International Journal of Advanced Computer Science and Applications • *Open Access* • Volume 10, Issue 6, Pages 459 - 464 • 2019

Document type
Article • *Gold Open Access*

Source type
Journal

ISSN
2158107X

DOI
10.14569/ijacsa.2019.0100659

View more

The mathematical model of hybrid schema matching based on constraints and instances similarity

Sutanta E., Nurnawati E.K., Kumalasanti R.A.

Save all to author list

[a] Department of Informatics Engineering, Institut Sains and Teknologi AKPRIND Yogyakarta, Yogyakarta, Indonesia

10
Views count

View all metrics

View PDF Full text options

Abstract
Author keywords
Indexed keywords
SciVal Topics
Metrics
Funding details

Abstract

Schema matching is a crucial issue in applications that involve multiple databases from heterogeneous sources. Schema matching evolves from a manual process to a semi-automated process to effectively guide users in finding commonalities between schema elements. New models are generally developed using a combination of methods to improve the effectiveness of schema matching results. Our previous research has developed a prototype of hybrid schema matching utilizing a combination of constraints-based method and an instance-based method. The innovation of this paper presents a mathematical

Cited by 0 documents

Inform me when this document is cited in Scopus:

Set citation alert

Related documents

A hybrid model schema matching using constraint-based and instance-based
Sutanta, E. , Wardoyo, R. , Mustofa, K.
(2016) International Journal of Electrical and Computer Engineering

Using Variation in Weighting Criteria and String Size Matching on Hybrid Model Schema Matching
Sutanta, E. , Nurnawati, E.K. , Kumalasanti, R.A.
(2021) International Journal on Advanced Science, Engineering and Information Technology

Survey: Models and prototypes of schema matching
Sutanta, E. , Wardoyo, R. , Mustofa, K.
(2016) International Journal of Electrical and Computer Engineering

View all related documents based on references

Find more related documents in Scopus based on:

Authors Keywords

formulation of a hybrid schema matching model so it can be run for different cases and becomes the basis of development to improve the effectiveness of output and or efficiency during schema matching process. The developed mathematical model serves to perform the main task in the schema matching process that matches the similarity between attributes, calculates the similarity value of the attribute pair, and specifies the matching attribute pair. Based on the test results, a hybrid schema matching model is more effective than the constraints-based method or instance-based method run individually. The more matching criteria used in the schema matching provide better mapping results. The model developed is limited to schema matching processes in the relational model database. © 2019 International Journal of Advanced Computer Science and Applications.

Author keywords

Constraint-based; Hybrid schema matching model; Instance-based; Mathematical model

Indexed keywords

SciVal Topics

Metrics

Funding details

References (31)

View in search results format >

All

CSV export | Print | E-mail | Save to PDF

Create bibliography

1 Leme, L.A.P.P., Casanova, M.A., Breitman, K.K., Furtado, A.L.

OWL schema matching (Open Access)

(2010) *Journal of the Brazilian Computer Society*, 16 (1), pp. 21-34. Cited 4 times.
doi: 10.1007/s13173-010-0005-3

View at Publisher

2 Cruz, I.F., Antonelli, F.P., Stroe, C.

AgreementMaker: Efficient matching for large real-world schemas and ontologies

(2009) *Proceedings of the VLDB Endowment*, 2 (2), pp. 1586-1589. Cited 212 times.
http://www.vldb.org/pvldb/2/vldb09-1003.pdf
doi: 10.14778/1687553.1687598

View at Publisher

3 Kavitha, C., Sadasivam, G.S., Shenoy, S.N.

Ontology based semantic integration of heterogeneous databases

(2011) *European Journal of Scientific Research*, 64 (1), pp. 120-127. Cited 11 times.
http://www.europeanjournalofscientificresearch.com/ISSUES/EJSR_64_1_11.pdf

4 Hamill, S., Dixon, M., Read, B.J., Kalmus, J.R.
Interoperating Database Systems: Issues and Architectures
(1997) *Council for the Central Laboratory of The Research Councils* United Kingdom, Technical Report

5 (2002) *SIFONAS Sebagai Tulang Punggung e-Governance (SIFONAS as The Backbone of e-Governance)*

6 Martinek, P.
Schema Matching Methodologies and Runtime Solutions in SOA Based Enterprise Application Integration
(2009) *Department of Electronics Technology, Faculty of Electrical Engineering & Informatics*. Cited 3 times.
Budapest University of Technology and Economics, Hungary, Ph.D. Thesis

7 Sutanta, E., Wardoyo, R., Mustofa, K., Winarko, E.
A hybrid model schema matching using constraint-based and instance-based

(2016) *International Journal of Electrical and Computer Engineering*, 6 (3), pp. 1048-1058. Cited 7 times.
http://www.iaescore.com/journals/index.php/IJECE/article/view/458/333
doi: 10.11591/ijece.v6i3.9790

View at Publisher

8 Alani, H., Saad, S.
Schema matching for large-scale data based on ontology clustering method (Open Access)

(2017) *International Journal on Advanced Science, Engineering and Information Technology*, 7 (5), pp. 1790-1797. Cited 5 times.
http://www.insightsociety.org/ojaseit/index.php/ijaseit/article/download/2133/1232
doi: 10.18517/ijaseit.7.5.2133

View at Publisher

9 Shafa'amri, K.H., Atoum, J.O.
A Framework for Improving the Performance of Ontology Matching Techniques in Semantic Web
(2012) *International Journal of Advanced Computer Science and Applications (IJACSA)*, 3 (1), pp. 8-14. Cited 5 times.

10 Azizul, N.H., Zin, A.M., Sundararajan, E.
The design and implementation of middleware for application development within honeybee computing environment

(2016) *International Journal on Advanced Science, Engineering and Information Technology*, 6 (6), pp. 937-943. Cited 4 times.
http://www.insightsociety.org/ojaseit/index.php/ijaseit/article/download/1415/904
doi: 10.18517/ijaseit.6.6.1415

View at Publisher

11 Sulisworo, D., Tawar, A.U.
ICT Based Information Flows and Supply Chain in Integrating Academic Business Process
(2012) *International Journal on Advanced Science, Engineering and Information Technology (IJASEIT)*, 2 (6), pp. 44-48. Cited 3 times.

12 Mohammadi, M., Mukhtar, M.
Business Process Modelling Languages in Designing Integrated Information Systems for Supply Chain Management
(2012) *International Journal on Advanced Science, Engineering and Information Technology (IJASEIT)*, 2 (6), pp. 464-467. Cited 8 times.

13 Sutanta, E., Wardoyo, R., Mustofa, K., Winarko, E.
Survey: Models and prototypes of schema matching

(2016) *International Journal of Electrical and Computer Engineering*, 6 (3), pp. 1011-1022. Cited 15 times.
http://www.iaescore.com/journals/index.php/IJECE/article/view/455/330
doi: 10.11591/ijece.v6i3.9789

View at Publisher

14 Li, W.S., Clifton, C.
Semantic Integration in Heterogeneous Databases Using Neural Networks
(1994) *in The 20th International Conference on Very Large Data Bases (VLDB)*, pp. 1-12. Cited 125 times.
Santiago de Chile, Chile

15 Rahm, E.
Schema Matching and Mapping: Towards Large-Scale Schema and Ontology Matching
(2011) *in Data Centric Systems and Applications*, pp. 3-27. Cited 282 times.
Z. Bellahsene, A. Bonifati, and E. Rahm, Eds. New York, USA: Springer

16 Do, H.-H., Melnik, S., Rahm, E.
Comparison of schema matching evaluations (Open Access)

(2003) *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2593, pp. 221-237. Cited 187 times.
https://www.springer.com/series/558
doi: 10.1007/3-540-36560-5_17

View at Publisher

17 Do, H.H.
Schema Matching and Mapping-Based Data Integration
(2005) *Interdisciplinary Center for Bioinformatics and Department of Computer Science*. Cited 62 times.
University of Leipzig, Leipzig, Germany, Ph.D. Thesis

18 Engmann, D., Massmann, S.
Instance Matching with COMA++
(2007) *in Datenbank Systeme in Business*, pp. 28-37. Cited 48 times.
Technologie und Web (BTW Workshop): Model Management and Metadata, Aachen, Germany
https://dbs.uni-leipzig.de/file/BTW-Workshop_2007_EngmannMassmann.pdf

19 Li, J., Tang, J., Li, Y., Luo, Q.
RiMOM: A dynamic multistrategy ontology alignment framework (Open Access)

(2009) *IEEE Transactions on Knowledge and Data Engineering*, 21 (8), art. no. 4633358, pp. 1218-1232. Cited 323 times.
doi: 10.1109/TKDE.2008.202

View at Publisher

20 Bernstein, P.A., Harry, B., Sanders, P., Shutt, D., Zander, J.
The microsoft repository

(1997) *Proceedings of the 23rd International Conference on Very Large Databases, VLDB 1997*, pp. 3-12. Cited 21 times.
https://www.onepetro.org/conferences/SPE/16SATS
ISBN: 1558604707; 978-155860470-4

21 Birney, E., Andrews, T.D., Bevan, P., Caccamo, M., Chen, Y., Clarke, L., Coates, G., (...), Clamp, M.
An overview of Ensembl (Open Access)

(2004) *Genome Research*, 14 (5), pp. 925-928. Cited 324 times.
doi: 10.1101/gr.1860604

View at Publisher

22 Madhavan, J., Bernstein, P.A., Rahm, E.
Generic schema matching with cupid

(2001) *VLDB 2001 - Proceedings of 27th International Conference on Very Large Data Bases*, pp. 49-58. Cited 1042 times.
ISBN: 1558608044; 978-155860804-7

23 Melnik, S., Rahm, E., Bernstein, P.A.
Rondo: A Programming Platform for Generic Model Management

(2003) *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 193-204. Cited 233 times.

View at Publisher

24 Bernstein, P.A., Madhavan, J., Rahm, E.
Generic schema matching, ten years later

(2011) *Proceedings of the VLDB Endowment*, 4 (11), pp. 695-701. Cited 205 times.
http://www.vldb.org/pvldb/vol4/p695-bernstein_madhavan_rahm.pdf

View at Publisher

25 Alwan, A.A., Nordin, A., Alzeber, M., Abualkishik, A.Z.
A Survey of Schema Matching Research using Database Schemas and Instances
(2017) *International Journal of Advanced Computer Science and Applications (IJACSA)*, 8 (10), pp. 102-111. Cited 15 times.

26 Bergamaschi, S., Castano, S., Vincini, M.
Semantic integration of semistructured and structured data sources

(1999) *SIGMOD Record (ACM Special Interest Group on Management of Data)*, 28 (1), pp. 54-59. Cited 236 times.
http://dl.acm.org/citation.cfm?id=J689
doi: 10.1145/309844.309897

View at Publisher

27 Milo, T., Zohar, S.
Using Schema Matching to Simplify Heterogeneous Data Translation
(1998) *in Proceedings of The 24th International Conference on Very Large Data bases (VLDB)*, pp. 122-133. Cited 244 times.
New York, USA
http://dl.acm.org/citation.cfm?id=645924.671326

28 Doan, A., Domingos, P., Halevy, A.
Reconciling schemas of disparate data sources: A machine-learning approach

(2001) *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 509-520. Cited 420 times.

29 Gross, A., Hartung, M., Kirsten, T., Rahm, E.
On matching large life science ontologies in parallel
(Open Access)

(2010) *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6254 LNBI, pp. 35-49. Cited 34 times.
ISBN: 3642151191; 978-364215119-4
doi: 10.1007/978-3-642-15120-0_4

View at Publisher

30 Karasneh, Y., Ibrahim, H., Othman, M., Yaakob, R.
An approach for matching schemas of heterogeneous relational databases

(2010) *Journal of Digital Information Management*, 8 (4), pp. 260-269. Cited 3 times.
http://www.dline.info/fpaper/jdim/v8i4/6.pdf

31 Shuaibu, M.B.
Determining an Appropriate Weight Attribute in Fraud Call Rate Data Using Case-Based Reasoning
(2014) *International Journal on Advanced Science, Engineering and Information Technology (IJASEIT)*, 4 (1), pp. 34-36. Cited 2 times.

also developed by scimago: SCIMAGO INSTITUTIONS RANKINGS

SJR Scimago Journal & Country Rank

Enter Journal Title, ISSN or Publisher Name

Home Journal Rankings Country Rankings Viz Tools Help About Us

Among One of the Best Journals

Reputed Double-blind Peer Reviewed Widely Indexed Multidisciplinary International Journal

IJIRMPS

Open

International Journal of Advanced Computer Science and Applications

COUNTRY

United Kingdom

Universities and research institutions in United Kingdom

SUBJECT AREA AND CATEGORY

Computer Science
Computer Science (miscellaneous)

PUBLISHER

Science and Information Organization

Among One of the Best Journals

Open

PUBLICATION TYPE

Journals

ISSN

21565570, 2158107X

COVERAGE

2017-2020

IJIRMPS

SCOPE

Information not localized

Join the conversation about this journal

CV. Exindo Pratama | Kontraktor Pameran, Event Organizer, Desain Booth Pameran
pembuatan stand kuliner
vendor Berpengalaman Kami Siap Membantu Anda. Lihat Layanan Kami Sekarang.

Directions Website

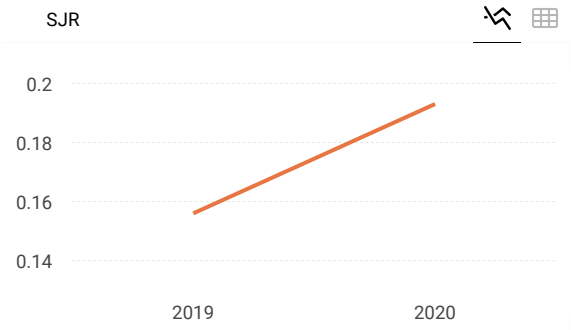
Quartiles

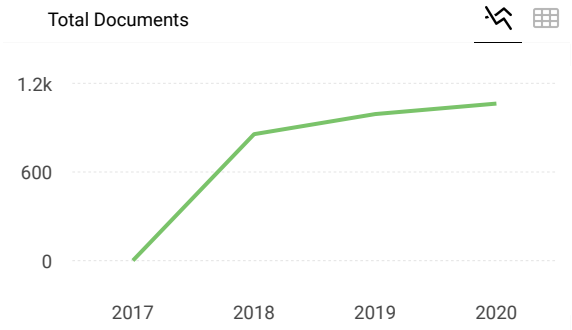
FIND SIMILAR JOURNALS

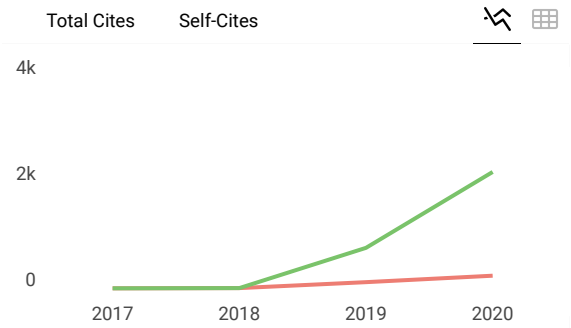
Among One of the Best Journals

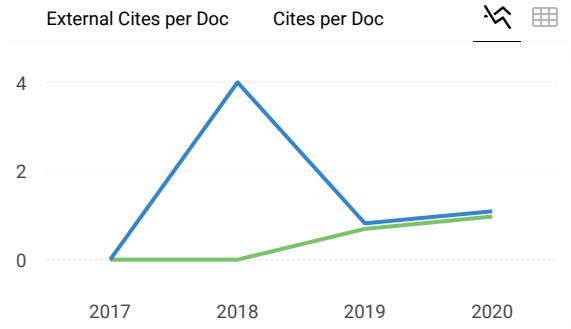
IJIRMPS

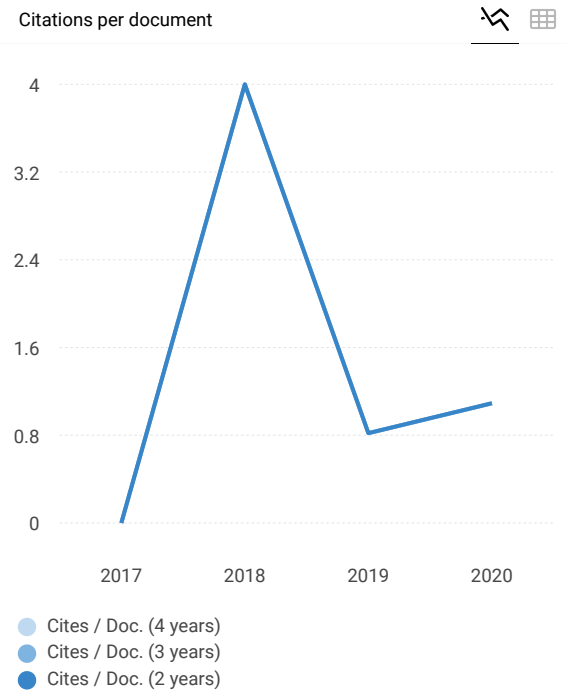
1	2	3	4
Journal of Computer Science	**Information (Switzerland)**	**Journal of King Saud University - Computer and**	**Computers**
USA	CHE	SAU	CHE
77% similarity	**76%** similarity	**76%** similarity	**72%** simila

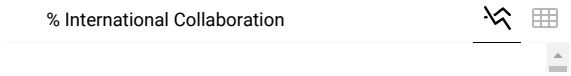


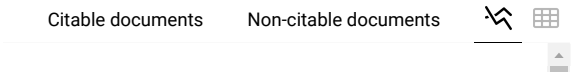


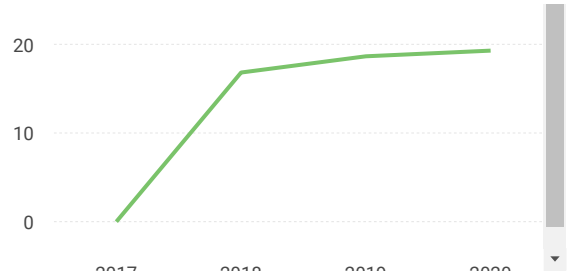


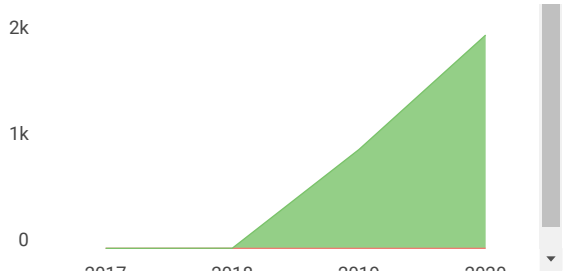


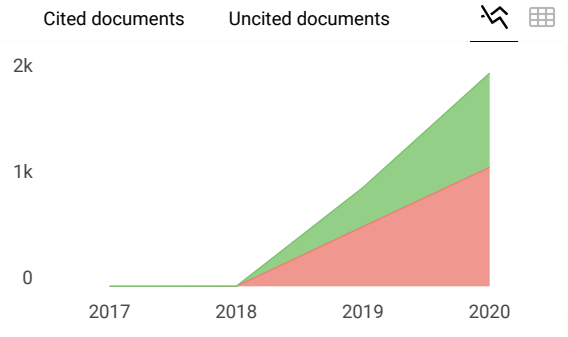












← Show this widget in your own website

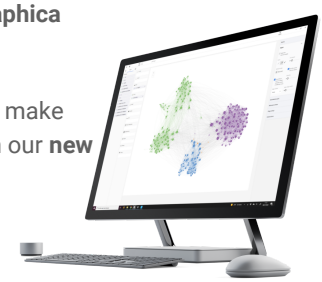
Just copy the code below and paste within your html code:

```
<a href="https://www.scimag
```

SCImago Graphica

Explore, visually communicate and make sense of data with our **new free tool**.

Get it



Metrics based on Scopus® data as of April 2021

N **Nesrine Abajadddi** 1 month ago

Please I want know the Q (Q1 Q2 Q3?) of this journal in 2021.

reply

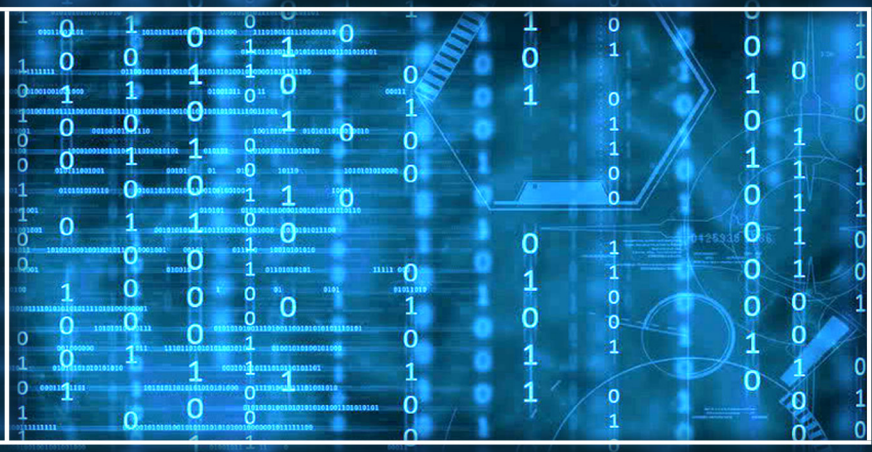
SCImago Team

Melanie Ortiz 4 weeks ago

Dear Nesrine,
Thank you for contacting us. Our data come from Scopus, they annually send us an update of the data. This update is sent to us around April / May every year. The SJR for 2020 was released on 17 May 2021. Therefore, the indicators for 2021 will be available in

IJACSA
WHERE WISDOM SHARES
International Journal of Advanced Computer Science and Applications

Volume 10 Issue 6
June 2019



ISSN 2156-5570(Online)
ISSN 2158-107X(Print)

SAI
www.ijacsa.thesai.org

Editorial Preface

From the Desk of Managing Editor…

It may be difficult to imagine that almost half a century ago we used computers far less sophisticated than current home desktop computers to put a man on the moon. In that 50 year span, the field of computer science has exploded.

Computer science has opened new avenues for thought and experimentation. What began as a way to simplify the calculation process has given birth to technology once only imagined by the human mind. The ability to communicate and share ideas even though collaborators are half a world away and exploration of not just the stars above but the internal workings of the human genome are some of the ways that this field has moved at an exponential pace.

At the International Journal of Advanced Computer Science and Applications it is our mission to provide an outlet for quality research. We want to promote universal access and opportunities for the international scientific community to share and disseminate scientific and technical information.

We believe in spreading knowledge of computer science and its applications to all classes of audiences. That is why we deliver up-to-date, authoritative coverage and offer open access of all our articles. Our archives have served as a place to provoke philosophical, theoretical, and empirical ideas from some of the finest minds in the field.

We utilize the talents and experience of editor and reviewers working at Universities and Institutions from around the world. We would like to express our gratitude to all authors, whose research results have been published in our journal, as well as our referees for their in-depth evaluations. Our high standards are maintained through a double blind review process.

We hope that this edition of IJACSA inspires and entices you to submit your own contributions in upcoming issues. Thank you for sharing wisdom.

Thank you for Sharing Wisdom!

Managing Editor
IJACSA
Volume 10 Issue 6 June 2019
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)
©2013 The Science and Information (SAI) Organization

Editorial Board

Editor-in-Chief

Dr. Kohei Arai - Saga University
Domains of Research: Technology Trends, Computer Vision, Decision Making, Information Retrieval, Networking, Simulation

---

Associate Editors

Chao-Tung Yang
Department of Computer Science, Tunghai University, Taiwan
Domain of Research: Software Engineering and Quality, High Performance Computing, Parallel and Distributed Computing, Parallel Computing

Elena SCUTELNICU
"Dunarea de Jos" University of Galati, Romania
Domain of Research: e-Learning, e-Learning Tools, Simulation

Krassen Stefanov
Professor at Sofia University St. Kliment Ohridski, Bulgaria
Domains of Research: e-Learning, Agents and Multi-agent Systems, Artificial Intelligence, Big Data, Cloud Computing, Data Retrieval and Data Mining, Distributed Systems, e-Learning Organisational Issues, e-Learning Tools, Educational Systems Design, Human Computer Interaction, Internet Security, Knowledge Engineering and Mining, Knowledge Representation, Ontology Engineering, Social Computing, Web-based Learning Communities, Wireless/ Mobile Applications

Maria-Angeles Grado-Caffaro
Scientific Consultant, Italy
Domain of Research: Electronics, Sensing and Sensor Networks

Mohd Helmy Abd Wahab
Universiti Tun Hussein Onn Malaysia
Domain of Research: Intelligent Systems, Data Mining, Databases

T. V. Prasad
Lingaya's University, India
Domain of Research: Intelligent Systems, Bioinformatics, Image Processing, Knowledge Representation, Natural Language Processing, Robotics

CONTENTS

Paper 1: AHP-based Security Decision Making: How Intention and Intrinsic Motivation Affect Policy Compliance
Authors: Ahmed Alzahrani, Christopher Johnson
PAGE 1 – 8

Paper 2: Prediction of Visibility for Color Scheme on the Web Browser with Neural Networks
Authors: Miki Yamaguchi, Yoshihisa Shinozawa
PAGE 9 – 18

Paper 3: Application of the Scattering Bond Graph Methodology for Composite Right/Left Handed Transmission Lines
Authors: Islem Salem, Hichem Taghouti, Ahmed Rahmani, Abdelkader Mami
PAGE 19 – 24

Paper 4: Expert System for Milk and Animal Monitoring
Authors: Todor Todorov, Juri Stoinov
PAGE 25 – 30

Paper 5: Innovative Means of Medical Students Teaching through Graphical Methods for Cardiac Data Estimating and Serious Games
Authors: Galya N. Georgieva-Tsaneva
PAGE 31 – 39

Paper 6: Muscles Heating Analysis in Sportspeople to Prevent Muscle Injuries using Thermal Images
Authors: Brian Meneses-Claudio, Witman Alvarado-Díaz, Fiorella Flores-Medina, Natalia I. Vargas-Cuentas, Avid Roman-Gonzalez
PAGE 40 – 44

Paper 7: Advanced Metaheuristics-based Tuning of Effective Design Parameters for Model Predictive Control Approach
Authors: Mohamed Lotfi Derouiche, Soufiene Bouallègue, Joseph Haggège, Guillaume Sandou
PAGE 45 – 53

Paper 8: Secure Medical Internet of Things Framework based on Parkerian Hexad Model
Authors: Nidal Turab, Qasem Kharma
PAGE 54 – 62

Paper 9: Two Dimensional Electronic Nose for Vehicular Central Locking System (E-Nose-V)
Authors: Mahmoud Zaki Iskandarani
PAGE 63 – 70

Paper 10: Forensic Analysis using Text Clustering in the Age of Large Volume Data: A Review
Authors: Bandar Almaslukh
PAGE 71 – 76

Paper 11: Internet of Things (IOT): Research Challenges and Future Applications
Authors: AbdelRahman H. Hussein
PAGE 77 – 82

Paper 12: Classification of Hand Movements based on Discrete Wavelet Transform and Enhanced Feature Extraction
Authors: Jingwei Too, Abdul Rahim Abdullah, Norhashimah Mohd Saad
PAGE 83 – 89

Paper 13: Modified Graph-theoretic Clustering Algorithm for Mining International Linkages of Philippine Higher Education Institutions
Authors: Sheila R. Lingaya, Bobby D. Gerardo, Ruji P. Medina
PAGE 90 – 95

Paper 14: Digital Preservation of Cultural Heritage: Terengganu Brassware Craft Knowledge Base
Authors: Wan Malini Wan Isa, Nor Azan Mat Zin, Fadhilah Rosdi, Hafiz Mohd Sarim
PAGE 96 – 102

Paper 15: Computer Students Attitudes on the Integration of m-Learning Applications
Authors: Abdulmohsin S. Alkhunaizan
PAGE 103 – 108

Paper 16: Exploring the Use of Digital Games as a Persuasive Tool in Teaching Islamic Knowledge for Muslim Children
Authors: Madihah Sheikh Abdul Aziz, Panadda Auyphorn, Mohd Syarqawy Hamzah
PAGE 109 – 113

Paper 17: Brain-Controlled for Changing Modular Robot Configuration by Employing Neurosky's Headset
Authors: Muhammad Haziq Hasbulah, Fairul Azni Jafar, Mohd. Hisham Nordin, Kazutaka Yokota
PAGE 114 – 120

Paper 18: Analysis of Spatially Modelled High Temperature Polymer Electrolyte Membrane Fuel Cell under Dynamic Load Conditions
Authors: Jagdesh Kumar, Jherna Devi, Ghulam Mustafa Bhutto, Sajida Parveen, Muhammad Shafiq
PAGE 121 – 130

Paper 19: A Collaborative Filtering Recommender System Model for Recommending Intervention to Improve Elderly Well-being
Authors: Aini Khairani Azmi, Noraswaliza Abdullah, Nurul Akmar Emran
PAGE 131 – 138

Paper 20: Forecasting Feature Selection based on Single Exponential Smoothing using Wrapper Method
Authors: Ani Dijah Rahajoe
PAGE 139 – 145

Paper 21: A Review on the Verification Approaches and Tools used to Verify the Correctness of Security Algorithms and Protocols
Authors: Mohammed Abdulqawi Saleh Al-humaikani, Lukman Bin Ab Rahim
PAGE 146 – 152

Paper 22: Performance Comparison of Detection, Recognition and Tracking Rates of the different Algorithms
Authors: Meghana Kavuri, Kolla Bhanu Prakash
PAGE 153 – 158

Paper 23: QoS-based Semantic Micro Services Discovery and Composition using ACO Algorithm
Authors: Ahmed ESSAYAH, Mohamed Youssfi, Omar Bouattane, Khalifa Mansouri, Elhocein Illoussamen
PAGE 159 – 168

Paper 24: An Approach to Control the Positional Accuracy of Point Features in Volunteered Geographic Information Systems
Authors: Mennatallah H. Ibrahim, Nagy Ramadan Darwish, Hesham A. Hefny
PAGE 169 – 175

Paper 25: Android Security Development: SpywareDetection, Apps Secure Level and Data Encryption Improvement
Authors: Lim Wei Xian, Chan Shao Hong, Yap Ming Jie, Azween Abdullah, Mahadevan Supramaniam
PAGE 176 – 187

Paper 26: An Aspect Oriented Programming Framework to Support Transparent Runtime Monitoring of Applications
Authors: Abdullah O. AL-Zaghameem
PAGE 188 – 193

Paper 27: A Novel Intelligent Cluster-Head (ICH) to Mitigate the Handover Problem of Clustering in VANETs
Authors: A.H. Abbas, Mohammed I. Habelalmateen, L. Audah, N.A.M. Alduais
PAGE 194 – 203

Paper 28: Weld Defect Categorization from Welding Current using Principle Component Analysis
Authors: Hayri Arabaci, Salman Laving
PAGE 204 – 211

Paper 29: Moving Object Detection in Highly Corrupted Noise using Analysis of Variance
Authors: Asim ur Rehman Khan, Muhammad Burhan Khan, Haider Mehdi, Syed Muhammad Atif Saleem
PAGE 212 – 216

Paper 30: Smart Home Energy Management System Design: A Realistic Autonomous V2H / H2V Hybrid Energy Storage System
Authors: Bassam Zafar, Ben Slama Sami, Sihem Nasri, Marwan Mahmoud
PAGE 217 – 223

Paper 31: Evaluation of the Performance of the University Information Systems: Case of Moroccan Universities
Authors: Ayoub Gacim, Hicham Drissi, Abdelwahed Namir
PAGE 224 – 230

Paper 32: Immuno-Computing-based Neural Learning for Data Classification
Authors: Ali Al Bataineh, Devinder Kaur
PAGE 231 – 237

Paper 33: Predictive Method for Service Composition in Heterogeneous Environments within Client Requirements
Authors: Saleh M. Altowaijri
PAGE 238 – 244

Paper 34: Spectral Classification of a Set of Hyperspectral Images using the Convolutional Neural Network, in a Single Training
Authors: Abdelali Zbakh, Zoubida Alaoui Mdaghri, Abdelillah Benyoussef, Abdellah El Kenz, Mourad El Yadari
PAGE 245 – 250

Paper 35: Optical Recognition of Isolated Machine Printed Sindhi Characters using Fourier Descriptors
Authors: Nasreen Nizamani, Mujtaba Shaikh, Jawed Unar, Ehsan Ali, Ghulam Mustafa Bhutto, Abdul Rafay
PAGE 251 – 255

Paper 36: Junction Point Detection and Identification of Broken Character in Touching Arabic Handwritten Text using Overlapping Set Theory
Authors: Inam Ullah, Mohd Sanusi Azmi, Mohamad Ishak Desa
PAGE 256 – 260

Paper 37: Implementation of Machine Learning Model to Predict Heart Failure Disease
Authors: Fahd Saleh Alotaibi
PAGE 261 – 268

Paper 38: Hyperparameter Optimization in Convolutional Neural Network using Genetic Algorithms
Authors: Nurshazlyn Mohd Aszemi, P.D.D Dominic
PAGE 269 – 278

Paper 39: New Method of Faults Diagnostic based on Neuro-Dynamic Sliding Mode for Flat Nonlinear Systems
Authors: O.Dhaou, L.Sidhom, A.Abdelkrim
PAGE 279 – 291

Paper 40: Big Data Technology-Enabled Analytical Solution for Quality Assessment of Higher Education Systems
Authors: Samiya Khan, Xiufeng Liu, Kashish Ara Shakil, Mansaf Alam
PAGE 292 – 304

Paper 41: A Novel Network user Behaviors and Profile Testing based on Anomaly Detection Techniques
Authors: Muhammad Tahir, Mingchu Li, Xiao Zheng, Anil Carie, Xing Jin, Muhammad Azhar, Naeem Ayoub, Atif Wagan, Muhammad Aamir, Liaquat Ali Jamali, Muhammad Asif Imran, Zahid Hussain Hulio
PAGE 305 – 324

Paper 42: Experimental Analysis of Color Image Scrambling in the Spatial Domain and Transform Domain
Authors: R. Rama Kishore, Sunesh
PAGE 325 – 333

Paper 43: Hijaiyah Letter Interactive Learning for Mild Mental Retardation Children using Gillingham Method and Augmented Reality
Authors: Irawan Afrianto, Agung Faishal Faris, Sufa Atin
PAGE 334 – 341

Paper 44: Causal Impact Analysis on Android Market
Authors: Hadiqa AmanUllah, Mishal Fatima, Umair Muneer, Sadaf Ilyas, Rana Abdul Rehman, Ibraheem Afzal
PAGE 342 – 350

Paper 45: Emotion Detection in Text using Nested Long Short-Term Memory
Authors: Daniel Haryadi, Gede Putra Kusuma
PAGE 351 – 357

Paper 46: Dynamic Matrix Control DMC using the Tuning Procedure based on First Order Plus Dead Time for Infant-Incubator
Authors: J. El Hadj Ali, E. Feki, A. Mami
PAGE 358 – 367

Paper 47: Dense Hand-CNN: A Novel CNN Architecture based on Later Fusion of Neural and Wavelet Features for Identity Recognition
Authors: Elaraby A. Elgallad, Wael Ouarda, Adel M. Alimi
PAGE 368 – 378

Paper 48: Detection of Suspicious of Diabetic Feet using Thermal Image
Authors: Brian Meneses-Claudio, Witman Alvarado-Díaz, Fiorella Flores-Medina, Natalia I. Vargas-Cuentas, Avid Roman-Gonzalez
PAGE 379 – 383

Paper 49: Blind Image Quality Evaluation of Stitched Image using Novel Hybrid Warping Technique
Authors: Sanjay T. Gandhe, Omkar S. Vaidya
PAGE 384 – 389

Paper 50: Bio-inspired Think-and-Share Optimization for Big Data Provenance in Wireless Sensor Networks
Authors: Adel Alkhalil, Rabie Ramadan, Aakash Ahmad
PAGE 390 – 397

Paper 51: Fragile Watermarking based on Linear Cellular Automata using Manhattan Distances for 2D Vector Map
Authors: Saleh AL-ardhi, Vijey Thayananthan, Abdullah Basuhail
PAGE 398 – 403

Paper 52: A Watermarking System Architecture using the Cellular Automata Transform for 2D Vector Map
Authors: Saleh AL-ardhi, Vijey Thayananthan, Abdullah Basuhail
PAGE 404 – 410

Paper 53: Electromyography Signal Acquisition and Analysis System for Finger Movement Classification
Authors: Alvarado-Díaz Witman, Meneses-Claudio Brian, Roman-Gonzalez Avid
PAGE 411 – 416

Paper 54: Blood Vessels Segmentation in Retinal Fundus Image using Hybrid Method of Frangi Filter, Otsu Thresholding and Morphology
Authors: Wiharto, YS. Palgunadi
PAGE 417 – 422

Paper 55: A New Image Inpainting Approach based on Criminisi Algorithm
Authors: Nouho Ouattara, Georges Laussane Loum, Ghislain Koffi Pandry, Armand Kodjo Atiampo
PAGE 423 – 433

Paper 56: Improving Knowledge Sharing in Distributed Software Development
Authors: Sara Waheed, Bushra Hamid, NZ Jhanjhi, Mamoona Humayun, Nazir A Malik
PAGE 434 – 443

Paper 57: IRPanet: Intelligent Routing Protocol in VANET for Dynamic Route Optimization
Authors: Rafi Ullah, Shah Muhammad Emad, Taha Jilani, Waqas Azam, Muhammad Zain uddin
PAGE 444 – 450

Paper 58: Depth Limitation and Splitting Criteria Optimization on Random Forest for Efficient Human Activity Classification
Authors: Syarif Hidayat, Ahmad Ashari, Agfianto Eko Putra
PAGE 451 – 458

Paper 59: The Mathematical Model of Hybrid Schema Matching based on Constraints and Instances Similarity
Authors: Edhy Sutanta, Erna Kumalasari Nurnawati, Rosalia Arum Kumalasanti
PAGE 459 – 464

Paper 60: The Role of Technical Analysis Indicators over Equity Market (NOMU) with R Programing Language
Authors: Mohammed A. Al Ghamdi
PAGE 465 – 471

Paper 61: A Circular Polarization RFID Tag for Medical Uses
Authors: Nada Jebali, Ali Gharsallah
PAGE 472 – 479

Paper 62: Passenger and Luggage Weight Monitoring System for Public Transport based on Sensing Technology: A Case of Zambia
Authors: Apolinalious Bwalya, Jackson Phiri, Monica M. Kalumbilo, David Zulu
PAGE 480 – 489

Paper 63: Survey Energy Management Approaches in Data Centres
Authors: Bouchra Morchid, Siham Benhadou, Mariam Benhadou, Abdellah Haddout, Hicham Medromi
PAGE 490 – 495

Paper 64: Comparative Study of Methods that Detect Levels of Lead and its Consequent Toxicity in the Blood
Authors: Kevin J. Rodriguez, Alicia Alva, Virginia T. Santos, Avid Roman-Gonzalez
PAGE 496 – 499

Paper 65: Smart Smoking Area based on Fuzzy Decision Tree Algorithm
Authors: Iswanto, Kunnu Purwanto, Weni Hastuti, Anis Prabowo, Muhamad Yusvin Mustar
PAGE 500 – 504

Paper 66: A Comprehensive Collaborating Filtering Approach using Extended Matrix Factorization and Autoencoder in Recommender System
Authors: Mahamudul Hasan, Falguni Roy, Tasdikul Hasan, Lafifa Jamal
PAGE 505 – 513

Paper 67: Satellite Image Enhancement using Wavelet-domain based on Singular Value Decomposition
Authors: Muhammad Aamir, Ziaur Rahman, Yi-Fei Pu, Waheed Ahmed Abro, Kanza Gulzar
PAGE 514 – 519

Paper 68: Convolutional Neural Networks in Predicting Missing Text in Arabic
Authors: Adnan Souri, Mohamed Alachhab, Badr Eddine Elmohajir, Abdelali Zbakh
PAGE 520 – 527

Paper 69: Multi-Modal Biometric: Bi-Directional Empirical Mode Decomposition with Hilbert-Hung Transformation
Authors: Gavisiddappa, Chandrashekar Mohan Patil, Shivakumar Mahadevappa, Pramod KumarS
PAGE 528 – 537

Paper 70: Heuristics Applied to Mutation Testing in an Impure Functional Programming Language
Authors: Juan Guti´errez-C´ardenas, Hernan Quintana-Cruz, Diego Mego-Fernandez, Serguei Diaz-Baskakov
PAGE 538 – 548

Paper 71: Exploiting the Interplay among Products for Efficient Recommendations
Authors: Anbarasu Sekar, Sutanu Chakraborti
PAGE 549 – 556

Paper 72: An Assessment of Open Data Sets Completeness
Authors: Abdulrazzak Ali, Nurul A. Emran, Siti A. Asmai, Amelia R. Ismail
PAGE 557 – 562

Paper 73: Design and Application of a Smart Diagnostic System for Parkinson's Patients using Machine Learning
Authors: Asma Channa, Attiya Baqai, Rahime Ceylan
PAGE 563 – 571

Paper 74: An Effective Framework for Tweet Level Sentiment Classification using Recursive Text Pre-Processing Approach
Authors: Muhammad Bux Alvi, Naeem A. Mahoto, Mukhtiar A. Unar, M. Akram Shaikh
PAGE 572 – 581

Paper 75: Competitive Algorithms for Online Conversion Problem with Interrelated Prices
Authors: Javeria Iqbal, Iftikhar Ahmad, Asadullah Shah
PAGE 582 – 589

Paper 76: Introducing Multi Shippers Mechanism for Decentralized Cash on Delivery System
Authors: Hai Trieu Le, Ngoc Tien Thanh Le, Nguyen Ngoc Phien, Nghia Duong-Trung, Ha Xuan Son, Thai Tam Huynh, The Phuc Nguyen
PAGE 590 – 597

Paper 77: Content-based Automatic Video Genre Identification
Authors: Faryal Shamsi, Sher Muhammad Daudpota, Sarang Shaikh
PAGE 598 – 607

Paper 78: A Comparison of Sentiment Analysis Methods on Amazon Reviews of Mobile Phones
Authors: Sara Ashour Aljuhani, Norah Saleh Alghamdi
PAGE 608 – 617

Paper 79: Virtualizing a Cluster to Optimize the Problems of High Scientific Complexity within an Organization
Authors: Enrique Lee Huamaní, Patricia Condori, Avid Roman-Gonzalez
PAGE 618 – 622

Paper 80: School Manager System based on a Personal Information Architecture
Authors: Elena Fabiola Ruiz Ledesma, Elizabeth Moreno Galván, Juan Jesús Gutiérrez García, Chadwick Carreto Arellano
PAGE 623 – 629

Paper 81: Fusing Identity Management, HL7 and Blockchain into a Global Healthcare Record Sharing Architecture
Authors: Mohammad A. R. Abdeen, Toqeer Ali, Yasar Khan, M. C.E. Yagoub
PAGE 630 – 636

Paper 82: An Efficient Machine Learning Technique to Classify and Recognize Handwritten and Printed Digits of Sudoku Puzzle
Authors: Sang C. Suh, Aghalya Dharshni Manmatharaj
PAGE 637 – 642

The Mathematical Model of Hybrid Schema Matching based on Constraints and Instances Similarity

Edhy Sutanta[1], Erna Kumalasari Nurnawati[2], Rosalia Arum Kumalasanti[3]
Department of Informatics Engineering, Institut Sains & Teknologi AKPRIND Yogyakarta Yogyakarta, Indonesia

***Abstract*—Schema matching is a crucial issue in applications that involve multiple databases from heterogeneous sources. Schema matching evolves from a manual process to a semi-automated process to effectively guide users in finding commonalities between schema elements. New models are generally developed using a combination of methods to improve the effectiveness of schema matching results. Our previous research has developed a prototype of hybrid schema matching utilizing a combination of constraints-based method and an instance-based method. The innovation of this paper presents a mathematical formulation of a hybrid schema matching model so it can be run for different cases and becomes the basis of development to improve the effectiveness of output and or efficiency during schema matching process. The developed mathematical model serves to perform the main task in the schema matching process that matches the similarity between attributes, calculates the similarity value of the attribute pair, and specifies the matching attribute pair. Based on the test results, a hybrid schema matching model is more effective than the constraints-based method or instance-based method run individually. The more matching criteria used in the schema matching provide better mapping results. The model developed is limited to schema matching processes in the relational model database.**

Keywords*—Constraint-based; hybrid schema matching model; instance-based; mathematical model

I. Introduction

Schema matching is a crucial issue in applications that involve multiple databases from heterogeneous sources, e.g., for query mediation and data warehouse [1]. The problem has emerged since the early 1980s [2]. Technically schema matching is a process of database integration that results in generalization or specialization [3]. The method of database integration generally faces the constraints caused by heterogeneity problems [4]. Database integration solutions consist of 3 levels, namely database, middleware, and applications [5]. The task of schema matching is limited to detecting the similarities and relationships between the elements of two schemas [6]. An example of an integration effort at the database level for a relational database schema is performed on [7] while using ontology is found in [8] and [9]. Integration at the middleware level is developed [10], while integration at the application level through business process integration is among others performed by [11] and [12].

Studies at [13] have at least found 36 schema matching models ever developed before. The schema matching prototype first appeared on SemInt [14], while the latest is COMA 3.0 [15]. Schema matching evolved from manual to semi-automatic ways. Until the end of the 2002 year, the schema matching model has been still done mainly by using manual methods [16], only a small developed model for the most familiar domain and suitable for applications with different schema languages [17]. The manual approach has disadvantages that are time-consuming, tedious, and impractical when it is applied, which involve many schemas [16]. The manual process is also expensive and error-prone. So, it needed new methods that are semi-automated to reduce manual effort [18]. The goal is to expertly guide users in solving schema matching problems [17]. The issue of schema matching is how to arrange a mapping between two elements of schema or ontology that have in common. The mapping process involves two schemas or ontology, one of which acts as a source and the other as a target. The schema matching model cannot be fully automated because it still encounters conflict problems at the naming level and data abstraction before it can be generalized for database integration [3].

According to [19], schema matching is a work similar to pairing, whereas according to [18], [20], and [21] schema matching is a process for finding the relation between elements in two schemas. The purpose of schema matching is given two schema input and or additional information, then determining the result of mapping the similarity of schema elements entered after verification by a user [22]. Generally, schema matching requires knowledge that is not always available in the schema so that the process cannot be performed automatically and requires user interaction to verify or provide suggestions for the model results [23].

Schema matching models can be developed using one or a combination of methods. The methods of schema matching are classifying in different ways, e.g. [7], [24], and [25] distinguishing into seven categories, i.e., linguistic-based, structure-based, constraint-based, instance-based, rule-based, hybrid, and auxiliary information dictionary, WordNet, or Corpus. In the case of schema matching performed using more than one method, the way of combining is doing using a hybrid or composite. The hybrid model runs several methods simultaneously [26], [27], while the composite runs the ways independently on each schema matched and combines on the result [28]. The term hybrid matcher is synonymous with intra-matcher parallelism, composite matcher equivalent with inter-matcher parallelism, while hybrid schema matching is also known as a mixed strategy [15], [29].

Based on survey results [13], research [7] developed a hybrid schema matching model by combining simultaneously constraints-based method approaches and instance-based methods. The model was tested 36 times and yielded the interval parameter values Precision (P) between 71.43% - 100.00%, Recall I at 75.00% -100.00% intervals, and F-measure (F) at intervals 81.48% -100.00 %. The focus on [7] is to develop a logical model and operational architecture for the hybrid schema matching model. The novelty of this paper is to present a mathematical formulation for the hybrid schema matching model [7], so it can be run for different cases and the basis of development to improve the effectiveness of output and or efficiency of schema matching process. The developed mathematical model serves to perform the primary task in the schema matching process that matches the similarity between attributes, calculates the similarity value of the attribute pair, and specifies the matching attribute pair.

The remainder of this contribution is structured as follows: Section 2 describes the material and method which is already applied in several real-world relational databases. Section 3 discusses the results and discussion of the results of the experiment, and finally, Section 4 contains the conclusion and opportunities for further research.

II. The Material and Method

The process in schema matching requires two types of input, i.e., database pairs to match and user verification of the mapping output generated by the model. In this research, the test data consist of a simulation database and test database. Both were developed using a relational database model. The input database is called DBSource (database matched) and DBTarget (database for matching reference).

The simulation database is prepared specially by the researcher for logical model validity testing. Constraints and instances in the simulation database are arranged to vary in a controlled manner so that when used for testing, it can display possible errors. The simulation database consists of 4 db_loc1, db_loc2, db_loc3, and db_loc4, each containing the master data of the provincial, district, and sub-district codes used in e-government applications in Indonesia and developed using MySQL. Each of the simulation databases is composed of 3 tables, eight attributes, and 9.953 instances. The test database is the result of a survey that meets the various criteria on aspects of DBMS, application domain, and size, used as many as 30 pieces. Based on the DBMS used, the test database consists of 22 databases developed using MySQL and eight others developed using MS Access. Based on the application domain, it consists of 8 college educational applications, 12 high school academic applications, eight e-government applications, and two e-commerce applications. By size, the most significant test database composed of 204 tables, the largest attribute count is 1,851, the largest instance count is 232,893 items, and the largest capacity is 79,769 Kb while the smallest test database contains 1 table, with 16 attributes, 480 instances, measuring 115 Kb.

The process sequence in our model is:

1) accept DBSource and DBTarget input;

2) extracting constraints and instances;

3) perform matching and computation of similarity values (SIM) in the attribute pair in DBSource and DBTarget;

4) specify the matching attribute pairs;

5) displaying the initial result of mapping the similarity of the attribute pair for the DBSource and DBTarget pair;

6) receive user verification of the initial finding of mapping the similarity of the attribute pair;

7) displaying the final result of mapping the similarity of the attribute pair that has been verified by the user; and calculate and display the values of effectiveness parameters, i.e., P, R, and F.

The output generated by the model includes:

1) general information about DBSource and DBTarget, including constraint and instance;

2) attribute values (SIM) of each pair of attributes in DBSource and DBTarget;

3) initial results mapping the similarity of the attribute pair as an output model;

4) the result of the mapping of the similarity of attribute pair which has been verified by the user; and

5) the effectiveness parameter values P, R, and F.

The values of Precision (P), Recall (R), and F-measure (F) are calculated using the following formula [30]:

$$P = \frac{TP}{TP+FP} \quad (1)$$

$$R = \frac{TP}{TP+FN} \quad (2)$$

$$F = \frac{2*P*R}{P+R} \quad (3)$$

The TP represents the relevant model output, and the user accepts it. TN is outward of the relevant model, and the user does not accept it. The FP is an outsource of the model that is irrelevant and accepted by the user, and the FN is the outward model that is irrelevant and is not accepted by the user.

Precision reflects the share of real correspondences among all found ones, Recall specifies the share of real correspondences that found, and F-measure represents the harmonic mean of Precision and Recall and is the most common variant of F-measure in Information Retrieval.

Overall this research consists of 7 stages, namely, 1) development of logical model, 2) development of model architecture, 3) development of procedures, 4) development of prototype model, 5) testing the validity of the model using the database simulation, 6) the development of mathematical models, and 7) testing model using the test database. Stage 1-5 has been performed on [7], and the focus of this paper is to present the mathematical model (steps 6) and show the comparison of the effectiveness of the hybrid schema matching model over the constraint-based method and the instance-based method performed separately (stage 7). The matching mechanism and the similarity value calculation of the attribute pair in the hybrid schema matching model shown in Fig. 1.

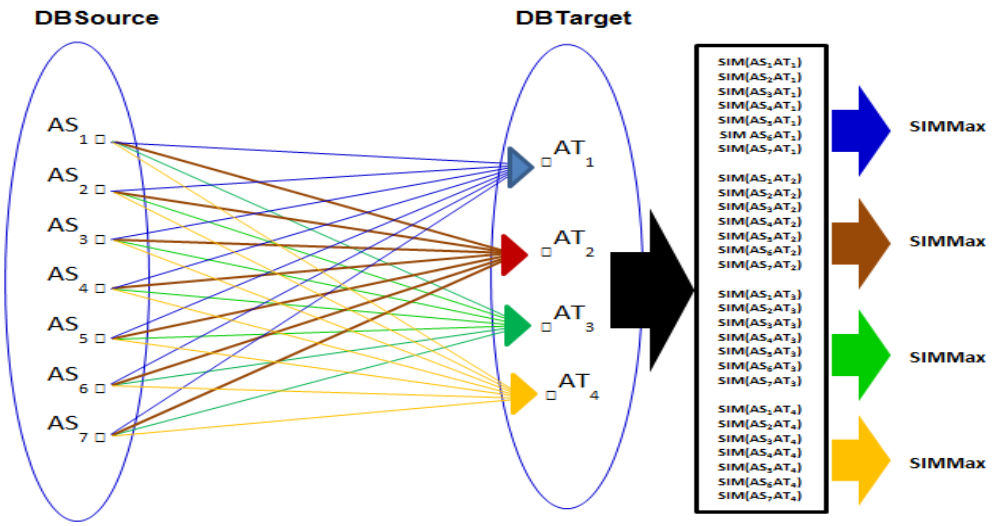


Fig. 1. Matching Mechanism and Computing the Similarity Value.

III. Result and Discussion

A. A Sample of Hybrid Schema Matching

To provide an overview process of the hybrid schema matching model, the following given a simple example of schema matching process for the employee (Table I and Table II) and religion (Table III and Table IV).

Table I to Table IV show the constraints and instances of employee and religion. Each possible attribute pair on employee and religion is then matched and calculated by the value of the similarity to determine the matching pair. Then it is verified to ascertain whether the pair that is declared to be matched by the model strictly matches, or must be revised, or otherwise unpaired. In this example, the calculation of the similarity value of each pair of attributes is based on the assumption that each matching criterion has the same weight so that each value is 0.166. The matching criteria on the constraint are of type, width, nullable, unique, and domain, whereas in an instance the attribute pair is the same if found at least one value of the same value. In each matching process using a specific criterion, if found similarity, then the value of similarity on the criterion is 0.166, otherwise the value of the similarity is 0.000. The attribute pair SIM value calculates by summing the similarity values across all criteria. Based on the SIM attribute pair values, then determined the attribute pair that is declared suitable that has the highest value or 1.00.

TABLE I. The Table Structure of Employee

Column	Type	Width	Null	Unique	Domain
employee_id	char	3	no	yes	-
employee_name	varchar	100	no	no	-
religion_id	char	1	no	no	-

TABLE II. The Instance of Employee

employee_id	employee_name	religion_id
emp_01	Edhy Sutanta	M
emp_02	Erna Kumalasari N	M
emp_03	Rosalia Arum K	C

TABLE III. The Table Structure of Religion

Column	Type	Width	Null	Unique	Domain
religion_id	Char	1	no	yes	-
religion_name	varchar	10	no	yes	-

TABLE IV. Instance of Religion

employee_id	religion_name
emp_01	Moslem
emp_02	Christian

There are three attributes on employee and two attributes in religion, so the matching process will doing six times. The matching pair of attributes is employee_id in employee and religion_id in religion, employee_name in employee and religion_id in religion, religion_id in employee and religion_id in religion, employee_id in employee and religion_name in religion, employee_name in employee and religion_name in religion. The result obtained three pairs of attributes that are otherwise suitable, is religion_id in employee match with religion_id in religion on SIM = 0.664, employee_id in employee match with religion_name in religion on SIM = 0.332, and employee_name in employee match with religion_name in religion on SIM = 0.332. Then, when verification, only one pair of attributes are found to be suitable is religion_id in employee match with religion_id in religion. The effectiveness parameter values obtained are P = 1.00, R = 0.25, and F = 0.40.

B. The Hybrid Schema Matching Logical Test

Testing the validity of the model is done to ensure that the logical model, model architecture, procedures, and prototypes developed are valid logically. Testing is done in 2 ways, manually and using a software of model prototype. The test was conducted 16 times using a combination of 4 databases simulated as source and target. The values of TP, FP, FN, and TN, and P, R, and F in each test using the developed and manually model are the same as the results obtained in the test using the model prototype so that the developed model is

validly valid logically. The end result obtained the average value of P = 94.42%, R = 100.00%, and F = 97.02%.

C. The Hybrid Schema Matching Logical Test

The description of the mathematical model for our hybrid matching schema is as follows:

$DS = \{RS_1, RS_2, .. RS_m\}$, RS_i is a table in DS

$RS_i = (AS_{i1}, AS_{i2}, .., AS_{ik_I})$, $1 \le I \le n$, representation of Rsi as attribute pair

$tS_{ij} = (vS_{i1j}, vS_{i2j}, .., vS_{ik_Ij})$, $1 \le j \le l_i$, tuple to-j in Rsi

where

$tS_{ij} \in r(RS_i) \subset DOM(AS_{i1})_x DOM(AS_{i2})_x .. _x DOM(AS_{ik_I})$

$DT = \{RT_1, RT_2, .. RT_n\}$, RT_i is a table in DT

$RT_p = (AT_{p1}, AT_{p2}, .., AT_{pr_p})$, $1 \le p \le m$, representation of Rti as attribute pair

$tT_{pq} = (vT_{p1q}, vT_{p2q}, .., vT_{pr_pq})$, $1 \le q \le s_p$, tuple to-q in Rti

where

$tT_{pq} \in r(RT_p) \subset DOM(AT_{p1})_x DOM(AT_{p2})_x .. _x DOM(AT_{pr_p})$

DS declares the source database to be matched (DBSource). RS is a relation in DS, and AS is an attribute in the RS. tS is a tuple in RS. vS is the data value in tS. M is the relation cache in DS. DT is the reference target database for matching (DBTarget). RT is a relation in the DT, and AT is an attribute in the RT. tT is a tuple in RT. vT is the value of data in tT, and n is a relation count in DT. While a DMATCH is representing the result of schema matching for the pair of DS and DT, where x is a cache attribute in DS, and y is a cache attribute in DT, then;

$DMATCH = \{(AS_1, AT_1, AS_1, AT_2), .. (AS_x, AT_y)\}$

The C represents the set of the match criteria;

C = {T, W, N, U, D, I}

where the T, W, N, U, D, and I represent the type, the width, the nullable, the unique, the domain, and the instance, respectively.

The similarity value for any pair of attributes in the DS and the b attribute in the DT value is calculated as follows:

$$SIMT(AS_a, AT_b) = \begin{cases} 1, T(AS_a) = T(AT_b) \\ 0, others \end{cases} \quad (4)$$

$$SIMW(AS_a, AT_b) = \begin{cases} 1, W(AS_a) = W(AT_b) \\ 0, others \end{cases} \quad (5)$$

$$SIMN(AS_a, AT_b) = \begin{cases} 1, N(AS_a) = N(AT_b) \\ 0, others \end{cases} \quad (6)$$

$$SIMU(AS_a, AT_b) = \begin{cases} 1, U(AS_a) = U(AT_b) \\ 0, others \end{cases} \quad (7)$$

$$SIMD(AS_a, AT_b) = \begin{cases} 1, D(AS_a) = D(AT_b) \\ 0, others \end{cases} \quad (8)$$

$$SIMI(AS_a, AT_b) = \begin{cases} 1, \exists\, I(AS_a) = I(AT_b) \\ 0, others \end{cases} \quad (9)$$

Where the SIMT is the value of similarity for the T criteria. The SIMI is the value of similarity for I criteria. The SIMW is the value of similarity for W criteria. The SIMN is the value of similarity for N criteria. The SIMU is the value of similarity for U criteria. The SIMD is the value of similarity for D criteria, and the SIMI is the value of similarity for I criteria. In general, the calculation of the similarity value of the attribute pair for any C member is:

$$SIM_c(AS_a, AT_b) = \begin{cases} 1, x(AS_a) = x(AT_b) \\ 0, others \end{cases} \quad (10)$$

where x(A)= criteria x for A.

While WI is the weight of the instance, WT is the weight of type, WW is the width weight, WN is the nullable weight, WU is a specific weight, and WD is the domain weight. Thus, the AS_a and AT_b pair similarity values are calculated as follows:

$$SIM(AS_a, AT_b) = \sum_{y \in C} SIM_y (AS_a, AT_b)\, W_y \quad (11)$$

Paired attributes that are otherwise matched by the model, if the AT_b matched to AS_a then AT_b are taken which meets the following conditions:

$$SIM(AS_a, AT_b) = \max_{z=1}^{m} SIM(AS_a, AT_z) \quad (12)$$

where

z = 1, 2, .., m.

D. The Hybrid Schema Matching Test Result

A valid hybrid schema matching model, then tested 32 times using a combination of randomly chosen test database pairs, and the results displayed in Table V.

Based on the test results in Table V and Table VI, it is known the highest P-value is 93.04% in the hybrid model, the lowest P-value is 33.28% in the constraint-based method, while the instance-based method is between constraint-based and hybrid. These results show that hybrid models provide the best results from the precision. Compared to the instance-based method, there was an increase of 31.94%, while compared to the constraint-based approach, there was an increase of 59.76%. This condition occurs because the hybrid model matches by involving more criteria (6 criteria at once), compared to the constraint-based method using five criteria or the instance-based method using one criterion. The more criteria applied for matching between attributes, and the results increase the value of the parameter P because the pair declared matched by the model have a higher chance of being accepted by the user as the right pair. These results indicate that the similarity of instances can be the basis for finding the matching attribute pair with an average P of 61.10%. Lowest P occurs in the constraint-based method, which is 33.28%. These results show that, even if the matching attribute pair has the same constraint, the result is still lower than the hybrid model and the instance-based method individually.

Based on the experiments on each model, effectiveness obtained, as shown in Table VI.

The highest R-value is 100.00% that is in the constraint-based method and the instance-based method, and both are equal. The lowest R-value is 99.83%, in the hybrid model. The value of R = 100.00% indicates that all attribute pairs declared matched by the model (in the initial result) are perfectly matched and accepted by the user (TP = N), and the attribute pair stated unmatched by the model (in the initial result) is received by the user (FN = 0).

TABLE V. THE RESULT OF HYBRID SCHEMA MATCHING

DBSource	DBTarget	Test Result (%)		
		P	R	F
sipt_admision	sipt_admision	89.90	100.00	94.68
sipt_admision	sipt_academic	85.11	100.00	91.95
sipt_academic	sipt_payroll	85.42	98.80	91.62
sipt_academic	sipt_employee	83.08	96.43	89.26
sipt_academic	sipt_tax_pph	91.37	99.45	95.24
sipt_academic	sipt_workshop	82.95	100.00	90.68
sipt_academic	sipt_library	88.46	100.00	93.88
sipt_academic	sipt_user	92.98	100.00	96.36
egov_dptkp	license	94.64	100.00	97.25
egov_dptkp	license_oln	88.48	100.00	93.89
egov_dptkp	egov_dptbgcpt	100.00	100.00	100.00
egov_dptkp	quickcount_bgcpt	94.01	100.00	96.91
egov_dptkp	egov_dptbtl	100.00	100.00	100.00
egov_dptkp	egov_dptkp	100.00	100.00	100.00
egov_dptkdy	ecomm_rsmitra	94.12	100.00	96.97
egov_dptkdy	ecomm_motorcred	90.55	100.00	95.04
nuptk	nuptk	99.41	100.00	99.71
nuptk	hs_sinisa	89.37	100.00	94.39
nuptk	hs_sipp	93.15	100.00	96.45
nuptk	hs_psb	98.01	100.00	99.00
hs_sipp	hs_sinisa	99.69	99.97	99.83
hs_sipp	hs_sipp	99.35	100.00	99.68
hs_sipp	hs_psb	93.73	100.00	96.76
hs_sipp	hs_grade	93.96	100.00	96.89
hs_sipp	hsgrade_ol	90.45	100.00	94.99
hs_sipp	hs_report	98.76	100.00	99.38
hs_sipp	hs_hspwt	94.49	99.98	97.16
hs_sipp	hs_forum	90.06	100.00	94.77
hs_sipp	hs_announcement	87.52	100.00	93.34
hs_sipp	hs_webinfo	98.24	100.00	99.11
hs_sipp	hs_osis	91.05	100.00	95.32
hs_sipp	hs_elearning	98.99	100.00	99.49
Average:		93.04	99.83	96.25

TABLE VI. THE RESULT OF HYBRID SCHEMA MATCHING

Method	Test Result Average (%)		
	P	R	F
Hybrid (constraint-based and instance-based)	93.04	99.83	96.25
Constraint-based method	33.28	100.00	38.60
Instance-based method	61.10	100.00	69.87

The highest F value is 96.25%, i.e. in the hybrid model, while the lowest is 38.60% in the constraint-based method. Value F = 96.25% indicates that the estimated level of effort required to add FN and eliminate FP has reached the best condition. The value of F in the hybrid model, compared with the instance-based method has increased by 26.38%, and when compared with the constraint-based method, there is an increase of 57.65%. This increase occurs because hybrid models perform matching by involving more criteria that are using six criteria at once, so the attribute pair declared fit by the model has a higher chance of being accepted by the user as a matching pair.

The experimental results in this study indicate that the mathematical model for hybrid schema matching has functioned as expected and provided better effectiveness than the original constraint-based method and instance-based method. The model that we have developed is still likely to be further investigated, at least to improve the effectiveness of output, process efficiency, and modification so that the model can be applied to non-relational databases.

This hybrid schema matching model still contains problems related to the effectiveness of the outcome because the constraint matching criteria and instance are assumed to have equal weight when calculating the similarity values (SIM) of each pair of attributes. Each criterion on the constraint is also considered to have the same weight. These criteria can have different weights in determining the value of similarity (SIM); one of the weighting ideas ever done [31]. The results of the survey at the time of collecting the test database also found the fact that the database designer has the freedom to make the database schema definition, including determining the data size (width) in the string data type. In this test, the string attribute pair is declared the same if it has the same size; it has not yet accommodated the freedom of the database designer.

Another problem is related to process efficiency. The developed model requires repetition of matching steps and simulated (SIM) value calculations on all possible attribute pairs. When schema matching performed on a database pair it involves many foreign keys, the model still encounters efficiency problems in matching and simulated values (SIM) and user verification steps that can only be done manually.

IV. CONCLUSIONS

The main contribution of this paper is to present the proposed mathematical model for hybrid schema matching. The model is developed based on a combination of two methods, namely the constraints-based method and an instance-based method. The model has been tested using a relational database, and the results are more effective than the original constraint-based method or instance-based method. Our research further refines the model by adding features to improve output effectiveness and process efficiency. Also, the model developed to run in non-relational database formats.

ACKNOWLEDGMENT

The authors would like to thanks the Directorate of Research and Community Service, Directorate General of Research and Development Strengthening at the Ministry of

Research, Technology and Higher Education of the Republic of Indonesia under Fundamental Research Grant Contract No.: 109/SP2H/LT/DRPM/2018.

REFERENCES

[1] L. A. P. P Leme, M. A. Casanova, K. K. Breitman, and A. L Furtado, "OWL Schema Matching," Journal of the Brazilian Computer Society, 16(5), pp. 21-34, April 2010, DOI: 10.1007/s13173-010-0005-3.

[2] I. F. Cruz, F. P. Antonelli, and C. Stroe, "AgreementMaker: Efficient Matching for Large Real-World Schemas and Ontologies (demo paper)," in International Conference on Very Large Data Bases (VLDB), 2009, DOI: 10.14778/1687553.1687598.

[3] C. Kavitha, G. S. Sadasivam, and S. N. Shenoy, "Ontology-Based Semantic Integration of Heterogeneous Databases," European Journal of Scientific Research, 64(1), pp. 115-122, November 2011.

[4] S. Hamill, M. Dixon, B. J. Read, and J. R. Kalmus, "Interoperating Database Systems: Issues and Architectures," Council for the Central Laboratory of The Research Councils, United Kingdom, Technical Report 1997.

[5] Ministry of Communication and Information Republic of Indonesia, "SIFONAS Sebagai Tulang Punggung e-Governance (SIFONAS as The Backbone of e-Governance)," 2002.

[6] P. Martinek, "Schema Matching Methodologies and Runtime Solutions in SOA Based Enterprise Application Integration," Department of Electronics Technology, Faculty of Electrical Engineering & Informatics, Budapest University of Technology and Economics, Hungary, Ph.D. Thesis, 2009.

[7] E. Sutanta, R. Wardoyo, K. Mustofa, and E. Winarko, "A Hybrid Model Schema Matching using Constraint-Based and Instance-Based," International Journal of Electrical and Computer Engineering (IJECE), 6(3), pp.1048-1058, June 2016, DOI: 10.11591/ijece.v6i3.pp 1048-1058.

[8] H. Alani and S. Saad, "Schema Matching for Large-Scale Data Based on Ontology Clustering Method," International Journal on Advanced Science, Engineering and Information Technology (IJASEIT), 7(5), pp. 1790-1797, 2017, DOI: 10.18517/ijaseit.7.5.2133.

[9] K. H. Shafa'amri and J. O. Atoum, "A Framework for Improving the Performance of Ontology Matching Techniques in Semantic Web," International Journal of Advanced Computer Science and Applications (IJACSA), 3(1), pp. 8-14, 2012.

[10] N. H. Azizul, A. M. Zin, and E. Sundararajan, "The Design and Implementation of Middleware for Application Development within Honeybee Computing Environment," International Journal on Advanced Science, Engineering and Information Technology (IJASEIT), 6(6), pp. 937-943, 2016, DOI: 10.18517/ijaseit.6.6.1415.

[11] D. Sulisworo, Tawar, and U. Ahdiani, "ICT Based Information Flows and Supply Chain in Integrating Academic Business Process," International Journal on Advanced Science, Engineering and Information Technology (IJASEIT), 2(6), pp. 44-48, 2012, DOI: 10.18517/ijaseit.2.6.243.

[12] M. Mohammadi and Muriati Mukhtar, "Business Process Modelling Languages in Designing Integrated Information Systems for Supply Chain Management," International Journal on Advanced Science, Engineering and Information Technology (IJASEIT), 2(6), pp. 464-467, 2012, DOI: 10.18517/ijaseit.2.6.245.

[13] E. Sutanta, R. Wardoyo, K. Mustofa, and E. Winarko, "Survey: Models and Prototypes of Schema Matching," International Journal of Electrical and Computer Engineering (IJECE), 6(3), pp. 1011-1022, June 2016, DOI: 10.11591/ijece.v6i3.pp1011-1022.

[14] W. S. Li and C. Clifton, "Semantic Integration in Heterogeneous Databases Using Neural Networks," in The 20th International Conference on Very Large Data Bases (VLDB), Santiago de Chile, Chile, 1994, pp. 1-12.

[15] E. Rahm, "Schema Matching and Mapping: Towards Large-Scale Schema and Ontology Matching," in Data Centric Systems and Applications, Z. Bellahsene, A. Bonifati, and E. Rahm, Eds. New York, USA: Springer, 2011, pp. 3-27, DOI: 10.1007/978-3-642-16518-4_1.

[16] H. H. Do, S. Melnik, and E. Rahm, "Comparison of Schema Matching Evaluations," in Proceedings of The 2nd International Workshop Web and Databases, In Lecture Notes In Computer Science (LNCS) 2593, Springer-Verlag, Germany, 2002, pp. 221-237, DOI: 10.1007/3-540-36560-5_17.

[17] H. H. Do, "Schema Matching and Mapping-Based Data Integration," Interdisciplinary Center for Bioinformatics and Department of Computer Science, University of Leipzig, Leipzig, Germany, Ph.D. Thesis, 2005.

[18] D. Engmann and S. Massmann, "Instance Matching with COMA++," in Datenbank Systeme in Business, Technologie und Web (BTW Workshop): Model Management and Metadata, Aachen, Germany, 2007, pp. 28-37, https://dbs.uni-leipzig.de/file/BTW-Workshop_2007_EngmannMassmann.pdf.

[19] J. Li, J. Tang, Y. Li, and Q. Luo, "RiMOM: A Dynamic Multistrategy Ontology Alignment Framework," IEEE Transaction Knowledge Data Engineering, 21(8), pp. 1218-1232, August 2009, DOI: 1109/TKDE .2008.202.

[20] P. A. Bernstein, B. Harry, P. Sanders, D. Shutt, and J. Zander, "The Microsoft Repository," in Proceedings of The 23rd International Conference Very Large Data Bases (VLDB), Athens, Greece, 1997, pp. 3-12.

[21] A. Stabenau, G. McVicker, C. Melsopp, G. Proctor, M. Clamp, and E. Birney, "An Overview of ENSEMBL," Genome Research Journal, 14(5), pp. 929-933, May 2004, DOI: 10.1101/gr.1860604.

[22] J. Madhavan, P. A. Bernstein, and E. Rahm, "Generic Schema Matching with CUPID," in Proceedings of The 27th International Conference on Very Large Data Bases (VLDB), Roma, Italy, 2001, pp. 49-58, http://dl.acm.org/citation.cfm?id=645927.67219.

[23] S. Melnik, E. Rahm, and P. Bernstein, "RONDO: A Programming Platform for Generic Model Management," in Proceedings of The ACM-SIGMOD Conference on Management of Data (SIGMOD), San Diego, California, USA, 2003, pp. 193-204, DOI: 10.1145/872757.872782.

[24] P. A. Bernstein, M. Jayant, and E. Rahm, "Generic Schema Matching, Ten Years Later," in The 33 International Conference on VLDB Endowment, Seattle, Washington, 2011, pp. 695-701, http://www.vldb.org/pvldb/vol4/p695-bernstein_madhavan_rahm.pdf.

[25] A. A. Alwan, A. Nordin, M. Alzeber, and A. Z. Abualkishik, "A Survey of Schema Matching Research using Database Schemas and Instances," International Journal of Advanced Computer Science and Applications (IJACSA), 8(10), pp. 102-111, 2017.

[26] S. Bergamaschi, S. Castano, and M. Vincini, "Semantic Integration of Semistructured and Structured Data Sources," SIGMOD Record, 28(1), pp. 54-59, March 1999, DOI: 10.1145/309844.309897.

[27] T. Milo and S. Zohar, "Using Schema Matching to Simplify Heterogeneous Data Translation," in Proceedings of The 24th International Conference on Very Large Data bases (VLDB), New York, USA,1998, pp.122-133, http://dl.acm.org/citation.cfm?id=645924.671326.

[28] A. H. Doan, P. Domingos, and A. Y. Halevy, "Reconciling Schemas of Disparate Data Sources-A Machine-Learning Approach," in Proceedings of The ACM SIGMOD International Conference Management of Data, Santa Barbara, California, 2001, pp. 509-520, DOI: 10.1145/376284.375731.

[29] A. Gross, M. Hartung, T. Kirsten, and E. Rahm, "On Matching Large Life Science Ontologies in Parallel," in Proceedings of The 7th International Conference Data Integration in the Life Sciences (DILS), Gothenburg, Sweden, 2010, pp. 35-49, http://dl.acm.org/citation.cfm?id= 1884477.1884483.

[30] Y. Karasneh, H. Ibrahim, M. Othman, and R. Yaakob, "An Approach for Matching Relational Database Schemas," Journal of Digital Information Management, 8(4), pp. 260-269, 2010, https://dblp.org/rec/bib/journals/jdim/KarasnehIOY10.

[31] M. B. Shuaibu, "Determining an Appropriate Weight Attribute in Fraud Call Rate Data Using Case-Based Reasoning," International Journal on Advanced Science, Engineering and Information Technology (IJASEIT), 4(1), pp. 34-36, 2014, DOI: 10.18517/ijaseit.4.1.357.