Intelligent Query Processing for Semantic Interoperable Information Systems

Benharzallah Saber[#1], Kazar Okba[*2], Caplat Guy[#3]

[#]*Department of Computer science, University of Mohamed Khider Biskra, 07000 Algeria,*
INSA de Lyon, Villeurbanne Cedex – France
[1]sbharz@yahoo.fr
[3]guy.caplat@insa-lyon.fr
[*]*Department of Computer science, University of Mohamed Khider Biskra, 07000 Algeria*
[2]kazarokba@yahoo.fr

***Abstract*— We propose a efficient query processing approach for semantic interoperable information systems. We propose also a generic multi agent architecture that supports our approach. Our approach consists in the exploitation of intelligent agents for query reformulation and the use of a new technology for the semantic representation. The algorithm is self-adapted to the changes of the environment, offers a wide aptitude and solves the various data conflicts in a dynamic way, it also reformulates the query using the schema mediation method for the discovered systems and the context mediation for the other systems.**

***Keywords*— Query answering; Semantic mediation; Multi-agent systems and OWL DL.**

I. Introduction

Interoperability has been a basic requirement for modern information systems environment. The cooperation of systems is confronted with many problems of heterogeneities and must take account of the open and dynamic aspect of modern environments. Querying the distributed ontologies is one major task in semantic interoperable information systems.

Various types of heterogeneity can be encountered cited as follow: technical, syntactic, structural and semantic heterogeneity. The resolution of semantic heterogeneity is becoming more important than before. Its types appear as: naming conflicts (taxonomic and linguistic problems), values conflicts (units and scales problems ...).

The solutions for the interoperability of the information systems evolved into the semantic mediation of the systems [1] [5] (all processes of interoperability of database allow to solve the semantic conflicts related to objects values and objects structure of the real-world).

The high number of the information sources implies the increase and the diversification of the conflicts number, as well as an increase in the time of localization of relevant information. It increases also the time of transmission of the queries towards all these information sources and the time response of the information sources. Therefore, the solutions of semantic interoperability should have an intelligent processor for query processing that allows the adaptation of the environment's changes and solves the various data conflicts in a dynamic way. Each solution provides some advantages to the detriments of others. Each one of them treats just one part of the data conflicts.

In this paper; we propose a efficient query processing approach for semantic interoperable information systems. We present also a generic multi agent architecture that supports our approach. In the following, section 2 presents a synthesis of the various existing approaches. Sections 3.A and 3.B describe the architecture of the mediation and the query processing. The section 3.C presents the technical aspects and prototype implementation.

II. Related words

As the query processing problem in distributed systems has been discussed in traditional databases and Semantic Web, two possible orientations have been proposed: the integration guided by the sources (schema mediation), and the integration guided by the queries (context mediation) [21][5] [6][8][4][10][11].

The schema mediation is a direct extension of the federate approach. Data conflicts are statically solved. In the schema mediation; the mediator should be associated with a knowledge set (mapping rules) for locating the data sources. The query processing follows an execution plan established by rules which determine the relevant data in order to treat a query (static resolution of queries). It requires a pre-knowledge on the systems participating in the cooperation. The mediator's role is to divide (according to the global schema) the user query in several sub-queries supported by the sources and gathers the results. The global schema is generally specified by object, logic, XML or OWL interfaces [24][17][3][5][22]. In all these works, the objective is to build a global schema which integrates all the local schemas. When one operates in an evolutionary world where sources can evolve all the time, the elaboration of a global schema is a difficult task. It would be necessary to be able to reconstruct the integrated schema each time a new source is considered or each time an actual source makes a number of changes [4]. Generally, the time response of the queries of this approach is better than the context mediation which requires much time (it uses the semantic reconciliation). In this approach; the transparency (is to give the illusion to the users whom they interact with a local system, central and homogeneous) is assured. The degree of automation of the resolution of the data conflicts is weak, and the scalability (the system effectiveness should be not degraded and the query processing remains independent of the addition or the suppression of systems in a given architecture) and evolutionarity (to control the update, the remove and the addition of information systems) are less respected compared to the context mediation.

Many works are dedicated to the proposition of automatic approaches for schemas/ontologies integration [30][31]. The schemas mapping notion have been particularly investigated in many studies, therefore it leads to the elaboration of several systems such as DIKE [7], COMA [13], CUPID [14]. It is possible to find analyses and comparisons of such systems in [18]. Several ontologies based approaches for integration of information were suggested. In [20] and [4] survey of this subject is presented. Among the many drawbacks of these works is that they do not describe the integration process in a complete way; they always use assumptions like pre-existence mappings [23][33] from a part, and from another part, they provide methods to calculate mappings between general or specific ontologies [30] and they do not indicate how to really exploit it for automatic integration or for the query reformulation [22][33].

In [21][3] the authors have proposed an extended schema mediation named DILEMMA based on the static resolution of queries. The mediation is ensured by a couple mediator/wrapper and a knowledge base associated with each system that takes part in the cooperation. The mediator comprises a queries processor and a facilitator. This approach provide a better transparency and makes it possible to solve the semantic values conflicts, but in a priori manner. The automation degree of the resolution of the data conflicts is enhanced compared to the schema mediation. This later involves always the recourse of an expert of the domain. It has a low capacity to treat evolutionarity and the scalability.

The first introduction of the context concept appeared in Kashyap and Seth pioneer work (1994). The role of the mediator in the context mediation approach is to identify, locate, transform and integrate the relevant data according to semantics associated with a query [21][3]. The resolution of data conflicts is dynamic and does not require the definition of a mediation schema. The user's queries are generally formulated in terms of ontologies. The data are integrated dynamically according to the semantic information contained in the description of the contexts. This approach provides a best evolutionarity of the local sources and the automation degree of the resolution of the data conflicts is better compared to schema mediation. Two categories of context mediation are defined: - the single domain approach SIMS [9], COIN [10] working on a single domain where all the contexts are defined by using a universal of consensual speech. The scalability and evolutionarity are respected but remains limited by the unicity of the domain. - Multi-domains approaches Infosleuth [11], Observer [12] they use various means to represent and connect heterogeneous semantic domain: ontologies, hierarchy of ontologies and method of statistical analysis.

In the context mediation approach the data conflicts are dynamically solved during the execution of the queries (dynamic query resolution), allowing the best evolution of the local sources and the automation degree is enhanced compared to the schema mediation, this to the detriment of time response of the queries (it uses the semantic reconciliation). Concerning the semantic conflicts, the majority of the projects solve only the taxonomic conflicts (Coin [10]). The resolution of the values conflicts is either guided by the user (Infosleuth [11]), or unsolved in the majority of cases (Observer [12] [28]).

The agent paradigm gives a new insight for the systems nature development such as: complex, heterogeneous, distributed and/or autonomous [15][34][26][2]. Several works of semantic interoperability use the agent paradigm [16][11][25][29] [32].

Infosleuth project [11] is used to implement a set of cooperative agents which discover, integrate and present information according to the user or application needs for which they produce a simple and coherent interface. The Infosleuth's architecture project consists of a set of collaborative agents, communicating with each other using the agent communication language KQML. Users express their queries on a specific ontology using KIF and SQL. The queries are dispatched to the specialized agents (agent broker, ontological, planner...) to retrieve data on distributed sources. The resolution of many semantic conflicts remains guided by the user [3]. They use specialized agents seen as threads which are widely different from the usual definition of the cognitive agent given in the distributed artificial intelligence.

In [25], the authors propose a multi-agent system to achieve semantic interoperability and to resolve semantic conflicts related to evolutive ontologies domain. In this approach, the query processing and the validation of the mappings are completely related to the users. [29] propose an agent based intelligent meta-search and recommendation system for products through consideration of multiple attributes by using ontology mapping and Web services. This framework is intended for an electronic commerce domain.

III. Approach description

Our objective is to realize a semantic mediation having the following characteristics:

- Give permission to the system which provides its context of application to find the information systems and information shared by those systems. This information is integrated dynamically for the system for which it can use them transparently.
- To ensure the advantages of the context and schema mediation, and to avoid their disadvantages. Our approach focuses on the dynamic change of the mediation system, of the context mediation to the schema mediation. This change is done by the use of the intelligent agents, in order to ensure the open aspect of the context mediation (high automation degree of the resolution of the data conflicts...) and the robustness of

the schema mediation (the formulation of the queries in schema terms...).

- To solve the majority of the semantic conflicts by using new technologies for semantic representation, and by respecting a high automation degree of the query processing. The query processing algorithm reformulates the query by using the schema mediation method for the systems discovered and the context mediation for the others. So, our query processing approach is self-adapted to the changes of the environment.

A. *Generic Architecture based Agent for Context and Schema mediation (GAACSM)*

The cooperation suggested in our solution is based on:

- A preliminary construction of information before its integration in the architecture system,
- The static and dynamics query resolution.

An information system can play the role of information supplier and/or consumer (Fig1).

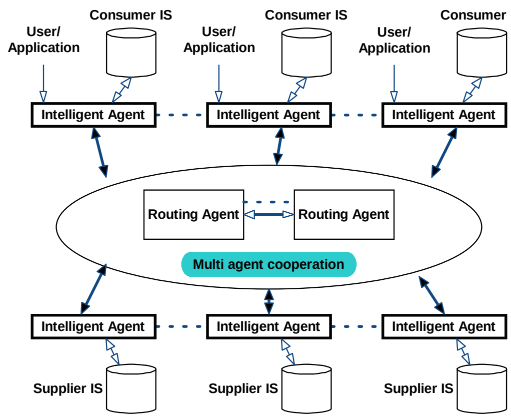

Fig 1. General architecture of the proposed approach

Our architecture consists of two types of agents: intelligent agents (IAs) and routings agents (described below)

The integration phase of a new information system (IS) in our proposed mediation system begins with the creation of an IA and continues with the fastening of this last to a routing agent (RA) which is *nearest semantically*.

Before creating an IA, we've to create its knowledge base (KB). An IA is an intermediary between an information system and the semantic mediation environment. Mainly, the KB of an IA contains: the context of its local information system, the name of the domain, the ontology which describes the name of the domain and the ontology of the semantic conflicts values (OSCV). This information makes it possible to prepare the IA to the semantic mediation.

The new IA integrated into the system of mediation applies the Contract Net protocol and sends an invitation describing its domain. The RAs receiving the call and provide their ability (*semantic proximity rate)*. As soon as, the IA receives answers from all RAs, then it evaluates these rates, and makes its choice on a RA which is the nearest semantically. The chosen RA adds the previous IA to its net contacts.

Our approach does not use a global schema or some predefined mappings. Users interrogate the consuming system (the queries formulated in term of the consuming schema).

At the beginning, the intelligent agent consuming (IAC) applies the dynamic query resolution protocol (context mediation) because it does not have information on the suppliers systems. This protocol is applied via the RA which is the nearest semantically with the IAC. During the dynamic evaluation of the query, the intelligent agent suppliers (IASs) update their histories and add information (mapping between terms of query ontology and their ontologies) to facilitate their dynamic integration with the IAC.

Each IAS replies with results, the RA updates its KB and reorders the list of IASs that are the most important to previous IAC (in other words; the IASs which contain results are at the head of the list). If no IAS replies, the RA sends the query to other RAs. If there are replies, the RA adds the IASs of other RAs to its KB (auto reorganization).

During the operation of the mediation system, the IAC applies the protocol to discover suppliers which are the nearest semantically to its domain, and to integrate them dynamically in order to use them in the schema mediation. For this aim, it cooperates with the RA. Indeed; the RA updates its KB during its communication with the other agents. Particularly, its KB contains for each IA an ordered list of its IASs which are not discovered yet. These IASs should be near semantically to it. The first IAS in the list is the one which has largest number of responses of IA. After that the first IAS becomes the next supplier solicited to the following dynamic integration done by IA.

After the dynamic integration, the IAC updates its knowledge base by mapping rules.

During the operation of the system, the IAC discovers some suppliers and adapts itself with the environment. So, to treat a query two protocols should be applied: the static query resolution protocol is adopted for the discovered systems and the dynamic query resolution for other systems.

B. *Queries processing*

The query processing is divided into several steps, and during this process, the multi-agents system uses a set of protocols (fig 2):

Algorithm . Query processing
Given L the list of discovered agents and their mappings
If QueryValidation() then
1:if L <> empty then
- QueryReformulation()
- StaticRecombiningResults()

2: *Dynamic query resolution*
- SemanticEnrichmentQuery()
- TransmissionSemanticallyEnrichedQuery()
- SemanticEvaluation()
- DynamicRecombiningResults()

Fig 2 : Query processing

1) Static query resolution: The static resolution is applied to the systems have been already discovered.

Step 1: query validation. The IAC checks the validity of the query. i.e. whether it is written in schema mediation terms or not,

Step2: query reformulation. The query is divided into a recombining query of the results and sub queries intended for the IAS which contain data necessary to the execution of the query. The decomposition of the query is done by the use of the mapping rules. The IAC applies the cooperation protocol of static query resolution.

Step3: recombining of the results. The IAC executes the recombining query for the results.

2) Dynamic query resolution: The dynamic resolution makes it possible to take into account the appearance of new IASs. The principal steps are:

Step 1: Semantic enrichment of a query. The IAC enriches the query semantically by using the ontology and the links schema-ontology which are in its own knowledge base.

Step 2: Transmission of the semantically enriched query. The IAC applies the cooperation protocol of dynamic query resolution. So it transmits the semantically enriched query to the routing agent which is nearest semantically. This latter sends it to all IASs of its net contacts.

Step 3: Semantic evaluation of the semantically enriched query. Each IAS answers according to its capacity to treat the query:

- To compare elements of the query with its ontology. The elements of the query and its ontology are compared by using a semantic distance. The identified elements as equivalent are retained.
- The query is rewritten in terms of the equivalent elements of its ontology (then interpreted on its schema) to take into account the semantic conflicts of values (each intelligent agent has library of functions for the conversion of the types).
- The answer is sent latter to the routing agent, indicating the manner of treating the query, so that this letter can build recombining queries of the results.

If no IAS answers, the routing agent sends the query to the other routings agents of other domains and if there are answers the routing agent updates its net contacts.

Stage 4: Results recombining the routing agent recomposes the results obtained by IASs. Then it sends the final result to the IAC, this latter recomposes the results of static and dynamic query resolution.

C. Prototype implementation

Our implementation is based on three class libraries: OntoSim [35], Alignment API [36] and Jade [37]. The cooperation protocols are implemented using the Jade platform. In order to implement the communication between agents we adopted the most effective method that uses ontologies for the descriptions of messages. The description of ontology in Jade is a solution given for the transport of objects. Concerning the local information systems, the local database of the consuming system and the database of the supplier system 1 are established under the Access DBMS and the Windows XP operating system. The database of the supplier system 2 is implemented in XML files and the same operating system.

When new IS wants to be recorded at the mediation system, that involves the creation of an IA in addition to a descriptive level of this IS. If no RA near semantically exists (the semantic proximity rate between IA and all RAs is a lower from predefined level), an RA is created. Jade gives the possibility for an agent to create another one (while passing by the container, agentContainer.createNewAgent). The scalability and the performances of the transport system of Jade message were treated in [27][28]. The obtained results confirm the fact that Jade deals well with the scalability according to several scenarios intra or inter framework. The fig 3 presents an example of comparison between two ontologies of the consuming system and the supplier system 2. It presents also the automatic generation of the mappings rules between the schemas of the two systems.

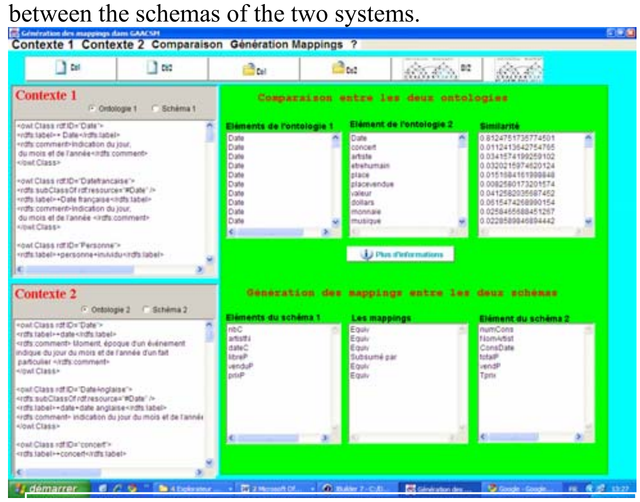

Fig 3 : automatic mapping generation

In order to facilitate the implementation of our prototype we suppose that the IAC is queried via a graphical interface where the user can insert their queries. The queries are formulated in terms of the schema of IAC. Figure 4 presents the graphical interface, an example of query and the obtained results. In this example, the IAS1 is discovered by agent IAC. This last applies the schema mediation in order to reformulate the query. The IAC applies the context mediation for other agents, which are not yet discovered (IAS2). It communicates with the agent RA.

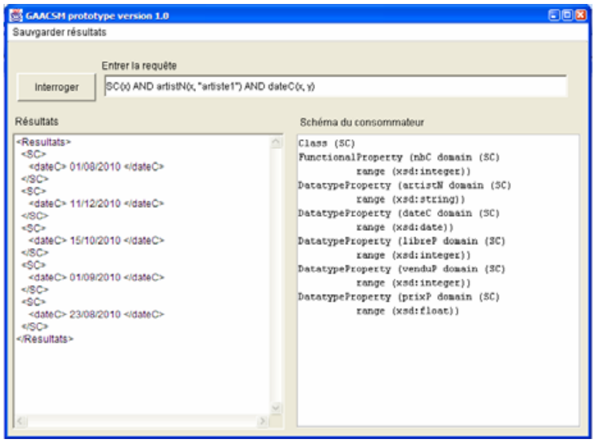

Fig 4 : a simple exemple

D. Conclusion and future research

In this paper we presented a efficient query answering for semantic interoperable information systems. We proposed a generic multi agent architecture supporting our approach.

The main advantage of our query answering is its robustness with regard to the evolution of systems, adaptation to the changes of environment and solves the most various data conflicts in a dynamic way. The developed prototype shows us the functionality of architecture suggested. As prospect we try to slacken our data mediation towards service mediation in general and to use intelligent methods to reduce ontologies to be compared for not to influence the scalability of architecture suggested.

REFERENCES

[1] Journal article : E. Sciore, M. Siegel, A. Resenthal, Using Semantic Values to Facilitate Interoperability Among Heterogeneous Information Systems, ACM Transaction on Database Systems, Vol. 19, N.2, pp. 254-290, Jun. 1994

[2] LI Ai-Ping, Jia Yan and WU Quan-Yuan, "A Study on Organizational Knowledge in Multi-Agent System", JCIT:Journal of Convergence Information Technology, Vol.2 No.1, pp.61-65, 2007

[3] Thesis: Fabrice Jouanot, « DILEMMA : vers une coopération de systèmes d'informations basée sur la médiation sémantique et la fusion d'objets », université de bourgougne, novembre 2001.

[4] Michel Schneider, Lotfi Bejaoui, Guillaume Bertin, A Semantic Matching Approach for Mediating heterogeneous Sources, metadata and semantic, springer US, p 537-548, ISBN 978-0-387-77744-3, 2009

[5] Article: Susanne Busse, Ralf-Detlef Kutsche, Ulf Leser, Herbert Weber, "Federated Information Systems: Concepts, Terminology and Architectures", Technical Report Nr.99-9 Berlin University,1999

[6] Farshad Hakimpour, thesis Using Ontologies to Resolve Semantic Heterogeneity for Integrating Spatial Database Schemata, Universität Zürich Zürich, 2003

[7] Palopoli, L., Terracina, G., Ursino, D.: DIKE: a system supporting the semi-automatic construction of cooperative information systems from heterogeneous databases. Softw., Pract. Exper. 33(9), pp. 847-884 (2003).

[8] Eric LECLERCQ, Djamal BENSLIMANE, Kokou Yétongnon, "ISIS: A Semantic Mediation Model and an Agent Based Architecture for GIS Interoperability", in Proceedings of the International Database Engineering and Applications Symposium (IDEAS'99), Montreal, CANADA, August 2-4, 1999. Published by IEEE Computer Society Press, ISBN 0-7695-0265-2, p. 81--92

[9] ARENS Y. et al , Query processing in the SIMS Information Mediator. In TATE A. Ed. Advanced planning technology, Menlo park, CA, AAAI Press, 1996.

[10] Bressan S, et al., "Context Knowledge Representation and Reasoning in the Context Interchange System", Applied Intelligence, vol. 13, n°2, September 2000, p. 165-180.

[11] Article: R. J. Bayardo Jr.et al, "InfoSleuth: Agent-Based Semantic Integration of Information in Open and Dynamic Environments", Microelectronics and Computer Technology Corporation,1997

[12] Article: Mena E., Illarramendi A., Kashyap V., Sheth A., "OBSERVER : An approach for Query Processing in Global Information Systems based on Interoperation across Pre-existing Ontologies", International journal on Distributed And Parallel Databases DAPD, vol.8, n°2, April 2000, p. 223-272.

[13] Do, H. H., Rahm, E.: COMA - A System for Flexible Combination of Schema Matching Approaches. VLDB 2002, pp. 610-621 (2002).

[14] Madhavan, J., Bernstein, P.A., Rahm, R.: Generic Schema Matching with Cupid. VLDB 2001, pp. 49-58 (2001).

[15] R. GIRARDI, "An analysis of the contributions of the agent paradigm for the development of complex systems", In (SCI 2001) and (ISAS 2001), Orlando, Florida. 2001.

[16] Purvis M, Cranefield S., Bush G., Carter D., McKinlay B., Nowostawski M. Ward R., The NZDIS Project: an Agent-Based Distributed Information Systems Architecture, Proceedings of the Hawai'i International Conference On System Sciences, January 4- 7, 2000, Maui, Hawaii.

[17] Seksun Suwanmanee et al, OWL Based Approach for Semantic Interoperability, Proceedings of the 19th International Conference on Advanced Information Networking and Applications (AINA 05)

[18] Mohsenzadeh, M., Shams, F., Teshnehlab, M.: Comparison of Schema Matching Systems. WEC (2), pp 141-147 (2005).

[19] Deborah L. McGuinness and Frank van Harmelen. « Owl web ontology language ». Technical report, 2004.

[20] Wache, H., Vogele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., Hubner, S.:Ontology-based integration of information - a survey of existing approaches. In Stuckenschmidt, H., ed., IJCAI-01 Workshop: Ontologies and Information Sharing, pp 108-117 (2001).

[21] Fabrice Jouanot, Nadine Cullot, and Kokou Y´etongnon, Context Comparison for Object Fusion CAiSE 2003, LNCS 2681, pp. 536−551, 2003.

[22] Rubén Tous, Thesis: Data Integration with XML and Semantic web Technologies. Universitat Pompeu Fabra, Barcelona, June 2006

[23] François Goasdoué and Marie-Christine Rousset, Querying Distributed Data through Distributed Ontologies: ASimple but Scalable Approach, IEEE INTELLIGENT SYSTEMS,p60-65, 2003

[24] Nang Tom Hawm Saw, OWL-Based Approach for Semantic teroperating and Accessing Heterogeneous XML Sources, 0-7803-9521-2/06/$20.00 §2006 IEEE.

[25] Guilaine Talens, Danielle Boulanger, and Magali Séguran, Domain Ontologies Evolutions to Solve Semantic Conflicts M. Collard (Ed.): ODBIS 2005/2006, LNCS 4623, pp. 51–67, 2007., Springer-Verlag Berlin Heidelberg 2007

[26] J. G. Carbonell and J. Siekmann, Lecture Notes in Artificial Intelligence, Agent-Oriented Information Systems III 7th International Bi-ConferenceWorkshop, AOIS 2005 Utrecht, ISBN-10 3-540-48291-1 Springer Berlin Heidelberg NewYork, 2006

[27] E. Cortese, F. Quarta, G. Vitaglione. Scalability and Performance of Jade Message Transport System. Presented at AAMAS Workshop on AgentCities, Bologna, 16 th July, 2002.

[28] Thesis: Magali Séguran, Résolution des conflits sémantiques dans les systèmes d'information coopératifs : proposition d'un modèle d'interaction entre agents, Université Jean Moulin, Lyon3, 2003

[29] Wooju Kim & Dae Woo Choi & Sangun Park, Agent based intelligent search framework for product information using ontology mapping, J Intell Inf Syst (2008) 30:227–247 DOI 10.1007/s10844-006-0026-8, Springer Science + Business Media, LLC 2008

[30] Fleur Mougin, Anita Burgun, Olivier Bodenreider, Julie Chabalier, Olivier Loréal4, and Pierre Le Beux1, Automatic Methods for Integrating Biomedical Data Sources in a Mediator-Based System, DILS 2008, LNBI 5109, pp. 61–76, 2008. Springer-Verlag Berlin Heidelberg 2008

[31] H. L. Hansen, K. Ingvaldsen , Dynamic Schema Integration The ACRoS Prototype THESIS, The Faculty of Social Sciences, Department of Information Science and Media Studies, may 2005

[32] Avigdor Gal and Aviv Segev, Agent Oriented Data Integration, J. Akoka et al. (Eds.): ER Workshops 2005, LNCS 3770, pp. 98–108, 2005.Springer-Verlag Berlin Heidelberg 2005

[33] Peter haase,dissertation, semantic technologies for distributed information systems, universitatsverlag karlsruhe 2007, ISBN 978-3-86644-100-2.

[34] Ning Zhong, Jiming Liu, Setsuo Ohsuga, Jeffrey Bradshaw, Intelligent Agent Research and Development 2nd Asia-Pacific Conference on IAT , ISBN 981-02-4706-0 World Scientific Publishing Co. Pte. Ltd., 2001

[35] https://gforge.inria.fr/projects/ontosim/

[36] http://alignapi.gforge.inria.fr/lib.html

[37] http://jade.tilab.com/