.

[44] Li W. S., *et al.*, "Database integration using neural networks: implementation and experiences," *Knowledge and Information Systems Journal*, vol/issue: 2(1), pp. 73-96, 2000. DOI: 10.1007/s101150050004.

[45] Do H. H. and Rahm E., "COMA: A system for flexible combination of schema matching approach," *The 28th Conference on Very Large Data Bases (VLDB). Hong Kong, China,* pp. 610-621, 2002. URL: http://dbs.uni-leipzig.de/file/COMA.pdf.

[46] Madhavan J., *et al.*, "Corpus-Based Schema Matching," *The IJCAI-03 Workshop on Information Integration on the Web (IIWeb)*. Acapulco, Mexico, pp. 59-63, 2003. DOI: 10.1109/ICDE.2005.39.

[47] Dhamankar R., *et al.*, "IMAP: discovering complex semantic matches between database schemas," *The ACM SIGMOD International Conference Management of Data*. Paris, France, pp. 383-394, 2004. DOI: 10.1145/1007568.1007612.

[48] Bernstein P. A., *et al.*, "Industrial-strength schema matching," *ACM SIGMOD Record*, vol/issue: 33(4), pp. 38-53, 2004. DOI: 10.1145/1041410.1041417.

[49] Dragut E. and Lawrence R., "Composing mappings between schemas using a reference ontology," *The International Conference on Ontologies, Databases, and Applications of Semantics (ODBASE). Larnaca, Cyprus,* pp. 783-800, 2004. DOI: 10.1007/978-3-540-30468-5_50.

[50] Mork P. and Bernstein P. A., "Adapting a generic match algorithm to align ontologies of human anatomy," *The 20th International Conference on Data Engineering (ICDE). Boston, Massachusetts, USA,* pp. 787-790, 2004. DOI: 10.1109/ICDE.2004.1320047.

[51] Tu K. W. and Yu Y., "CMC: combining multiple schema-matching strategies based on credibility prediction," *The 10th International Conference on Database Systems for Advanced Applications (DASFAA). Beijing, China,* pp. 888-893, 2005. DOI: 10.1007/11408079_80.

[52] Jian N., *et al.*, "Falcon-AO: Aligning ontologies with Falcon," *The K-CAP Workshop on Integrating Ontologies (K-CAP'05). Banff, Canada,* pp. 85-91, 2005. DOI: 10.1016/j.websem.2008.02.006.

[53] Jean-Mary Y. R., *et al.*, "Ontology matching with semantic verification," *Web Semantics Journal,* vol/issue: 7(3), pp. 235-251, 2009. DOI: 10.1016/j.websem.2009.04.001.

[54] Palopoli L., *et al.*, "Semi-automatic, semantic discovery of properties from database schemes," *The International Database Engineering and Applications Symposium (IDEAS)*. Cardiff, Wales, United Kingdom, pp. 244-253, 1998. DOI: 10.1109/IDEAS.1998.694384.

[55] Palopoli L., *et al.*, "A unified graph-based framework for deriving nominal interscheme properties, type conflicts and object cluster similarities," *The 4th IFCIS International Conference on Cooperative Information Systems (CoopIS).Edinburgh, Scotland,* pp. 34-35, 1999. DOI: 10.1109/COOPIS.1999.792152.

[56] Palopoli L., *et al.*, "The system Dike: Towards the semi-automatic synthesis of cooperative information systems and data warehouses," *The Chalenges: 2000 ADBIS-DASFAA Symposium on Advances in Databases and Information Systems, Enlarged 4th East-European Conference on Advances in Databases and Information Systems. Prague, Czech Republic,* pp. 108-117, 2000. URL: http://dblp.uni-trier.de/rec/bib/conf/adbis/PalopoliTU00.

[57] Castano S. and Antonellis V. D., "A schema analysis and reconciliation tool environment for heterogeneous databases," *The International Database Engineering and Applications Symposium (IDEAS). Montreal, Que,* pp. 53-62, 1999. DOI: 10.1109/IDEAS.1999.787251.

[58] Bergamaschi S., *et al.*, "Semantic integration of heterogeneous information sources," *Data and Knowledge Engineering Journal,* vol/issue: 36(3), pp. 215-249, 2001. DOI: 10.1016/S0169-023X(00)00047-1.

[59] Mitra P., *et al.*, "Semi-automatic integration of knowledge sources," *The 2nd International Conference on Information Fusion (FUSION). Sunnyvale, California, USA,* 1999. URL: https://www.researchgate.net/publication/2630475_ Semi-automatic_Integration_of_Knowledge_Sources.

[60] Mitra P., *et al.*, "Graph-oriented model for articulation of ontology interdependencies," *The 7th International Conference Extending Database Technology (EDBT). Konstanz, Germany,* pp. 86-100, 2000. DOI: 10.1007/3-540-46439-5_6.

[61] Mitra P. and Wiederhold G., "Resolving terminological heterogeneity in ontologies," *The ECAI-02 Workshop on Ontologies and Semantic Interoperability, European Conference on Artificial Intelligence (ECAI). Lyon, France,* pp. 45-50, 2002. DOI: 10.1145/505168.505196.

[62] Giunchiglia F., *et al.*, "A large scale taxonomy mapping evaluation," *The 4th International Conference Semantic Web Conference (ISWC). Galway, Ireland,* pp. 67-81, 2005. DOI: 10.1007/11574620_8.

[63] Palopoli L., *et al.*, "Uniform techniques for deriving similarities of objects and subschemes in heterogeneous databases," *IEEE Transaction Knowledge and Data Engineering,* vol/issue: 15(2), pp. 271-294, 2003. DOI: 10.1109/TKDE.2003.1185834.

[64] He H., *et al.*, "Wise-Integrator-An automatic integrator of web search interfaces for e-commerce," *The 29th International Conference on Very Large Data Bases (VLDB). Berlin, Germany,* pp. 357-368, 2003. URL: www.vldb.org/conf/2003/papers/S12P01.pdf.

[65] He B., *et al.*, "Discovering complex matchings across web query interfaces-a correlation mining approach," *The 10th ACM SIGKDD International Conference Knowledge Discovery and Data Mining. Seattle, Washington, USA,* pp. 148-157, 2004. *DOI:* 10.1145/1014052.1014071.

[66] Noy N. F. and Musen M. A., "The Prompt Suite: Interactive tools for ontology merging and mapping," *International Journal of Human-Computer Studies*, vol/issue: 59(6), pp. 983-1024, 2003. DOI: 10.1016/j.ijhcs.2003.08.002.

[67] Noy N. F. and Musen M. A., "Using prompt ontology-comparison tools in the EON ontology alignment contest," *The 3rd International Workshop Evaluation of Ontology-Based Tools (EON). Hiroshima, Japan,* pp. 79-90, 2004. DOI: 10.1.1.91.1763.

[68] Euzenat J., *et al.*, "Ontology alignment with OLA," *The 3rd International Workshop Evaluation of Ontology Based Tools (EON). Hiroshima, Japan,* pp. 56-88, 2004. URL: ceur-ws.org/Vol-128/EON2004_EXP_Euzenat.pdf

[69] Ehrig M. and Staab S., "QOM-quick ontology mapping," *The 3rd International Semantic Web Conference (ISWC). 2004, Hiroshima, Japan,* pp. 683-697, 2004. DOI: 10.1007/978-3-540-30475-3_47.

[70] Ehrig M. and Sure Y., "Ontology alignment-Karlsruhe," *The 3rd International Workshop Evaluation of Ontology-Based Tools (EON) Hiroshima, Japan,* pp. 48-55, 2004. URL: sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-128/EON2004_EXP_Ehrig.pdf.

[71] Giunchiglia F., *et al.*, "S-Match: an algorithm and an implementation of semantic matching," *The 1st European Semantic Web Symposium (ESWS). Heraklion, Crete, Greece,* pp. 61-75, 2004. DOI: 10.1007/978-3-540-25956-5_5.

[72] Giunchiglia F., *et al.*, "Efficient semantic matching," *The 2nd European Semantic Web Conference (ESWC). Heraklion, Crete, Greece,* pp. 272-289, 2005. DOI: 10.1007/11431053_19.

[73] Benkley S., *et al.*, "Data Element Tool-Based Analysis (DELTA)," *MITRE Technical Report MTR'95 B147*, 1995.

[74] Clifton C., *et al.*, "Experience with a combined approach to attribute-matching across heterogeneous databases," *The 7th IFIP 2.6 Working Conf. on Database Semantics. Leysin, Switzerland,* pp. 428-451, 1997. DOI: 10.1007/978-0-387-35300-5_18.

[75] Melnik S., *et al.*, "Similarity Flooding: A versatile graph matching algorithm," *The 18th International Conference on Data Engineering (ICDE). San Jose, California, USA,* pp. 117-128, 2002. DOI: 10.1109/ICDE.2002.994702.

[76] Lee M. L., *et al.*, "XClust: Clustering XML schemas for effective integration," *The 11th International Conference on Information and Knowledge Management (CIKM'02). Virginia, USA,* pp. 292-299, 2002. DOI: 10.1.1.62.1532.

[77] Lu J., *et al.*, "An experiment on the matching and reuse of XML schemas," *The 5th International Conference on Web Engineering (ICWE). Sydney, Australia,* pp. 273-284, 2005. DOI: 10.1007/11531371_38.

[78] Berlin J. and Motro A., “Automatch: database schema matching using machine learning with feature selection,” *The 6th International Conference Cooperative Information Systems (CoopIS). In Cooperation with VLDB 2001. Trento, Italy*, pp. 108-122, 2001. URL: http://cs.gmu.edu/~ami/teaching/infs797/current/autoplex-derivatives.pdf.

[79] Berlin J. and Motro A., “Autoplex: Automated discovery of content for virtual databases,” *The 9th International Conference Cooperative Information Systems (CoopIS). In Cooperation with VLDB 2001. Trento, Italy*, pp. 108-122, 2001. DOI: 10.1007/3-540-44751-2_10.

[80] Embley D. W., *et al.*, “Multifaceted exploitation of metadata for attribute match discovery in information integration,” *The 1st International Workshop Information Integration on the Web (WIIW). Rio deJaneiro, Brazil*, pp. 110-117, 2001. URL: http://disi.unitn.it/~accord/RelatedWork/Matching/15-David_Embley(29).pdf.

[81] Xu L. and Embley D., “Discovering direct and indirect matches for schema elements,” *The 8th International Conference on Database Systems for Advanced Applications (DASFAA). Kyoto, Japan*, pp. 39-46, 2003. DOI: 10.1109/DASFAA.2003.1192366.

[82] Doan A. H., *et al.*, “Learning to map between ontologies on the semantic web,” *The 11th International Conference on World Wide Web (WWW). Honolulu, Hawaii, USA*, pp. 662-673, 2002. DOI: 10.1145/511446.511532.

[83] Wang J., *et al.*, “Instance-based schema matching for web databases by domain-specific query probing,” *The 13th International Conference on Very Large Data-Bases (VLDB). Toronto, Canada*, pp. 408-419, 2004. URL: http://www.isys.ucl.ac.be/vldb04/eProceedings/contents/pdf/RS10P4.PDF.

[84] Hoshiai T., *et al.*, “A semantic category matching approach to ontology alignment,” *The 3rd International Workshop Evaluation of Ontology-Based Tools (EON). Hiroshima, Japan*, pp. 67-78, 2004. URL: http://ceur-ws.org/Vol-128/EON2004_EXP_Hoshiai.pdf.

[85] Bilke A. and Naumann F., “Schema matching using duplicates,” *The 21th International Conference on Data Engineering (ICDE). Tokyo, Japan*, pp. 69-80, 2005. DOI: 10.1109/ICDE.2005.126.

[86] Samini S., *et al.*, “Bridging XML and relational databases: An effective mapping scheme based on persistent,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol/issue: 2(2), pp. 239-246, 2012. DOI: 10.11591/ijece.v2i2.215.

[87] Win L. H., “XML-based RDF data management for XPath query language,” *International Journal of Informatics and Communication Technology (IJ-ICT)*, vol/issue: 2(1), pp. 1-8, 2013. DOI: 0.11591/ij-ict.v2i1.1503.

BIOGRAPHIES OF AUTHORS

Edhy Sutanta. Currently is pursuing his doctoral program in computer science at Department of Computer Sciences & Electronics Instrumentations, Faculty of Mathematics & Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia. He took his undergraduate (S.T.) in from IST AKPRIND Yogyakarta Indonesia in 1996, and Master (M.Kom.) in Computer Science from Universitas Gadjah Mada, Yogyakarta, Indonesia in 2006. His research areas of interest are database concept, database analysis and design, information systems, information systems integration, and e-goverment.

Email : edhy_sst@akprind.ac.id, edhy_sst@yahoo.com

Retantyo Wardoyo. He had his undergraduate (Drs) in Mathematics from Universitas Gadjah Mada, Yogyakarta, Indonesia in 1982, and Master (MSc.) in Computer Science from the University of Manchester, UK in 1990. He had his Doctor (PhD.) in Computation from University of Manchester Institute of Science and Technology, UK in 1996. His research areas of interest are data base systems, operating systems, management information systems, fuzzy logics, and software engineering.

Email : rw@ugm.ac.id

Dr.Techn. Khabib Mustofa, S.Si., M.Kom. He had his undergraduated (S.Si) in Computer Science from Universitas Gadjah Mada, Yogyakarta, Indonesia in Pebruari 1997. In May 2001 he was granted a degree of Master level by Universitas Gadjah Mada, Yogyakarta, Indonesia. And he presented his dissertation and was declared to pass the exams Ph.D. on 24 April 2007 in Computer Science from Institute for Software Engineering and Interactive Systems, Technische Universit at Wien (Vienna University of Technology), Vienna, Austria. His research area of interest includes database system**,** web technology, semantic web, software engineering and information management.

Email: khabib@ugm.ac.id

Drs. Edi Winarko, M.Sc., Ph.D.
He took his undergraduated (Drs.) Degree in Statistics from Universitas Gadjah Mada, Yogyakarta, Indonesia at 1986, and Master (MSc.) in Computer Sciences from Queen's University, Canada at 2002, and Ph.D. in Computer Sciences from Computer Science School of Informatics and Engineering Flinders University, Australia at 2007. Her research areas of interest are data warehousing and data mining, and information retrieval.

Email: edwin@ugm.ac.id