

KAJIAN MODEL DAN PROTOTYPE SCHEMA MATCHING

(Studi untuk Menemukan Peluang Pengembangan Model dan Prototipe Baru)

Edhy Sutanta

Mahasiswa S3 Jurusan Ilmu Komputer & Elektronika
Fakultas Matematika & IPA, UGM
Yogyakarta, Indonesia
e-mail: edhy_sst@yahoo.com

Retantyo Wardoyo

Jurusan Ilmu Komputer & Elektronika
Fakultas Matematika & IPA, UGM
Yogyakarta, Indonesia
e-mail: rw@ugm.ac.id

Khabib Mustofa

Jurusan Ilmu Komputer & Elektronika
Fakultas Matematika & IPA, UGM
Yogyakarta, Indonesia
e-mail: khabib@ugm.ac.id

Edi Winarko

Jurusan Ilmu Komputer & Elektronika
Fakultas Matematika & IPA, UGM
Yogyakarta, Indonesia
e-mail: edwin@ugm.ac.id

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

Keywords—*model, hybrid model, prototype, schema matching.*

I. PENDAHULUAN

Isu tentang *schema matching* telah muncul sejak awal tahun 1980-an [1], merupakan permasalahan mendasar dalam banyak aplikasi untuk integrasi data, yaitu bagaimana menyusun pemetaan antar dua elemen *schema* atau ontologi yang memiliki kesamaan. *Schema matching* merupakan permasalahan penting untuk integrasi informasi dari berbagai sumber yang heterogen [2], untuk mewujudkan interoperabilitas, dan melaksanakan integrasi data dari berbagai bidang [3], dan juga penting dalam evolusi *schema*, serta penggunaan kembali *software* [4]. *Schema matching* sebagai bagian dari topik *Enterprise Application Integration* khususnya *Enterprise Information Integration*, merupakan tugas integrasi pada level rendah untuk mengatasi permasalahan yang diakibatkan oleh *schematic heterogeneity* dan dilaksanakan oleh level *back end* [5], yaitu menyusun pemetaan dan pencocokan antar *schema* dalam sistem aplikasi satu dengan aplikasi lainnya [6]. *Schematic heterogeneity* adalah perbedaan penamaan dalam definisi *schema*, meliputi tipe, format, dan presisi data [7].

Perkembangan model *schema matching* hingga akhir tahun 2002, sebagian besar masih dilakukan secara manual [8], hanya sebagian kecil yang dikembangkan untuk domain yang lebih umum dan sesuai untuk aplikasi dan bahasa *schema* yang berbeda [5]. Kelemahan model manual adalah membutuhkan waktu lama, membosankan, dan tidak praktis jika diterapkan dalam kasus yang melibatkan banyak *schema* [8]. Model manual juga mahal dan kemungkinan kesalahan relatif besar, sehingga dibutuhkan metode baru yang bersifat semi otomatis [6] untuk mengarahkan secara efektif kepada para pengguna dalam memecahkan masalah *schema matching* [3]. *Schema matching* merupakan obyek penelitian yang menarik [3], dan masih terus terbuka untuk menemukan cara-cara cerdas dalam mengembangkan model maupun perangkat lunak dengan menggabungkan metode yang sudah ada saat ini [9].

Makalah ini merupakan ringkasan hasil kajian pustaka yang memaparkan kegiatan penelitian dan publikasi tentang model dan prototipe *schema matching* selama 25 tahun terakhir untuk menggambarkan perkembangan dan peluang penelitian baru.

II. KONSEP SCHEMA MATCHING

A. Definisi Schema Matching

Istilah *schema matching* telah didefinisikan oleh para peneliti/pakar dengan cara yang berbeda, namun memiliki kemiripan makna. Menurut [10], *schema matching* merupakan sebuah pekerjaan yang mirip dengan pencocokan, sementara [6], [11], [12], dan [13] mendefinisikan *schema matching* sebagai sebuah proses untuk menemukan kerelasian di antara elemen-elemen dalam dua *schema*. Tujuan *schema matching* adalah diberikan input dua *schema* yang berbeda, dan/atau informasi tambahan, dan input pemetaan *schema*, kemudian menentukan hasil pemetaan elemen *schema* yang diinputkan setelah melalui verifikasi pengguna [14]. *Schema matching*

melibatkan dua *schema* atau ontologi, salah satu sebagai *source* dan lainnya disebut *target* [1].

B. Klasifikasi Model Schema Matching

Proses dalam *schema matching* dapat melibatkan berbagai macam algoritma, misal untuk menentukan data yang akan dicocokkan, transformasi pemetaan, atau melakukan penggabungan [15]. Berdasarkan algoritma yang digunakan, model *schema matching* dapat diklasifikasikan ke dalam beberapa kategori.

Klasifikasi model *schema matching* oleh [14] dan [16] terdiri atas *schema based vs. instance based*, *element vs. structure granularity*, *linguistic based*, *constraint based*, *matching cardinality*, *auxiliary information*, serta *individual vs. combinational*. Klasifikasi model *schema matching* menurut [3] terdiri atas *schema based*, *instance based*, dan *reuse oriented*. Klasifikasi [17] membagi model berdasarkan level dan jenis informasi yang dieksplorasi meliputi elemen *schema* dan struktur *schema*, dan berdasarkan jenis informasi yang dieksplorasi meliputi aspek terminologi (melibatkan aspek linguistik (terdiri atas *language based* dan *linguistic based*) atau tidak melibatkan aspek linguistik (*string based*)), berdasarkan aspek struktural (meliputi aspek *internal* (*constraint based*) dan *relational* (terdiri atas *alignment reuse*, *graph based*, *taxonomy based*, serta *repository structure*)), dan semantik (terdiri atas *upper level formal ontology* dan *model based*). Dalam [1], model *schema matching* diklasifikasikan berdasarkan level komponen yang dicocokkan (konseptual dan struktur), tingkat intervensi pengguna (manual dan otomatis), metode yang digunakan (*stand-alone* dan gabungan), dan tipe komponen yang digunakan sebagai dasar pencocokan (menggunakan *schema* atau menggunakan *schema* dan *instance*). Klasifikasi model *schema matching* menurut [18] terdiri atas RSM (*relations schemas matcher*), ANM (*attribute relations name matcher*), DTM (*data type matcher*), CM (*constraint matcher*), dan IDM (*instance data matcher*). Dengan cara yang berbeda, [5] mengelompokkan algoritma *schema matching* menjadi tiga, yaitu *linguistic matcher* (NTA (*name, connected terms, attributes*) *linguistic matcher*, *prefix/suffix based matcher for name*, dan *prefix/suffix based matcher for types*), *vocabular matchers* (*WordNet based word matcher for names* dan NTA (*name, connected terms, attributes*) *related terms similarity*), dan *structural matchers* (*flooding similarity*, *WordNet based ancestor context similarity*, *string comparison based child context similarity*, *child context similarity*, dan *direct ancestor similarity using string comparison*). Klasifikasi model *schema matching* yang mirip dengan [14] dan [16] diberikan oleh [9], meliputi *linguistic matching*, *auxiliary information*, *instance based matching*, *structure based matching*, *constraint based matching*, *rule based matching*, dan *hybrid matching*.

Menurut [9], beberapa metode lain pernah dikembangkan termasuk yang memanfaatkan informasi tambahan, yaitu *graph matching*, *usage based matching*, *document content similarity*, dan *document link similarity*. Klasifikasi kombinasi model *schema matching* untuk *schema* berukuran besar diberikan oleh [9] yaitu strategi independen atau eksekusi berurutan atau eksekusi kombinasi, *parallel matching*, *self-*

tuning match work flow, *early search space pruning*, *partition based matching*, dan optimasi *schema*. Model *schema matching* dalam sebuah domain tertentu, yaitu *reuse based matching* dan pencocokan holistik merupakan kelompok pendekatan berbeda yang telah diusulkan [9]. Beberapa strategi yang menyertakan interaksi dan umpan balik dari pengguna dalam proses pencocokan dikelompokkan oleh [9], meliputi dukungan GUI (*graphical user interface*), pencocokan *incremental*, pencocokan *Top-k*, dan kolaboratif. Kelompok strategi lain adalah penggunaan algoritma untuk memperluas pencocokan *semantic*, misal *semantic tagging* dan *conditional tagging* [9].

Dalam referensi lain, [19] mengelompokkan model *schema matching* menjadi dua, yaitu *schema based* dan *instance based*. Model *schema based* terdiri atas model *element based* dan *structure based*. Model *element based* terdiri atas *linguistic based* dan *constraint based*, sementara level struktur dikembangkan berdasarkan *constraint based*. Model *instance based* dikembangkan berdasarkan *element level* yang terdiri atas *linguistic based*, *constraint based*, dan *learning based*.

C. Peran Pengguna dalam Schema Matching

Problem utama dalam *schema matching* adalah seringkali ditemukan penamaan yang tidak jelas dalam *schema*, kesulitan ditemukan sinonim penamaan, dan perbedaan bahasa sehingga model *schema matching* tidak mungkin menghasilkan keluaran yang 100% tepat dalam pemetaan *schema* [6]. Dengan alasan tersebut, maka *schema matching* tidak bisa sepenuhnya dilakukan secara otomatis, biasanya harus dikoreksi oleh pengguna/pakar untuk memperoleh hasil akhir yang benar [15]. Terdapat dua kasus di mana *schema matching* akan mengalami kegagalan sehingga memerlukan keterlibatan pengguna, pertama apabila elemen *schema* sumber tidak dapat dicocokkan dengan satupun elemen *schema* target menggunakan aturan yang digunakan, atau kedua, apabila elemen *schema* sumber menghasilkan beberapa elemen yang dianggap cocok dalam *schema* target dan sistem tidak dapat menentukan elemen yang cocok terbaik secara otomatis [20]. Menurut [21], umumnya pencocokan dua *schema* membutuhkan informasi yang tidak selalu ditampilkan di dalam *schema* dan tidak dapat dilakukan secara otomatis, sehingga memerlukan pengguna yang cakap untuk *me-review* dan menentukan saran pada hasil *schema matching*. Proses *schema matching* tidak akan pernah bisa dilakukan secara otomatis penuh selama belum ada model pencocokan semantik yang lengkap pada integrasi sistem informasi [22]. Alasan lain yang menyebabkan *schema matching* tidak bisa dilakukan secara otomatis karena adanya masalah konflik penamaan dan konflik level abstraksi [4].

D. Schema Matching dengan Satu atau Gabungan Metode

Model *schema matching* dapat dikembangkan secara *individual matcher* atau *combinational matchers* [14] dan [16]. Model *individual matcher* memiliki kelebihan dan kelemahan, dan hanya sesuai pada kasus tertentu, sehingga umumnya memerlukan lebih dari sebuah *matcher* yang dikombinasikan [19]. Menurut [16] dan [19], model *combinational matchers* dapat diimplementasikan secara terpisah (*composite*) atau

bersamaan (*hybrid*). Istilah *composite matchers* sinonim dengan *inter-matcher parallelism*, sedangkan *hybrid matcher* sinonim dengan *intra-matcher parallelism* [23]. Model *hybrid matcher* menggunakan beberapa kriteria pencocokan secara bersamaan [20], [24], dan [25], sedangkan *composite matchers* menjalankan algoritma secara terpisah atau independen dan mengkombinasikan pada hasilnya [26].

Menurut [16], *hybrid matcher* menggabungkan beberapa metode secara bersamaan untuk melakukan pencocokan antar elemen *schema*, dan harus memberikan hasil yang lebih baik dan perbaikan kinerja (efektivitas) daripada *individual matcher*. Secara sederhana, *hybrid matcher* menggabungkan dua metode berbeda yang diproses secara *simultaneous*, sedangkan *composite matcher* menggabungkan dua metode yang diproses secara berurutan yaitu sebuah metode akan dilaksanakan setelah metode yang lain selesai dilaksanakan.

III. PEMETAAN MODEL DAN PROTOTIPE SCHEMA MATCHING

Studi pustaka ini menemukan 34 model dan prototipe *schema matching* dalam 71 publikasi ilmiah. Model dan prototipe *schema matching* pertama muncul pada SEMINT [27], sedangkan yang terbaru adalah COMA 3.0 [28].

Masing-masing model dan prototipe dapat menggunakan input berupa *schema* model relasional (RDF/Relational Database Format), model XML (DTD/Document Type Definition atau XSD/W3C XML Schema), atau model ontologi (OWL/Web Ontology Language). Prototipe *schema matching* telah dievaluasi secara intensif oleh [3], [8], dan [29], hasilnya menunjukkan bahwa sebagian besar prototipe *schema matching* dikembangkan untuk lingkup sempit dan terbatas atau bersifat spesifik, dan sebagian dikembangkan secara khusus dengan memanfaatkan ontologi.

Dengan mengacu pada [18], [30], dan [31], penelitian [4] melakukan integrasi basisdata heterogen berdasarkan ontologi semantik, dan diujikan menggunakan basisdata akademik dengan DBMS MsAccess dan MySQL. Penemuan model untuk menghubungkan antar sumber data di *web* oleh [32] dan [33] juga merupakan area riset pengembangan di mana pendekatan hubungan semantik antar entitas diperlukan dan berlaku karena ketersediaan ontologi secara luas.

Dengan mengacu pada [28], [9] telah melakukan kajian prototipe model *schema matching* yang berkembang pada tahun 2001-2011 dan kemudian membandingkan prototipe (CUPID, COMA++, ASMOV, Falcon-AO, Rimon, AGREEMENTMAKER, dan OpenII) pada aspek arsitektur, representasi *schema*, representasi pemetaan *schema*, informasi input dan algoritma pencocokan, serta eksekusi pencocokan elemen *schema*. Menurut [9], prototipe tersebut menggunakan pendekatan yang berbeda dalam hal algoritma pencocokan antar elemen *schema*, sementara COMA++, Falcon-AO, Rimon, dan Agreement Maker merupakan prototipe yang menggabungkan tiga metode sekaligus, yaitu *linguistic based*, *structure based*, dan *instance based*. Penggunaan kamus eksternal, seperti daftar sinonim (*thesaurus*) biasanya digunakan untuk meningkatkan akurasi pencocokan *linguistic*.

Dukungan GUI terbatas telah disediakan oleh beberapa prototipe [34] (dalam [29]), dan sebagian mampu melakukan pencocokan ontologi pada dua *schema* [1], [35], dan [36]. Beberapa prototipe yang berperan dalam OAEI (*Ontology Alignment Evaluation Initiative*) meningkat signifikan dalam beberapa tahun terakhir, namun masih perlu dikembangkan untuk mengatasi permasalahan *schema matching* pada lingkup yang lebih luas [37]. Teknik canggih seperti partisi *schema*, *parallel matching*, penggunaan kembali dalam pemetaan dan kemampuan *self-tuning* (misal, pilihan pencocokan dinamis) hanya didukung sampai batas tertentu [28].

Model dan prototipe *hybrid schema matching* yang pernah dikembangkan sebelumnya adalah CLIO [38], [39], [40], [41], [42], dan pada penelitian [43]. Sedangkan SEMINT [25], [27], dan [44], LSD [26], CUPID [14], COMA [45], COMA++ [3], COMA 3.0 [28] dan [46], IMAP [47], PROTOPLASM [48], [49], [50], dan [51], FALCON-AO [6] dan [52], serta ASMOV [53] dikembangkan dengan menggunakan gabungan metode secara *composite*.

Metode *linguistic based matching* digunakan pada DIKE [54], [55], dan [56], MOMIS [57] dan [58], ONION [59], [60], dan [61], ARTEMIS [62], UNIFORM [63], WISE-INTEGRATOR [2], [64], dan [65], PROMPT [66] dan [67], RONDO [21], OLA [68], QOM [69] dan [70], S-MATCH [71], [62], dan [72], RiMOM [10], AGREEMENTMAKER [1], OPENII [36], serta [4]. DELTA [73] dan [74], SIMILARITYFLOODING/SF [75], XCLUST [76], dan penelitian [77] merupakan prototipe dan model yang mengimplementasikan metode *structure based matching*.

Penggunaan *constraint* dalam *schema matching* mengasumsikan bahwa *constraint* memiliki makna untuk menetapkan kesamaan elemen-elemen basisdata, misal, atribut A dalam tabel X bertipe karakter adalah sama dengan atribut B dalam tabel Y bertipe *text* [31]. Menurut [19], penggunaan *constraint based* merupakan salah satu kelompok model *schema matching* yang termasuk dalam level struktur, namun tidak dijelaskan lebih lanjut tentang *property* apa saja yang dieksplorasi dan termasuk sebagai *constraint*. Metode *instance based* digunakan pada model dan prototipe TRANSCM [20], AUTOPLEX [78], AUTOMATCH [79], [80], dan [81], GLUE [82] dan [83], SCM [84], serta DUMAS [85].

Auxiliary based matching berupa *dictionary*, *WordNet*, atau *Corpus* digunakan pada DIKE [54], [55] dan [56], MOMIS [57] dan [58], ONION [59], [60] dan [61], ARTEMIS [62], CUPID [14], [80] dan [81], COMA [45], XCLUST [76], UNIFORM [63], WISE-INTEGRATOR [2], [49], [64] dan [65], OLA [50], [68], S-MATCH [51], [62], [71], [72] dan [77], COMA++ [3] dan [6], OPENII [36], serta COMA 3.0 [28].

Statistik penelitian, model, dan prototipe *schema matching* yang berkembang dalam 25 tahun terakhir berdasarkan jenis input, metode yang digunakan (berdasarkan klasifikasi [9] ditampilkan pada Tabel 1. Berdasarkan Tabel 1, tampak bahwa metode *linguistic based* (76%) paling banyak diterapkan dalam model *schema matching*, diikuti oleh pengguna gabungan metode secara *composite* (73%), penggunaan *auxiliary based* (56%), *structure based* dan *instance based* (49%), *constraint*

based (20%), dan penggunaan gabungan metode secara *hibryd* paling sedikit (13%). Pemanfaatan metode *linguistic based* relatif paling pesat, hal ini berkembang seiring perkembangan ilmu dan ketersediaan ontologi yang semakin luas, sekalipun demikian eksplorasi pada metode *structure based* dan *instance based* masih terus diteliti oleh para peneliti.

Berdasarkan jenis masukan yang digunakan, model data OWL paling banyak dieksplorasi, diikuti oleh RDF, dan kemudian DTD/XSD. Berdasarkan domain aplikasinya, model *schema matching* berkembang menuju ke arah domain yang lebih umum dalam 10-15 tahun terakhir. Hal ini didukung oleh kebutuhan riil bahwa *schema matching* semakin dibutuhkan dalam banyak aplikasi, sehingga model yang dikembangkan dituntut untuk mampu digunakan dalam domain yang lebih luas. Penyediaan fitur GUI juga semakin berkembang juga semakin berkembang pada model dan prototipe yang dikembangkan dalam 10-15 tahun terakhir. Kondisi ini dimungkinkan karena adanya dukungan bahasa pemrograman yang memudahkan pengembangan aplikasi berbasis GUI.

Menurut [16], *hybrid matcher* menggabungkan beberapa metode secara bersamaan untuk melakukan pencocokan antar elemen *schema*, dan *hybrid matcher* harus memberikan hasil yang lebih baik dan perbaikan kinerja (efektivitas) daripada *individual matcher*. Pada sisi lain, penggunaan kombinasi metode secara *hybrid* masih relatif jarang dilakukan oleh peneliti, baru ada 9 penelitian dalam 25 tahun terakhir.

Dengan mengacu pada klasifikasi metode *schema matching* dalam [9], pengembangan *hybrid matcher* masih dapat dilakukan menggunakan variasi gabungan dua atau lebih metode *linguistic based*, *structure based*, *constraint based*, *instance based*, dan *auxiliary based* (penggunaan kamus/WordNet/Corpus). Hal ini menunjukkan adanya peluang penelitian dan pengembangan menggunakan *hybrid matcher*. Penelitian terbaru, saat ini penulis sedang mengembangkan sebuah model *hybrid schema matching* (sebutan yang diberikan oleh [16] dan [19]) atau *mixed strategy* (dalam klasifikasi [23] dan [28]) yang menggabungkan metode *constraint based* dan *instance based* (dalam klasifikasi [9] dan [19]), HybridScheMatch melibatkan DTM (*data type matcher*), CM (*constraint matcher*), dan IDM (*instance data matcher*) (dalam klasifikasi [18]).

Solusi *schema matching* ke depan juga dapat dikembangkan untuk meningkatkan konvergensi dan pendekatan resolusi pada entitas, misal pencocokan pada level *metadata* dan *instance* untuk mengidentifikasi hubungan semantik pada entitas atau *instance* [9].

TABEL 1. RINGKASAN PENELITIAN, MODEL, & PROTOTIPE SCHEMA MATCHING DALAM 25 TAHUN TERAKHIR

Tahun	Jml Publikasi	Input			Metode Schema Matching yang Digunakan						
		RDF	DTD / XSD	OWL	Linguistic Based	Structure Based	Constraint Based	Instance Based	Kombinasi Metode		Auxiliary based (kamus/WordNet/Corpus)
								Hybrid	Composit e		
1989 - 1994	5	1	0	0	0	0	5	1	0	1	0
1995 - 1999	7	4	1	1	4	2	0	1	0	4	4

2000 - 2004	42	18	12	14	20	11	9	16	5	22	17
2005 - 2009	13	5	4	8	10	7	0	6	1	9	7
2010 - 2014	4	4	9	16	20	15	0	11	3	16	12
Jml:	71	32	26	39	54	35	14	35	9	52	40

Sumber: Diolah dari berbagai sumber (2014)

IV. KESIMPULAN

Penelitian dan pengembangan model dan prototipe *schema matching* masih terbuka pada model *hybrid matcher* yang melibatkan dua atau lebih metode berbeda dan diproses secara *simultaneous*. Penggunaan *hybrid matcher* akan memberikan hasil yang lebih baik dan perbaikan kinerja (efektivitas) daripada *individual matcher*. Model dan prototipe baru juga dapat dikembangkan dengan mengeksplorasi hubungan semantik pada entitas atau *instance* sebagai dasar pemetaan hubungan antar elemen *schema*.

Penelitian terbaru, saat ini sedang dikembangkan sebuah model *hybrid schema matching* dan prototipe HybridScheMatch yang menggabungkan metode *constraint based* dan *instance based*, dengan melibatkan DTM (*data type matcher*), CM (*constraint matcher*), dan IDM (*instance data matcher*).

DAFTAR PUSTAKA

- [1] Cruz, I.F., Antonelli, F.P., and Stroe, C., 2009, Agreementmaker: Efficient matching for large real-world schemas and ontologies, Proc. of Int'l. Conf. on Very Large Data Base (VLDB) (demo paper), vol. 2, no. 2, pp. 1586-1589, August 24-28, 2009, Lyon, France.vol.
- [2] He, B. and Chang, K.C.C., 2003, Statistical schema matching across web query interfaces, Proc. of The ACM SIGMOD Int'l. Conf. Management of Data, pp. 217-228, June 9-12, 2003, San Diego, California, USA.
- [3] Do, H.H., 2005, Schema matching and mapping-based data integration, Ph.d. Thesis, Interdisciplinary Center for Bioinformatics and Dept. of Computer Science, University of Leipzig, Leipzig, Germany.
- [4] Kavitha, C., Sadasivam, G. S., and Shenoy, S. N., 2011, Ontology based semantic integration of heterogeneous databases, European Journal of Scientific Research, vol. 64, no. 1, November 2011, pp. 115-122.
- [5] Villanyi, B., Martinek, P., and Szikora, B., 2010, A novel framework for the composition of schema matchers, Proc. of The 14th WSEAS Int'l. Conf. on Computers, Latest Trends on Computers, pp. 379-384, July 23-25, 2010, Corfu Island, Greece.
- [6] Engmann, D. and Massmann, S., 2007, Instance matching with COMA++, Datenbank Systeme in Business, Technologie und Web (BTW Workshop) Proc.: Model Management and Metadata, pp. 28-37, March 5-6, 2007, Aachen, Germany.
- [7] Kim, W. and Seo, J., 1991, Classifying schematic and data heterogeneity in multidatabase systems, IEEE, vol. 24, no. 12, December 1991, pp. 12-18.
- [8] Do, H.H., Melnik, S., and Rahm, E., 2003, Comparison of schema matching evaluations, Proc. of 2nd Int'l. Workshop Web and Databases, In: Lecture Notes In Computer Science (LNCS) 2593, pp. 221-237, Springer-Verlag, Germany.
- [9] Bernstein, P.A., Madhavan, J., and Rahm, E., 2011, Generic schema matching, ten years later, Proc. of The VLDB Endowment, vol. 4, no. 11, pp. 695-701, August 29-September 3, 2011, Seattle, Washington.
- [10] Li, J., Tang, J., Li, Y., and Luo, Q., 2009, RiMOM: A dynamic multistrategy ontology alignment framework, Journal of IEEE

- Transaction Knowledge Data Engineering, vol. 21, no. 8, August 2009, pp. 1218-1232.
- [11] Bernstein, P., Harry B., Sanders, P., Shutt, D., and Zander, J., 1997, The Microsoft repository, Proc. of The 23rd Int'l. Conf. Very Large Databases (VLDB), pp. 3-12, August 25-29, 1997, Athens, Greece.
- [12] Bernstein, P.A., 2003, Applying model management to classical meta data problems, Proc. of The 1st Int'l. Conf. Innovative Data Systems Research (CIDR), pp. 209-220, January 5-8, 2003, Asilomar, California, USA.
- [13] Stabenau, A., McVicker, G., Melsopp, C., Proctor, G., Clamp, M., and Birney, E., 2004, An overview of ensembl, Genome Research Journal, vol. 14, no. 5, May 2004, pp. 929-933.
- [14] Madhavan, J., Bernstein, P.A., and Rahm, E., 2001, Generic schema matching with Cupid, Proc. of The 27th Int'l. Conf. on Very Large Data Bases (VLDB), pp. 49-58, September 11-14, 2001, Roma, Italy.
- [15] Massmann, S., Raunich, S., Aumueller, D., Arnold, P., and Rahm, E., 2011, Evolution of the Coma match system, OM-2011, Proc. of The 6th Int'l. Workshop on Ontology Matching (OM), October 24, 2011, Bonn, Germany.
- [16] Rahm, E. and Bernstein, P.A., 2001, A survey of approaches to automatic schema matching, Very Large Databases (VLDB) Journal, vol. 10, no. 4, pp. 334-350.
- [17] Shvaiko, P. and Euzenat, J., 2005, A survey of schema-based matching approaches, Journal on Data Semantics IV, vol. 4, LNCS 3730, pp. 146-171, December 2005.
- [18] Karasneh, Y., Ibrahim, H., Othaman, M., and Yaakob, R., 2009, Integrating schemas of heterogeneous relational databases through schema matching, Proc. of The 11th Int'l. Conf. on Information Integration and Web-based Applications and Services (iiWAS), pp. 209-216, December 14-16, 2009, Kuala Lumpur, Malaysia.
- [19] Özsü, M.T. and Valduriez, P.P., 2011, Principles of distributed database systems, 3rd edition, Pearson Education, Inc., Springer, New York, USA.
- [20] Milo, T. and Zohar, S., 1998, Using schema matching to simplify heterogeneous data translation, Proc. of The 24th Int'l. Conf. on Very Large Data Bases (VLDB), pp. 122-133, August 24-27, 1998, New York, USA.
- [21] Melnik, S., Rahm, E., and Bernstein, P., 2003, RONDO: A programming platform for generic model management, Proc. of The ACM-SIGMOD Conf. on Management of Data (SIGMOD), June 9-12 2003, pp. 193-204.
- [22] Banek, M., Vrdoljak, B., Tjoa, A. M., & Skocir, Z., 2008, Automated integration of heterogeneous data warehouse schemas, Int'l. Journal of Data Warehousing and Mining (IJDWM), vol. 4, no. 4, October-December 2008, pp. 1-21.
- [23] Gross, A., Hartung, M., Kirsten, T., and Rahm, E., 2010, On matching large life science ontologies in parallel, Proc. of The 7th Int'l. Conf. Data Integration in the Life Sciences (DILS), pp. 35-49, August 25-27, 2010, Gothenburg, Sweden.
- [24] Bergamaschi, S., Castano, S., and Vincini, M., 1999, Semantic integration of semistructured and structured data sources, SIGMOD Record, vol. 28, no. 1, March 1999, pp. 54-59.
- [25] Li, W.S. and Clifton, C., 2000, Semint: A tool for identifying attribute correspondences in heterogeneous databases using neural network, Data and Knowledge Engineering Journal, vol. 33, no. 1, April 2000, pp. 49-84.
- [26] Doan, A.H., Domingos, P., and Halevy, A. Y., 2001, Reconciling schemas of disparate data sources-a machine-learning approach, Proc. of The ACM SIGMOD Int'l. Conf. Management of Data, pp. 509-520, May 21-24, 2001, Santa Barbara, California, USA.
- [27] Li, W.S. and Clifton, C., 1994, Semantic integration in heterogeneous databases using neural networks, Proc. of 20th Int'l. Conf. on Very Large Data Bases (VLDB), pp. 1-12, September 12-15, 1994, Santiago de Chile, Chile.
- [28] Rahm, E., 2011, Towards large-scale schema and ontology matching, In: Bellahsene, Z., Bonifati, A., and Rahm, E., 2011, Schema matching and mapping, data-centric systems and applications, pp. 3-28, Springer, New York, USA.
- [29] Bellahsene, Z., Bonifati, A., and Rahm, E., 2011, schema matching and mapping, data-centric systems and applications, Springer, New York, USA.
- [30] Barrasa, J., Corcho, O., and Gómez-Pérez, 2004, R2O, An extensible and semantically based database to ontology mapping language, Proc. of The 2nd Workshop on Semantic Web and Databases (SWDB2004), pp. 92-119, 2004, Toronto, Canada.
- [31] Evermann, J., 2008, An exploratory study of database integration processes, IEEE Transactions on Knowledge and Data Engineering, vol. 20., no. 1, January 2008, pp. 99-115.
- [32] Bizer, C., Heath, T., and Berners-Lee, T., 2009, Linked data-the story so far, Int'l. Journal of Semantic Web Informations Systems, vol. 5, no. 3, October 2009, pp. 1-22.
- [33] Parundekar, R., Knoblock, C.A., and Ambite, J.L., 2010, Linking and building ontologies of linked data, Proc. of The 9th Int'l. Semantic Web Conf. (ISWC), pp. 598-614, November 7-11, 2010, Shanghai, China.
- [34] Falconer, S.M. and Noy, N.F., 2011, Interactive techniques to support ontology matching, In: Bellahsene, Z., Bonifati, A., and Rahm, E., 2011, Schema matching and mapping, data-centric systems and applications, pp. 29-52, Springer, New York, USA.
- [35] Aumueller, D., Do, H. H., Massmann, S., and Rahm, E., 2005, Schema and ontology matching with Coma++, Proc. of The SIGMOD, demo paper, pp. 906-908, June 14-16, 2005, Baltimore, Maryland, USA.
- [36] Seligman, L., Mork, P., Halevy, A.Y. Smith, K., Carey, M.J. Chen, K., Wolf, C., Madhavan, J., Kannan, A., and Burdick, D., 2010, OpenII: An open source information integration toolkit, Proc. of The ACM SIGMOD Int'l. Conf. on Management of Data (SIGMOD), pp. 1057-1060, June 6-10, 2010, Indianapolis, Indiana, USA.
- [37] Euzenat, J., Ferrara, A., Meilicke, C., Pane, J., Scharffe, F., Shvaiko, P., Stuckenschmidt, H., Šváb-Zamazal, O., Svátek, V., and Dos Santos, C.T., 2010, Results of the ontology alignment evaluation initiative 2010, Proc. of The 5th ISWC Workshop on Ontology Matching (OM), November 7, 2010, Shanghai, China.
- [38] Hernández, M.A., Miller, R.J., and Haas, L.M., 2001, Clio: A semi-automatic tool for schema mapping, Software Demonstration, Proc. of The ACM SIGMOD Int'l. Conf. Management of Data, pp. 607, June 2001, Santa Barbara, California, USA.
- [39] Naumann, F., Ho, C.T., Tian, X., Haas, L., and Megiddo, N., 2002, Attribute classification using feature analysis (poster), Proc. of The 18th Int'l. Conf. on Data Engineering (ICDE), pp. 271, February 26-March 1, 2002, San Jose, California, USA.
- [40] Popa, L., Hernández, M., Velegrakis, Y., and Miller, R., 2002, Mapping XML and relational schemas with Clio (software demonstration), Proc. of The Int'l. Conf. on Data Engineering (ICDE), pp. 498-499, February 26-March 1, 2002, San Jose, California, USA.
- [41] Haas, L.M., Hernández, M.A., Ho, H., Popa, L., and Roth, M., 2005, Clio grows up: from research prototype to industrial tool, Proc. of The ACM SIGMOD Int'l. Conf. Management of Data, pp. 805-810, June 13-16, 2005, Baltimore, Maryland, USA.
- [42] Kang, J. and Naughton, J., 2003, On schema matching with opaque column names and data values, Proc. of The ACM SIGMOD Int'l. Conf. Management of Data, pp. 205-216, June 9-12, 2003, San Diego, California, USA.
- [43] Chien, B.C. and He, S.Y., 2010, A hybrid approach for automatic schema matching, Proc. of the 9th Int'l. Conf. on Machine Learning and Cybernetics, 11-14 July 2010, pp. 2881-2886, Qingdao, China.
- [44] Li, W.S., Clifton, C., and Liu, S. Y., 2000, Database integration using neural networks: implementation and experiences, Knowledge and Information Systems Journal, vol. 2, no. 1., February 2000, pp. 73-96.
- [45] Do, H.H. and Rahm, E., 2002, Coma: A system for flexible combination of schema matching approach, Proc. of The 28th Conf. on Very Large Data Bases (VLDB), pp. 610-621, August 20-23, 2002, Hong Kong, China.
- [46] Madhavan, J., Bernstein, P.A., Chen, K., Halevy, A., and Shenoy, P., 2003, Corpus-Based Schema Matching, Proc. of The IJCAI-03 Workshop on Information Integration on the Web (IIWeb), pp. 59-63, August 9-10, 2003, Acapulco, Mexico.
- [47] Dhamankar, R., Lee, Y., Doan, A., Halevy, A., and Domingos, P., 2004, Imap: discovering complex semantic matches between database

- schemas, Proc. of The ACM SIGMOD Int'l. Conf. Management of Data, pp. 383-394, June 13-18, 2004, Paris, France.
- [48] Bernstein, P.A., Melnik, S., Petropoulos, M., and Quix, C., 2004, Industrial-strength schema matching, ACM SIGMOD Record, vol. 33, no. 4., December 2004, pp. 38-53.
- [49] Dragut, E. and Lawrence, R., 2004, Composing mappings between schemas using a reference ontology, Proc. of The Int'l. Conf. on Ontologies, Databases, and Applications of Semantics (ODBASE), pp. 783-800, October 25-29, 2004, Larnaca, Cyprus.
- [50] Mork, P. and Bernstein, P.A., 2004, Adapting a generic match algorithm to align ontologies of human anatomy, Proc. of The 20th Int'l. Conf. on Data Engineering (ICDE), pp. 787-790, Boston, MA, USA, 30 March-2 April 2004.
- [51] Tu, K.W. and Yu, Y., 2005, CMC: combining multiple schema-matching strategies based on credibility prediction, Proc. of The 10th Int'l. Conf. on Database Systems for Advanced Applications (DASFAA), pp. 888-893, April 17-20, 2005, Beijing, China.
- [52] Jian, N., Hu, W., Cheng, G., and Qu, Y., 20 05, Falcon-AO: Aligning ontologies with Falcon, Proc. of The K-CAP Workshop on Integrating Ontologies (K-CAP'05), pp. 85-91, October 2, 2005, Banff, Canada.
- [53] Jean-Mary, Y.R., Shironoshita, E.P., and Kabuka, M.R., 2009, Ontology matching with semantic verification, Web Semantics Journal, vol. 7, no. 3, September 2009, pp. 235-251.
- [54] Palopoli, L., Saccà, D., and Ursino, D., 1998, Semi-automatic, semantic discovery of properties from database schemes, Proc. of Int'l. Database Engineering and Applications Symposium (IDEAS), pp. 244-253, June 8-10, 1998, Cardiff, Wales, United Kingdom.
- [55] Palopoli, L., Sacca, D., Terracina, G., and Ursino, D., 1999, A unified graph-based framework for deriving nominal interscheme properties, type conflicts and object cluster similarities, Proc. of The 4th IFCIS Int'l. Conf. on Cooperative Information Systems (CoopIS), pp. 34-45, September 2-4, 1999, Edinburgh, Scotland.
- [56] Palopoli, L., Terracina, G., and Ursino, D., 2000, The system Dike: Towards the semi-automatic synthesis of cooperative information systems and data warehouses, Proc. of The Challenges: 2000 ADBIS-DASFAA Symposium on Advances in Databases and Information Systems, Enlarged 4th East-European Conf. on Advances in Databases and Information Systems, pp. 108-117, September 5-8, 2000, Prague, Czech Republic.
- [57] Castano, S., Antonellis, V.D., 1999, A schema analysis and reconciliation tool environment for heterogeneous databases, Proc. of The Int'l. Database Engineering and Applications Symposium (IDEAS), pp. 53-62, August 2-4, 1999, Montreal, Que.
- [58] Bergamaschi, S., Castano, S., Vincini, M., and Beneventano, D., 2001, Semantic integration of heterogeneous information sources, Data and Knowledge Engineering, vol. 36, no. 3, March 2001, pp. 215-249.
- [59] Mitra, P., Wiederhold, G., and Jannink, J., 1999, Semi-automatic integration of knowledge sources, Proc. of The 2nd Int'l. Conf. on Information Fusion (FUSION), July 6-8, 1999, Sunnyvale, California, USA.
- [60] Mitra, P., Wiederhold, G., and Kersten, M., 2000, Graph-oriented model for articulation of ontology interdependencies, Proc. of The 7th Int'l. Conf. Extending Database Technology (EDBT), pp. 86-100, March 27-31, 2000, Konstanz, Germany.
- [61] Mitra, P. and Wiederhold, G., 2002, Resolving terminological heterogeneity in ontologies, Proc. of The ECAI-02 Workshop on Ontologies and Semantic Interoperability, European Conf. on Artificial Intelligence, (ECAI), pp. 45-50, July 22, 2002, Lyon, France.
- [62] Giunchiglia, F., Yatskevich, M., and Avesani, P., 2005 (a), A large scale taxonomy mapping evaluation, Proc. of The 4th Int'l. Conf. Semantic Web Conf. (ISWC), pp. 67-81, November 6-10, 2005, Galway, Ireland.
- [63] Palopoli, L., Sacca, D., Terracina, G., and Ursino, D., 2003, Uniform techniques for deriving similarities of objects and subschemes in heterogeneous databases, IEEE Transaction Knowledge and Data Engineering, vol. 15, no. 2, Mar./Apr. 2003, pp. 271-294.
- [64] He, H., Meng, W., Yu, C., and Wu Z., 2003, Wise-Integrator-An automatic integrator of web search interfaces for e-commerce, Proc. of The 29th Int'l. Conf. on Very Large Data Bases (VLDB), pp. 357-368, September 9-12, 2003, Berlin, Germany.
- [65] He, B., Chang, K.C.C., and Han, J., 2004, Discovering complex matchings across web query interfaces-a correlation mining approach, Proc. of The 10th ACM SIGKDD Int'l. Conf. Knowledge Discovery and Data Mining, pp. 148-157, August 22-25, 2004, Seattle, Washington, USA.
- [66] Noy, N.F. and Musen, M.A., 2003, The Prompt Suite: Interactive tools for ontology merging and mapping, Int'l. Journal of Human-Computer Studies, vol. 59, no. 6, December 2003, pp. 983-1024.
- [67] Noy, N.F. and Musen, M.A., 2004, Using prompt ontology-comparison tools in the EON ontology alignment contest, Proc. of The 3rd Int'l. Workshop Evaluation of Ontology-Based Tools (EON), pp. 79-90, November 8th, 2004, Hiroshima, Japan.
- [68] Euzenat, J., Loup, D., Touzani, M., and Valtchev, P., 2004, Ontology alignment with Ola, Proc. of The 3rd Int'l. Workshop Evaluation of Ontology Based Tools (EON), pp. 56-88, November 8, 2004, Hiroshima, Japan.
- [69] Ehrig, M., and Staab, S., 2004, QOM-quick ontology mapping, Proc. of The 3rd Int'l. Semantic Web Conf. (ISWC), pp. 683-697, November 7-11, 2004, Hiroshima, Japan.
- [70] Ehrig, M. and Sure, Y., 2004, Ontology alignment-Karlsruhe, Proc. of The 3rd Int'l. Workshop Evaluation of Ontology-Based Tools (EON), pp. 48-55, November 8, 2004, Hiroshima, Japan.
- [71] Giunchiglia, F., Shvaiko, P., and Yatskevich, M., 2004, S-Match: An Algorithm and an Implementation of Semantic Matching, Proc. of The 1st European Semantic Web Symposium (ESWS), pp. 61-75, May 10-12, 2004, Heraklion, Crete, Greece.
- [72] Giunchiglia, F., Yatskevich, M., and Giunchiglia, E., 2005 (b), Efficient semantic matching, Proc. of The 2nd European Semantic Web Conf. (ESWC), pp. 272-289, May 29-June 1, 2005, Heraklion, Crete, Greece.
- [73] Benkley, S., Fandozzi, J., Housman, E., and Woodhouse, G., 1995, Data Element Tool-Based Analysis (DELTA), MITRE Technical Report MTR'95 B147.
- [74] Clifton, C., Housman, E., and Rosenthal, A., 1996, Experience with a combined approach to attribute-matching across heterogeneous databases, Proc. of The 7th IFIP 2.6 Working Conf. on Database Semantics, pp. 428-451, October 1997, Leysin, Switzerland.
- [75] Melnik, S., Molina, H.G., and Rahm, E., 2002, Similarity Flooding: A versatile graph matching algorithm, Proc. of The 18th Int'l. Conf. on Data Engineering (ICDE), pp. 117-128, February 26-March 1, 2002, San Jose, California, USA.
- [76] Lee, M.L., Yang, L.H., Hsu, W., and Yang, X., 2002, XClust: Clustering XML schemas for effective integration, Proc. of 11th Int'l. Conf. on Information and Knowledge Management (CIKM'02), pp. 292-299, November 4-9, 2002, Virginia, USA.
- [77] Lu, J., Wang, J., and Wang, S., 2005, An experiment on the matching and reuse of XML schemas, Proc. of 5th Int'l. Conf. on Web Engineering (ICWE), pp. 273-284, July 27-29, 2005, Sydney, Australia.
- [78] Berlin, J., and Motro, A., 2001 (b), Autoplex: Automated discovery of content for virtual databases, Proc. of The 9th Int'l. Conf. Cooperative Information Systems (CoopIS), In Cooperation with VLDB 2001, pp. 108-122, September 5-7, 2001, Trento, Italy.
- [79] Berlin, J., and Motro, A., 2001 (a), Automatch: database schema matching using machine learning with feature selection, Proc. of The 6th Int'l. Conf. Cooperative Information Systems (CoopIS), In Cooperation with VLDB 2001, pp. 108-122, September 5-7, 2001, Trento, Italy.
- [80] Embley, D.W., Jackmann, D., and Xu, L., 2001, Multifaceted exploitation of metadata for attribute match discovery in information integration, Proc. of 1st Int'l. Workshop Information Integration on the Web (WIIW), pp. 110-117, April 9-11, 2001, Rio de Janeiro, Brazil.
- [81] Xu, L. and Embley, D., 2003, Discovering direct and indirect matches for schema elements, Proc. of The 8th Int'l. Conf. on Database Systems for Advanced Applications (DASFAA), pp. 39-46, March 26-28, 2003, Kyoto, Japan.
- [82] Doan, A.H., Madhavan, J., Domingos, P., and Halevy, A.Y., 2002, Learning to map between ontologies on the semantic web, Proc. of 11th Int'l. Conf. on World Wide Web (WWW), pp. 662-673, May 7-11, 2002, Honolulu, Hawaii, USA.
- [83] Wang, J., Wen, J., Lochovsky, F., and Ma, W., 2004, Instance-based schema matching for web databases by domain-specific query probing,

Proc. of 13th Int'l. Conf. on Very Large Data-Bases (VLDB), pp. 408-419, August 31-September 3, 2004, Toronto, Canada.

The 3rd Int'l. Workshop Evaluation of Ontology-Based Tools (EON), pp. 67-78, November 8, 2004, Hiroshima, Japan.

[84] Hoshiai, T., Yamane, Y., Nakamura, D., and Tsuda, H., 2004, A semantic category matching approach to ontology alignment, Proc. of

[85] Bilke, A. and Naumann, F., 2005, Schema matching using duplicates, Proc. of The 21th Int'l. Conf. on Data Engineering (ICDE), pp. 69-80, April 5-8, 2005, Tokyo, Japan.