International Journal of Applied Science and Engineering

On hybrid schema matching modified model in minimizing user verification process in output validation

Edhy Sutanta[1*], Erna Kumalasari Nurnawati[1], Catur Iswahyudi[1], Rosalia Arum Kumalasanti[2]

[1] *Department of Informatics, Institut Sains & Teknologi AKPRIND Yogyakarta 28 Kalisahak Road, Yogyakarta, 55222, Indonesia*

[2] *Department of Informatics, Sanata Dharma University Yogyakarta, Paingan Road, Sleman, Yogyakarta, 55282, Indonesia*

ABSTRACT

The problem of alignment of the scheme is that the process that cannot be thoroughly performed automatically is the similarity of the results of the element output map scheme alignment that the user still needs to correct to obtain a valid end result. The correction process which is carried out at the stage of verification and evaluation can only be done manually by the user. In this research, a modification of the hybrid schema matching model was proposed that previously develop by adding three new features, namely the use of a similarity value limit (SVL), checking the inter-attribute similarity of the input database, and selecting the appropriate database to act as a DBSource during the matching process. Every new feature is tested using a relational database model (RDBM). Compare the yields of the original model and the modified model to determine the reduction in output of the user-performed model validation process. The test result shows that the addition of new features succeeded in minimizing the user verification process.

Keywords: Hybrid schema matching, Modified model, User verification.

OPEN ACCESS

Received: February 7, 2022
Revised: September 26, 2022
Accepted: November 21, 2022

Corresponding Author:
Edhy Sutanta
edhy_sst@akprind.ac.id

Publisher:
Chaoyang University of Technology
ISSN: 1727-2394 (Print)
ISSN: 1727-7841 (Online)

1. INTRODUCTION

Schema matching has been defined by experts in different ways but has a similar meaning. Generally, schema matching is a process of finding the similarity relationships between elements of a schema pair. Schema matching is the same job as matching (Li et al., 2010), or as a process for finding the relationship between elements of the pair schema (Kim and Seo, 1991; Bernstein et al., 1997; Bernstein, 2003). Schema matching models can be developed using one or a combination of methods (Kim and Seo, 1991; Madhavan, 2001; Rahm and Bernstein, 2001). According to Rahm and Bernstein (2001) and Özsu and Valduriez (2011), the combination of methods can be implemented as a composite or hybrid. A composite matcher runs algorithms separately and combines the results (Lee et al., 2009), while hybrid models use multiple criteria simultaneously (Li et al., 2010; Bergamaschi et al., 1999; Li and Clifton, 2000). Refers to the survey in the schema matching model by Sutanta et al. (2016a), research by Sutanta et al. (2016b) has developed a schema matching model by combining two methods, namely, constraints and instances that are run simultaneously or called a hybrid model. The mathematical model for the hybrid schema matching is found in Sutanta et al. (2019). The hybrid schema matching model has been modified to improve the effectiveness of the model outputs in Sutanta et al. (2021).

Referring to Sutanta et al. (2016b), the hybrid schema matching model consists of 4 steps, namely input, process, output, verification and evaluation. One of the problems in a schema matching is that a process cannot be thoroughly carried out automatically.

For small to medium sized enterprises matching schemas is still a time consuming manual task (Schmidts et al., 2019). Even expensive commercial solutions perform poorly, if the context is not suitable for the product (Bernstein et al., 1997; Schmidts et al., 2019). The similarity between the element output mapping schema matching results still needs to be corrected by the user to obtain a valid final result (Milo and Zohar, 1998; Massmann et al., 2011). The correction process can only be done manually by the user (Melnik et al., 2003; Banek et al., 2008; Kavitha et al., 2011). In Sutanta et al. (2016b), the correction process is carried out at the stage of verification and evaluation. In a schema matching that involves a pair of databases with numerous elements, this verification process will be very tedious, because the user must repeatedly verify a vast amount. It requires an effort to minimize the verification process on the output schema matching.

This paper proposes a model for reducing the user verification process. The proposed model is based on a modified hybrid schema matching. The model is a modification of our previous results presented in Sutanta et al. (2016b), Sutanta et al. (2019), and Sutanta et al. (2021). In Sutanta et al. (2016b), we propose a new model to run the schema matching process by combining two methods, which are constraint-based and instance-based. Those two combined methods are run simultaneously and meet the hybrid model criteria so that we named it hybrid model schema matching. Hybrid model schema matching shows an excellent result with precision (P) = 90.00%, Recall (R) = 80.00, and F-Measure (F) = 84.00%. The mathematical model for the hybrid model schema matching is presented in detail in Sutanta et al. (2019). The hybrid model schema matching in Sutanta et al. (2016b) needs to be improved in terms of the effectiveness and efficiency. The effort to increase the model effectiveness has been conducted through a new additional feature, which is uses a variation in weighting criteria and string size matching, which has been presented in Sutanta et al. (2021). This effort was successfully conducted in increasing the model effectiveness. Hence, values of P = 97.66%, R = 99.83%, and F = 98.74% are obtained. Referring to the research results, this research specifically reviews the efforts to increase the efficiency in hybrid model schema matching.

Unlike previous methods, the novelty of this work is the use of similar value limits (SVL), checking the input database for similarity between attributes, and selecting an appropriate database to serve as the DBSource during execution.

The rest of this paper is organized as follows. Section 2 presents the material used and our proposed modified hybrid schema matching. Section 3 presents the obtained results and following by discussion. Finally, Section 4 concludes this work.

2. MATERIALS AND METHODS

2.1 Hybrid Schema Matching Model

Schema matching requires pairs of input databases, one called DBSource as the database to be matched and the other as DBTarget, which serves as the reference when matching. Furthermore, the model will be a matching process and will determine the output of a similar mapping attribute pair using the hybrid schema matching model. Fig. 1 shows the original hybrid model schema matching in Sutanta et al. (2016b), which consists of four parts, namely:

1. Input: to take obtain from DBSource and DBTarget, DBMS types, extraction constraints, data type conversions, instance extractions, and similarity checks between attributes of DBSource and DBTarget.
2. Process: to handles and runs the matching process of each attribute on the DBSource with each attribute of the DBTarget, then computes a similarity score (SIM_{MN}) for each possible pair of matching attributes, and determines which pair of attributes to declare as a match.
3. Output: to display attribute pairs that map similarity to attribute pairs, that is, attribute pairs with $SIM_{MN}MAX$ and $SIM_{MN} = 1$, namely, preliminary results.
4. Verification and Evaluation. Verification is the process of determining whether the preliminary results generated by the model are correct or still need to be manually edited by the user. Thus, the process is a supervised approach. User-verified preliminary results produce verified results as a pair of valid attributes. The evaluation process is performed to calculate the value of the performance parameters of the model, that is, P (Precision), R (Recall) and F (F-measure). The values of P, R, and F are calculated by comparing preliminary and verified results.

2.2 Modified Hybrid Schema Matching Model

In this research, the hybrid schema model of Sutanta et al. (2016b) was modified by adding three new features. There is the use of the SVL on the attribute pair similarity value (SIM), checking the inter-attribute similarity on the DBSource and DBTarget, and selecting the appropriate database as the DBsource during the matching process. The modified hybrid schema matching model displayed in Fig. 2.

2.3 Dataset for Model Testing

Tests were performed on 32 database pairs using a dataset of 30 relational database models as used in Sutanta et al. (2016b). The test for each pair of databases was run 12 times using a combination of 3 variants of match criteria weight and 4 variants of matching string size as performed by Sutanta et al. (2016b), so the number of tests performed was 384 times. The following description shows DBSource and DBTarget used as test data. Based on DBMS, it includes 8 databases developed using MS Access and 22 databases developed using MySQL. Based on the application domain, it consists of 8 university academic applications, 12 high school academic applications, 8 e-government applications and 2 e-commerce applications.

The largest test database consists of 204 tables, while the largest number of attributes is 1,851, the largest number of data items is 232,893 entries, and the largest capacity of the database is 79,769 Kilo Bytes (KB). The smallest test database contains 1 table, the smallest number of attributes is 16, the smallest number of data items is 480 items, and the smallest size is 115 KB. The test results obtained with the original model are compared with the results of the original model. The model was revised to determine if there was any degradation in the new model validation process.

3. RESULTS AND DISCUSSION

3.1 Modified 1: Using A Similarity Value Limit (SVL) to Determine the Attribute Pair

The first modification of the hybrid schema-matching model performed was to use SVL to determine each pair of attributes that the model declared appropriate. This method has been applied to Cupid software (Kim and Seo, 1991). The modification is done by adding a step to find the attribute of the smallest SIM value pair declared good by the model and validated by the user. This value is then set as the SVL value. Then the entire attribute pair containing the SIM < SVL declares the SVL value to be unsuitable. The user validation process is adjusted only if it has to rely on attribute pairs with SIM ≥ SVL values. Table 1 displays the results of a comparative test of how many pairs of the attributes verification process as a whole, manually, and automatically on all tests using SVL = 0.75. We tested the hybrid schema matching model 384 times for 32 pairs of DBSource and DBTarget combining 3 variations of match criteria weights and 4 variations of string size matches, according to the previous description in Research Methods. Based on the test result, the smallest SIM value of the attribute pair, which is declared suitable by the model and declared valid by the user, is 0.76. We use this result as the basis for determining the accuracy limit value in this study, which is 0.75. Furthermore, all attribute pairs with SIM values < 0.75 are declared unsuitable by the model, attribute pairs with SIM values 0.75 is declared matched, and the verification process is only required for attribute pairs that are declared matched. From Table 1, the verification process of the model output required as a whole is 37,246,452 times. Compared with Sutanta et al. (2016b), by using SVL = 0.75, most verification processes can be carried out automatically, which is 36,378,155 or 97.67%. Users only should verify 868,297 attribute pairs. These results demonstrate that adding new functionality, or using SVL, can reduce the scope of the user authentication process.

Where the number of iterations of the manual validation process for attribute pairs is denoted by cvm, caDBs denotes the attribute count of the DBSource and caDBt denotes the attribute count of the DBTarget and the attribute pairs declared as good and accepted by the model. shows the count of if the user is represented as ctp, count the attribute pair repetitions in the manual validation process should be calculated by Equation (1).

$$\text{cvm} = (\text{caDBs} \times \text{caDBt}) - \text{ctp} \quad (1)$$

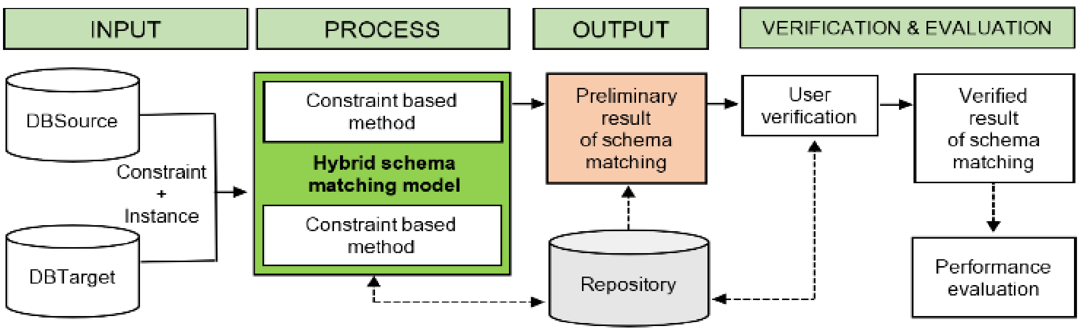

Fig. 1. Original hybrid schema matching model (Sutanta et al., 2016b)

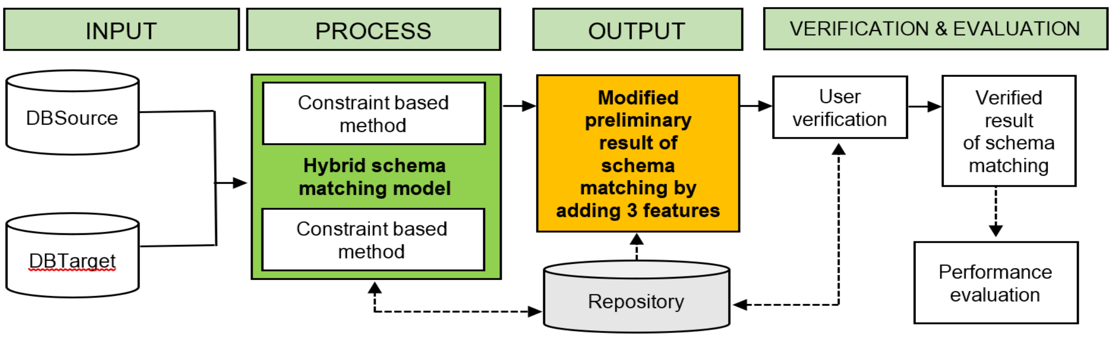

Fig. 2. Modified hybrid schema matching model

As a result of implementing SVL, the model should provide the user with a mechanism for adding attribute pairs that are incompatible with the model but matched by the user. The addition of an automatic validation mechanism is aimed at minimizing user interaction when validating model output. This modification does not affect the P, R, and F values in Sutanta et al. (2019).

3.2 Modified 2: Adding Checking the Inter-Attribute Similarity of The Database Input

Referring to the matching mechanism and calculating the SIM value of the attribute pair in the original hybrid schema matching model in Sutanta et al. (2016b), every possible attribute pair will be matched and the similarity value calculated, then verified by the user. The total number of matching processing steps and calculations performed on SIM attribute pairs can be calculated by Equation (2), where chn is the number of matching and similar pair value attribute calculation steps.

$$chn = caDBs \times caDBt \quad (2)$$

The second modification to the schema matching hybrid model is the addition of the checking process of inter-attribute similarity in the database, both in DBSource and DBTarget. The modification is done by adding a procedure to find attributes with the same definition and declaring them as the same attribute. Such attributes can be found in foreign keys (FK) in the database. Comparison and calculation of SIM values are performed only once for the same attribute.

Table 2 shows the test results for comparing the number of fitting steps of the original and modified models and calculating the SIM values. Based on Table 2 compared by Sutanta et al. (2016b), adding a function to check similarity between attributes in DBSource and DBTarget reduced SIM value matches and calculations from 37,246,452 to 35,997,940. Or there are 1,248,512 (= 3.35%) unnecessary attribute pairs and SIM value calculations.

Table 1. Comparison of model output overall needs verification and manual and automatic

DBSource	DBTarget	All Verification Needed	Manual Verification	Automatic Verification	
				Count	%
admission	admision	480,000	41,774	438,226	91.30
admission	academic	1,082,400	12	1,082,388	100.00
academic	payroll	524,964	1,856	523,108	99.65
academic	employ	1,076,988	7,032	1,069,956	99.35
academic	taxpph	308,484	4,186	304,298	98.64
academic	workshop	86,592	4,429	82,163	94.89
academic	library	2,305,512	588	2,304,924	99.97
academic	user	373,428	1,384	372,044	99.63
dptkp	lisence	12,276	796	11,480	93.52
dptkp	lisenceoln	26,928	2,713	24,215	89.92
dptkp	dptbgcpt	7,524	584	6,940	92.24
dptkp	quickcountbgcpt	23,364	1,482	21,882	93.66
dptkp	dptbtl	13,068	656	12,412	94.98
dptkp	dptkp	13,068	488	12,580	96.27
dptkdy	rsmitra	26,136	195	25,941	99.25
dptkdy	motorcred	7,524	748	6,776	90.06
nuptk	nuptk	4,523,952	313,778	4,210,174	93.06
nuptk	sinisa	523,128	41,253	481,875	92.11
nuptk	sipp	1,112,568	17,269	1,095,299	98.45
nuptk	psb	13,638,168	125,689	13,512,479	99.08
sipp	sinisa	128,652	10,060	118,592	92.18
sipp	sipp	273,612	236	273,376	99.91
sipp	psb	3,354,012	81,372	3,272,640	97.57
sipp	grade	344,280	11,512	332,768	96.66
sipp	gradeol	48,924	4,302	44,622	91.21
sipp	report	563,532	16,828	546,704	97.01
sipp	hspwt	2,975,304	60,527	2,914,777	97.97
sipp	forum	41,676	1,435	40,241	96.56
sipp	announcement	30,804	1,801	29,003	94.15
sipp	webinfo	576,216	38,390	537,826	93.34
sipp	osis	164,892	7,178	157,714	95.65
sipp	elearning	2,578,476	67,744	2,510,732	97.37
Sum or Average:		37,246,452	868,297	36,378,155	97.67

Table 2. Comparison of matching & calculation of SIM with and without involving the same attribute

DBSource			DBTarget			Number of Matching & SIM Calculation Steps		Step Reduction	
Database Name	Number of Attribute	Same Attribute	Database Name	Number of Attribute	Number of Attribute	Origin Model	Modified Model	Count	%
admision	200	73	admision	200	73	480,000	463,871	16,129	3.36
admision	200	73	academic	451	191	1,082,400	1,049,380	33,020	3.05
academic	451	191	payroll	97	36	524,964	509,104	15,860	3.02
academic	451	191	employ	199	75	1,076,988	1,044,748	32,240	2.99
academic	451	191	taxpph	57	11	308,484	296,524	11,960	3.88
academic	451	191	workshop	16	1	86,592	82,692	3,900	4.50
academic	451	191	library	426	180	2,305,512	2,241,552	63,960	2.77
academic	451	191	user	69	16	373,428	359,648	13,780	3.69
dptkp	33	0	dptkp	33	0	13,068	11,979	1,089	8.33
dptkp	33	0	dptbgcpt	19	3	7,524	6,996	528	7.02
dptkp	33	0	quickcountbgcpt	59	28	23,364	22,341	1,023	4.38
dptkp	33	0	lisence	31	0	12,276	11,253	1,023	8.33
dptkp	33	0	lisenceol	68	11	26,928	25,047	1,881	6.99
dptkp	33	0	dptbtl	33	0	13,068	11,979	1,089	8.33
dptkdy	33	0	rsmitra	66	15	26,136	24,453	1,683	6.44
dptkdy	33	0	motorcred	19	0	7,524	6,897	627	8.33
nuptk	614	95	sinisa	71	8	523,128	490,431	32,697	6.25
nuptk	614	95	sipp	151	38	1,112,568	1,053,921	58,647	5.27
nuptk	614	95	nuptk	614	95	4,523,952	4,254,591	269,361	5.95
nuptk	614	95	psb	1,851	1,064	13,638,168	13,229,715	408,453	2.99
sipp	151	38	sipp	151	38	273,612	260,843	12,769	4.67
sipp	151	38	gradeol	27	3	48,924	46,212	2,712	5.54
sipp	151	38	announcement	17	4	30,804	29,335	1,469	4.77
sipp	151	38	forum	23	6	41,676	39,755	1,921	4.61
sipp	151	38	sinisa	71	8	128,652	121,533	7,119	5.53
sipp	151	38	osis	91	10	164,892	155,739	9,153	5.55
sipp	151	38	webinfo	318	157	576,216	558,023	18,193	3.16
sipp	151	38	sma2pwt	1,642	1,311	2,975,304	2,937,901	37,403	1.26
sipp	151	38	grade	190	98	344,280	333,884	10,396	3.02
sipp	151	38	report	311	154	563,532	545,791	17,741	3.15
sipp	151	38	psb	1,851	1,064	3,354,012	3,265,081	88,931	2.65
sipp	151	38	elearning	1.423	788	2,578,476	2,506,721	71,755	2.78
Sum or Average:						37,246,452	35,997,940	1,248,512	96.65

These results show that adding a similarity check function between attributes can reduce the adjustment step and SIM value calculation. In general, the higher the number of attributes that are the same in both the DBSource and DBTarget, the more the SIM value attribute pair matching and calculation steps are reduced. The counting step of matching and calculating SIM values without checking features across common attributes, denoted by chtia, where chtia is the counting of a computational process involving checking features across common attributes, the same attributes are counted in the CASDBt state. The state of DBTarget and caSDBs count the same attributes in DBSource. Then the number of iteration steps to match and compute the original model and the modified SIM values can be calculated using Equations (3) and (4) as follows:

chtia = (caDBs) × (caDBt) (3)

chia = (caDBs – caSDBs) × (caDBt – caSDBt) (4)

If the matching process performed by the step calculation model and the reduction of SIM value attribute pairs are represented by Deltasim, it can be calculated by Equation (5).

deltachsim = chtia – chia (5)

3.3 Modified 3: Adding Selecting the Appropriate Database to Act as DBSource

The third change in hybrid model schema matching adds new functions for selecting and determining the appropriate database as the DBTarget. Modified by adding a step that finds a relatively small table that has a count attribute and matches the current SIM and calculates the value for each pair of tables. Then, in the process of matching and calculating SIM values, a table was determined and placed as a DBSource.

Table 3 summarizes the test results and compares the steps for user validation of attribute pairs based on transposed DBSource and DBTarget placements.

Table 3 shows that the whole test is about choosing the right database to play as the DBTarget. In comparison with Sutanta et al. (2016b) reduced the number of steps a user has to perform to validate an attribute pair from 88,152 steps to 38,232 steps. Or a reduction of 49,920 steps, or an average reduction of 19.41%. This result shows that setting a database as the DBTarget in the schema matching process affects the number of process output verification models by the user. Databases with fewer attributes than DBTarget are preferred to minimize the validation process on the model's output.

In general, if the user declares the number of validation operations as cvptt on the incorret database pair attribute pair and cvpt specifies the number of validation operations on the correct DBSource and DBTarget pair, the validation process will be rejected. The user calculates using Equation (6) as follows:

IF (caDBs – caSDBs) > (caDBt – caSDBt) THEN cvptt = (caDBs-caSDBs) × (caDBt-caSDBt) AND IF (caDBs – caSDBs) < (caDBt – caSDBt) THEN cvpt = ABS((caDBs-caSDBs) × (caDBt-caSDBt)) (6)

When the user performs a difference count check on the attribute pair represented by deltacvp, the value could be calculated using Equation (7):

deltacvp = ABS (cvptt – cvpt) (7)

Table 3. Comparison of output model verification process based on transposed DBSource and DBTarget placements

DBSource		DBTarget		User Verification Steps		User Verification Step Reduction
Database Name	Number of Attribute	Database Name	Number of Attribute	Number of Attribute (DBSource > DBTarget)	Number of Attribute (DBSource < DBTarget)	
admision	127	admision	127	1,524	1,524	0
admision	127	academic	260	3,120	1,524	1,596
academic	260	payroll	61	3,120	732	2,388
academic	260	employ	124	3,120	1,488	1,632
academic	260	tax_pph	46	3,120	552	2,568
academic	260	workshop	15	3,120	180	2,940
academic	260	library	246	3,120	2,952	168
academic	260	user	53	3,120	636	2,484
dptkp	33	dptkp	33	396	396	0
dptkp	33	dptbgcpt	16	396	192	204
dptkp	33	quickcountbgcpt	31	396	372	24
dptkp	33	lisence	31	396	372	24
dptkp	33	lisenceol	57	684	396	288
dptkp	33	dptbtl	33	396	396	0
dptkdy	33	rsmitra	51	612	396	216
dptkdy	33	motorcred	19	396	228	168
nuptk	519	sinisa	63	6,228	756	5,472
nuptk	519	sipp	113	6,228	1,356	4,872
nuptk	519	nuptk	519	6228	6,228	0
nuptk	519	psb	787	9,444	6,228	3,216
sipp	113	sipp	113	1,356	1,356	0
sipp	113	announcement	13	1,356	156	1,200
sipp	113	forum	17	1,356	204	1,152
sipp	113	sinisa	63	1,356	756	600
sipp	113	osis	81	1,356	972	384
sipp	113	webinfo	161	1,932	1,356	576
sipp	113	sma2pwt	331	3,972	1,356	2,616
sipp	113	grade	92	1,356	1,104	252
sipp	113	report	157	1,884	1,356	528
sipp	113	psb	787	9,444	1,356	8,088
sipp	113	elearning	635	7,620	1,356	6,264
Sum or Average:				88,152	38,232	49,920

4. CONCLUSION

Based on this research, three new features have been added. That is, using SVL, checking input database similarity between attributes, and choosing an appropriate database to act as the DBSource during the matching process. These features reduce the user validation process in hybrid schema matching. In our next work, we will improve the model so that it can be applied automatically to heterogeneous systems, both RDBMS and application domains. The model has also been modified to provide output in the form of one-to-many or many-to-one attribute pair affinity mappings.

ACKNOWLEDGEMENT

The authors would like to thanks to The Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, that support this research through grant funding under Fundamental Research Grant Contract Number: SP DIPA-023.17.1.690439/2021.

CONFLICT OF INTERESTS

The authors would like to declare no conflict of interest in the publication of this manuscript.

REFERENCES

Banek, M., Vrdoljak, B., Tjoa, A.M., Skočir, Z. 2008. Automated integration of heterogeneous data warehouse schemas. The International Journal of Data Warehousing and Mining, 4, 1–21.

Bergamaschi, S., Castano, S., Vincini, M. 1999. Semantic integration of semistructured and structured data sources. SIGMOD Record, 28, 54–59.

Bernstein, P.A. 2003. Applying model management to classical meta data problems. In CIDR, 2003, 209–220.

Bernstein, P.A., Harry, B., Sanders, P., Shutt, D., Zander, J. 1997. The microsoft repository. In VLDB, 97, 25–29.

Kavitha, C., Sadasivam, G.S., Shenoy, S.N. 2011. Ontology based semantic integration of heterogeneous databases. European Journal of Scientific Research, 64, 120–127.

Kim, W., Seo, J. 1991. Classifying schematic and data heterogeneity in multidatabase systems. Computer, 24, 12–18.

Lee, C.Y., Ibrahim, H., Othman, M., Yaakob, R. 2009. Reconciling semantic conflicts in electronic patient data exchange. Proceedings of the 11th International Conference on Information Integration and Web-based Applications and Services, Kuala Lumpur, Malaysia. 390–394.

Li, W.S., Clifton, C. 2000. SEMINT: A tool for identifying attribute correspondences in heterogeneous databases using neural networks. Data Knowledge Engineering. 33, 49–84.

Li, J., Tang, J., Li, Y., Luo, Q. 2009. RiMOM: A dynamic multistrategy ontology alignment framework. IEEE Transactions on Knowledge and Data Engineering, 21, 1218–1232.

Madhavan, J., Bernstein, P.A., Rahm, E. 2001. Generic schema matching with cupid. In VLDB, 1, 49–58.

Massmann, S., Raunich, S., Aumüller, D., Arnold, P., Rahm E. 2011. Evolution of the COMA match system. Proceedings of the 6th International Conference on Ontology Matching, Bonn, Germany, 814, 49–60.

Melnik, S., Rahm, E., Bernstein, P. 2003. Rondo: A programming platform for generic model management. Proceedings of the 2003 ACM SIGMOD International conference on Management of data, New York, USA. 193–204.

Milo, T., Zohar, S. 1998. Using schema matching to simplify heterogeneous data translation. In VLDB, 98, 24–27.

Özsu, M.T., Valduriez, P. 2011. Principles of distributed database systems. Third Edition. Springer. New York.

Rahm E., Bernstein, P.A. 2001. A survey of approaches to automatic schema matching. The International Journal on Very Large Data Bases, 10, 334–350.

Schmidts, O., Kraft, B., Siebigteroth, I., Zündorf, A. 2019. Schema matching with frequent changes on semi-structured input files: A machine learning approach on biological product data. In ICEIS, 208–215.

Sutanta, E., Nurnawati, E.K., Kumalasanti, R.A., 2019. The mathematical model of hybrid schema matching based on constraints and instances similarity. International Journal of Advanced Computer Science and Applications, 10, 459–464.

Sutanta, E., Nurnawati, E.K., Kumalasanti, R.A., 2021. Using variation in weighting criteria and string size matching on hybrid model schema matching. International Journal on Advanced Science, Engineering and Information Technology, 11, 326–333.

Sutanta, E., Wardoyo, R., Mustofa, K., Winarko, E., 2016a. Survey: Models and prototypes of schema matching. International Journal of Electrical and Computer Engineering, 6, 1011–1022.

Sutanta, E., Wardoyo, R., Mustofa, K., Winarko, E., 2016b. A hybrid model schema matching using constraint-based and instance-based. International Journal of Electrical and Computer Engineering, 6, 1048–1058.