Integrating Large Knowledge Repositories in Multiagent Ontologies

Herlina Jayadianti, Carlos Sousa Pinto, Lukito Edi Nugroho, Paulus Insap Santosa

herlinajayadianti@gmail.com
csp@dsi.uminho.pt
lukito@mti.ugm.ac.id
Insap@mti.ugm.ac.id

Abstract. Knowledge is people's personal map of the world. According to the knowledge differences, it is possible different groups of people have different perceptions about the same reality. Each perception can be represented by using ontologies. In the research underlying this paper we are dealing with a multiple ontologies. In that context, each agent explores its own ontology. The goal of this research is to generate a common ontology including a common set of terms, based on the several ontologies available, in order to make possible to share the common terminology (set of terms) that it implements, between different communities. In this paper we are presenting a real implementation of a system using those concepts. The paper provides a case study involving groups of people in different communities, managing data using different perceptions (terminologies), and different semantics to represent the same reality. Each user – belonging to a different community - uses different terminologies in collecting data and as a consequence they also get different results of that exercise. It is not a problem if the different results are used inside each community. The problem occurs if people need to take data from other communities, sharing, collaborating and using it to get a more global solution.

Keywords. Heterogeneity, Agents, SPARQL, Ontology alignment, Common ontology.

1 Introduction

In information technology, a repository is a central place in which an aggregation of data is kept and maintained in an organized way. Repository is a place where things are collected. Depending on how the term is used, a repository may be directly accessible to users or may be a place from which specific databases, files, or documents are obtained for further relocation or distribution in a network. As an example scenario, institution A, institution B and institution C are working in a domain D. Repositories which contain information about that domain can be scattered in different places. One of the main problems that we can find in such a scenario is related to the existence of different perceptions and to the use of different representations and terms in each repository in each institution. Our problem is how to combine different repositories from different institutions and how to manage knowledge between these different repositories. Heterogeneity in data, in semantic and in perception between each institution is the major problem we need to solve. We use ontologies to solve those problems. Using ontologies we can shared different conceptualizations, different terminologies, and different meanings between systems [18]. However, tasks on distributed and heterogeneous systems demands support from more than one ontology.

We can distingue four types of heterogeneity [1]: (1) Paradigm heterogeneity that occurs if distinct agents express their knowledge using different modelling paradigms; (2) Language heterogeneity which occurs if distinct agents express their knowledge in different representation languages; (3) Ontology heterogeneity that occurs if distinct agents make different ontological assumptions about their domain of knowledge; (4) Content heterogeneity which occurs if distinct agents express different knowledge the same reality. Ontology integration [4], [9-12] is one way to solve the problem of heterogeneity and it can be done using several approaches. For example, ontology merging, ontology matching or ontology alignment. The integration of ontologies creates a new ontology by reusing other available ontologies through assembling, extending, or specializing operations. In integration processes the source ontologies and the resultant ontology can have different amounts of information [2]. We need to map ontologies in order to make compatible different terminologies (sets of terms). While having some common ground, either within an application area or for some high-level general concepts, this could alleviate the problem of data and semantic heterogeneity [5].

Ontology alignment or ontology matching [3], [13], [14] is the process of determining correspondence between concepts. Given two ontologies $i = (C_i, R_i, I_i, A_i)$ and $j = (C_j, R_j, I_j, A_j)$, we can define different types of (inter ontology) relationships among their terms. If two ontologies have at least one common component (relation, hierarchy, type, etc.) then they may be compared. Since the characteristics (attributes) of concepts capture the details of those concepts, they provide a good opportunity to find similarities [1].

In this paper we describe an approach to solve the problem of data and semantic heterogeneity using a common ontology derived from several different ontologies, using an ontology alignment process. This paper is organized as follows: (1) Introduction; (2) In this section we present several definitions of the terms used in operations involving ontologies, in order to avoid possible misunderstandings; (3) In this section we present the case study that underlies the work described in the paper; (4) This section describes the implementation of the proposed solution; (5) In this section we refer the used technologies and preliminary results of our work ; and (6) the paper ends with the Conclusions.

2 Operations Involving Ontologies – Used Terminology

To avoid potential misunderstandings, we present the definitions of the terms used throughout this paper.

- **Ontology Combination** is the process of using two or more ontologies and can be used to implement alignment, merge or integration of different ontologies. The combined ontologies usually hold data which is relevant to all ontologies involved.[6], [7]
- **Ontology Merging** is the process of building a single ontology through the merging of several source ontologies. Usually the source ontologies cover similar or overlapping domains. [8]
- **Ontology Alignment** is the process of determining correspondence between concepts and the process of creating a new ontology from two or more ontologies by overlapping the common parts. The domains of the source ontologies are different from the domain of the resulting ontology, but there is a relation between these domains. [3], [13], [14]
- **Ontology Matching** is the process of reaching global compatibility between two or more ontologies so that the resulting ontology is consistent and coherent. [3]
- **Ontology Mapping** is the process of relating similar concepts or relations from different sources through some equivalence relation. Mapping allows finding correspondences between the concepts of two ontologies. If two concepts correspond, then they mean the same thing or closely related things. Currently, the mapping process is regarded as a promise to solve the problem between ontologies since it attempts to find correspondences between semantically related entities that belong to different ontologies. It takes as input two ontologies, each consisting of a set of components (classes, instances, properties, rules and axioms). [15], [16], [17]

3 Heterogeneity And Interoperability Problems

In this section, we describe the problem we are trying to solve and an approach to solve it. Considering some reality, different groups of people (different communities) have different opinions, use different sets of data about it and have diverse perceptions about that reality. Figure 1 represents several communities that faced reality with different perceptions *(Perception_1, Perception_2, and Perception_N)*. Perceptions are converted into data that is saved into separate storage devices not interconnected. Repositories *db1, db2, and dbN* contain different data, different concepts, different terms, and different semantics. It depends on people in the group who look at reality (policy makers) and people who create and store data (users that use technology). Users who deal with computers has a very important role in controlling and changing the terminology and semantic of the data. Each group (community) uses technology to find data. It is very difficult for those different groups to get similar results and the problem happens if people need to use data from another group in order to share, collaborate and use it to get a more global solution.

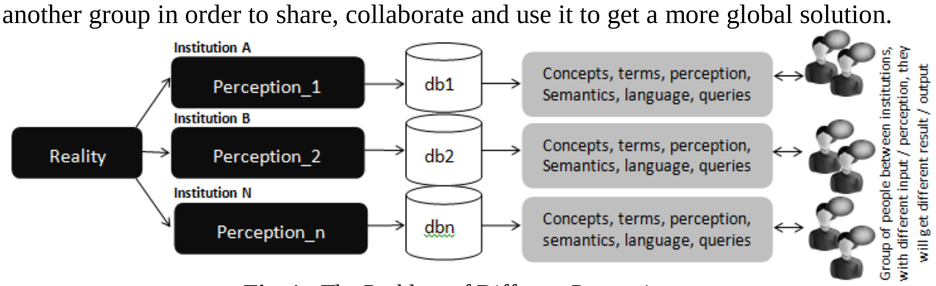

Fig. 1. The Problem of Different Perceptions

The solution presented in this paper is based on different knowledge about the same reality based on different perceptions and uses a mechanism that works with a set of common concepts, common terms, common semantics, common languages, and a set of common queries (See Figure 2). Users in each community still can use their different concepts, terms, and perceptions as inputs for querying the system. According to the proposed solution, we aim to get similar answers (output) from such a common layer that acts like an interface between the different systems and the users.

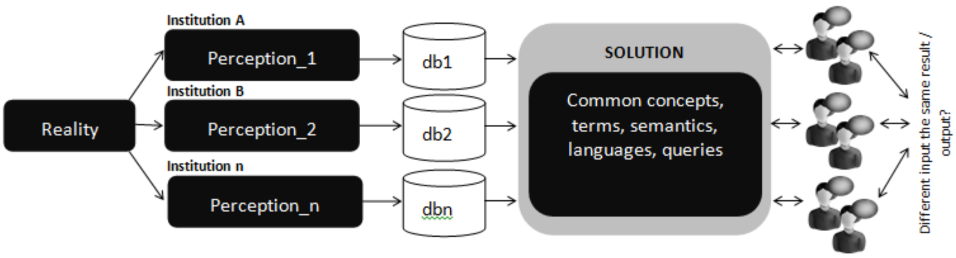

Fig. 2. Towards a Solution of Different Perceptions

4 Using Ontologies to implement the solution

Ontology is defined as a formal, explicit specification of a shared conceptualization [18]. Tasks on distributed and heterogeneous systems demand support from more than one ontology. Multiple ontologies need to be accessed by different systems. Different perceptions about the same reality led to dissimilar ontologies for the same domain. Thus, various organisms with different ontologies do not fully understand each other. To solve this problem, it is necessary to use ontology alignment geared for interoperability.

4.1 Ontology Alignment

Ontology Alignment [13], [14] is the process of creating a new ontology from two or more ontologies by overlapping common parts and determining correspondences between ontology entities. Entities of the source ontologies are different from entities of the resulting ontology, but there is a relation between these entities. Based on the fundamental concepts above and on Figure 2, the solution for solve the problem is to use ontology alignment (see Figure 3) to create a new ontology (a common ontology) by overlapping the common parts of the original ontologies. Common part is a common word recognized and used with the same meaning by different communities. CO (Common Ontology) is expected to overcome the differences that exist in the different source ontologies. In Figure 3 we use ontology UV_1 from institution A, UV_2 from institution B, and ontology UV_n from institution $_N$. CO will contain terms that will be equated with each term in the source UVs.

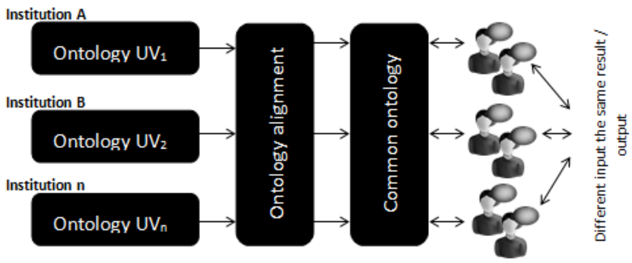

Fig. 3. Ontology Alignment

4.2 Dictionary and Search engine analysis

To get the CO terms we analyzed several dictionary such as WordNet[1] and Thesaurus[2] (See Table 1).

[1] Wordnet is a large lexical database or electronic dictionary for English. WordNet implements measure of similarity and relatedness among terms. Measures of similarity use information found in an *is–a* hierarchy of concepts, and quantify how much concept A is similar to concept B. http://wordnet.princeton.edu/

[2] Thesaurus is a reference work that lists words grouped together according to similarity of meaning (Synonym or antonym). http://en.wikipedia.org/wiki/Thesaurus/ and http://thesaurus.com/

Search string	*Synonym*	
	Wordnet 2.1	Thesaurus
People	Group, Family, Masses, Mass, Family Line	Citizens Community, Family, Folk, Folks, General Public, Heads, Persons, Population, Society
Person	Individual, Someone, Somebody	Human, Identity, Individual, Individuality

Table 1. Synonym results found by Wordnet and Thesaurus using "People" and "Person" as search string

There are four senses for the term people in Wordnet (version 2.1).
Sense 1 *people -- ((plural) any group of human beings (men or women or children) collectively) => group, grouping*
Sense 2 *citizenry, people -- (the body of citizens of a state or country) => group, grouping*
Sense 3 *people -- (members of a family line; "his people have been farmers for generations) => family, family line, folk*
Semantic Web Search Engines such as Swoogle[3], Watson[4], and Sindice[5] (See Table 2) accept queries in a format that varies from one tool to another.

Search string	Semantic Search Engine					
	Swoogle		Watson		Sindice	
	Number of references	Time	Number of Terms	Time	Number of references	Time
People	1,818	0.456	**12,348**	0	**12,709,732**	2.19
Person	**16,320**	0.237	3,046	0	77,724,899	0.04
Group	3,812	0,381	3,742	0	690,570	0.68
Family	2,209	0,565	7,326	0	7,081,244	2.24
Individual	1,010	0,469	854	0	237,692	2.05

Table 2. Different results found by several search engines using "People", "Person", "Group", "Family" and "Individual" as search strings

Different from other types of platforms that can be used to find suitable ontologies, which usually only provide browse functionalities, Semantic Web Search Engines (SWSE) permit another degree of automation. For instance, a query on Sindice for ontologies including the term "People", returned more than 12.699.661 results in 2.72 second, where near 4.568.172 documents (0.03 second) of them were RDF files. Data from Table 2 was taken on June 20, 2012.

4.3 A Case Study

To demonstrate the capabilities of the described mechanisms we implemented an alignment process between original ontologies using data about poverty. Poverty is not the focus of our research. We just use that case as a real scenario that allows us to demonstrate the validity of our approach. We combine different existing terminolo-

3 Swoogle is the first Web search engine dedicated to online semantic data. Its development was partially supported by DARPA and NFS (National Science Foundation). http://swoogle.umbc.edu/

4 Watson development was partially supported by the NeOn (http://www.neon-project.org) and the OpenKnowledge (http://www.openk.org) project. http://kmi-web05.open.ac.uk/WatsonWUI/

5 http://sindice.com/

gies about the same reality (poverty in this case) used by different communities in order to get a common set of terms that can be transparently used by those communities, while maintaining the original terms in the data sources. We use Indonesia as the country for the example because in that country there are several institutions in charge of dealing with poverty data, generating problems due to differences in the criteria used by them to make their surveys, even considering that the semantics of these different criteria are the same. For example, let's consider the two institutions, BKKBN[6] (institution A) and BPS[7] (institution B), that are responsible for collecting data on poverty. Each institution has a different system and use different sets of terms to describe the same domain and different criteria to classify people as poor or not. In fact, institution A uses 24 criteria and institution B has 14 criteria to define poverty.

Institution A: *"Normally all family members have* ***meal*** *two or more times a day"*

Institution B: "*Minimum two times per day the family have* ***food****"*

Meal and *food* have the same meaning, as well as *suit* and *clothes* or *clinic* and *hospital.* To be similar *(≅)* or not equal *(≠)* depend on several factors, such as the programmer's interpretation, the needs of the system itself, and last but not least the domain/area that we are talking about. One term has always a strong relationship with the domain. In this research, we focus on poverty domain, identifying terms that are most commonly used by users.

Table 3 shows some examples of criteria and terms in the domain of poverty from two different institutions. Currently, both institutions are working separately to collect and manage data on poverty. Each institution sends data to the government based on its perception. Institution A (BKKBN) is more focused on family welfare and institution B (BPS) is more concerned with basic needs. The major problem of this situation is the great impact on aid distribution.

	Criteria from Institution A	Criteria from Institution B
Classes	Area, Assets, Contraceptive, Education, FoodConsume, GovernmentAid, Hospital, HealthProblem, HouseCondition, Person	Asset, BirthControlMethod, EducationLevel, Food, GeographicArea, GovernmentHelp, HealthCondition, Clinic, HouseParameter, JobArea, Person.
Object Properties	isComposedBy, hasFrequentlyEat, PassTheStudyFrom, hasRarelyEat, has Assets, hasChildren, hasfamily, hasHouseCondition, hasJobPositionAs	EnergyUsedForCooking, hasEduBackground, hasFrequentlyEaten, hasLargestFloorMadeFrom, hasRarelyEaten,
Data properties	Address, has Age, FrequentlyEatenADay, hasMarriageStatus, hasSalary, hasaGoodHouseCondition	hasAge, DistrictCode, FloorArea, FullName, HouseCondition, JobsArea, NameOfFood, FloorArea, Salary ≈ hasWage, hasStatus.

Table 3. Example of Classes, Object Properties, and Data properties From Two Institutions

6 Badan Keluarga Berencana Nasional (BKKBN) or National Population and Family Planning Board is a governmental agencies that appointed to conduct a survey of poverty in Indonesia. http://www.**bkkbn**.go.id

7 Badan Pusat Statistik (BPS) or Central Berau of Statistic is a non departmental government institution directly responsible to the President of Indonesia. http://www.**bps**.go.id

Based on the criteria of both institution (see Table 1), we identify an example of Classes, ObjectProperties, and DataProperties to be used by institutions A and B (see Table 3). We can see that:

- **Terms (classes) in Ontology UV$_1$** = {Area, Assets, Contraceptive, Education, FoodConsume, GovernmentAid, Hospital , HealthProblem, HouseCondition, Person}
- **Terms (classes) in Ontology UV$_2$** = {Asset, BirthControlMethod, EducationLevel, Food, GeographicArea, GovernmentHelp, HealthCondition, Clinic, HouseParameter, JobArea, Person}.

By using WordNet, Thesaurus, and Swoogle, we identify common classes in CO, namely *People*, *Birth Control*, *Education*, *Food*, *Health*, *Property*, *Work*, *Hospital*, and *House Condition*. On the next stage, by overlapping the common parts, we determine the correspondence between classes in Ontology UV$_1$ (User view 1) and classes in ontology UV$_2$ (User view 2) with classes in CO. Figure 4, automatically generated in Protégé[8], show the relation between CO and UVs.

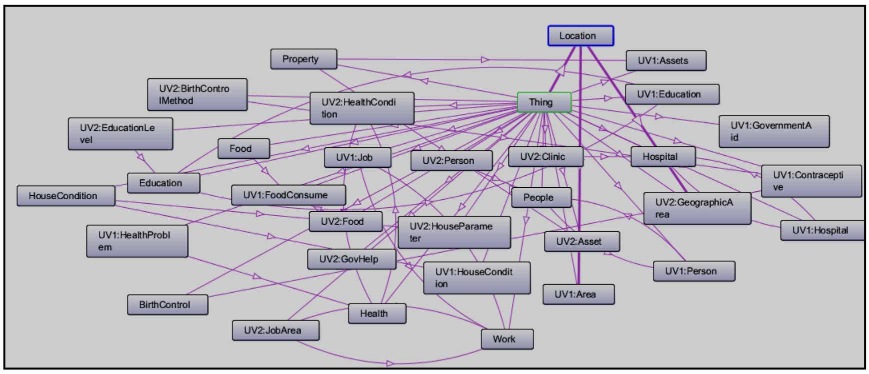

Fig. 4. The relation between UV's and CO

5 Used Technologies and Preliminary Results

Web Ontology Language (OWL) is a language for create ontologies to the web. OWL was designed for processing information and to provide a common way to process the content of web information. SPARQL[9] is a graph-matching query language. SPARQL can be used to express queries across diverse data sources. In Figures 5-7 we can see examples of the results of SPARQL queries. Based on Figure 5 we can see that Ontology UV$_1$ (data taken form Institution A) consists of classes Person, Food, Job, Floor and Area. UV$_1$ also includes the object properties "RarelyEat" (Chicken instance), "JobName" (Farmer instance) and TypeOfFloor (Soil instance). With SPARQL we get as result from UV$_1$ two people included in these criteria.

---

[8] http://protegewiki.stanford.edu/wiki/Protege4GettingStarted

[9] http://www.w3.org/TR/rdf-sparql-query/

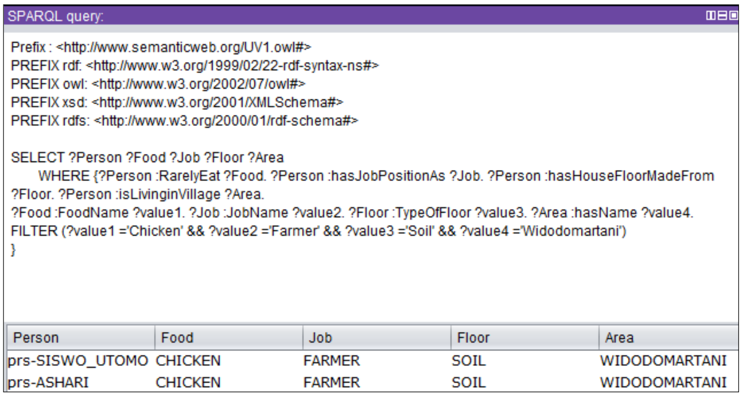

Fig. 5. SPARQL result using UV_1

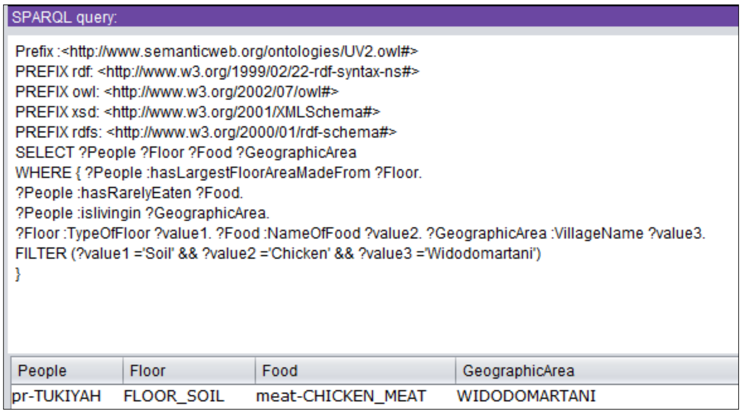

Fig. 6. SPARQL result using UV_2

As we can see in Figure 6 ontology UV_2 (data taken from Institution B) consists of classes Person, Food, GeographicArea, and Floor (subclass of class House Condition) and also consists of object properties hasRarelyEaten (Chicken instance), isLivingIn (Widodomartani instance) and hasLArgestFloorAreaMadeFrom (Soil instance). Using SPARQL we get as result from UV_2 one person included in these criteria. It should be highlighted that poverty data in UV_1 and UV_2 was taken from the same village, Widodomartani. Based on the criteria used by Institution A and Institution B, implemented in the ontologies UV1 and UV2, the results returned by SPARQL queries are: Siswo Utomo and Ashari are poor people considering the ontology UV1, and Tukiyah is a poor person when considered the ontology UV2.

With common term in CO (see Figure 7), we can see that Siswo Utomo, Ashari and Tukiyah are poor people. With ontology alignment we determine the correspondence among concepts and implement the process of creating a new ontology based on two ontologies (UV1 and UV2) by overlapping the common parts.

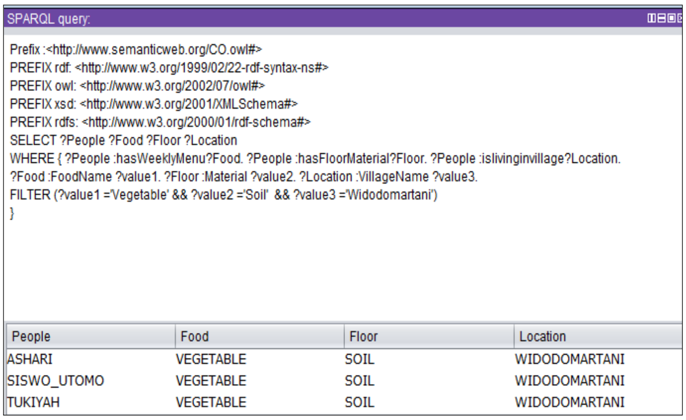

Fig. 7. SPARQL result in CO

Our future work will include functionalities that will allow users ask queries using JSP[10] (JavaServer Pages) and Jena[11] ontology API against OWL/RDF files. Through the ontology API, Jena provides a consistent programming interface for ontology applications.

6 Conclusion

Different communities have different perceptions and use different sets of terms (terminologies) to represent the same reality. The problem of it is how to share a different perception between communities and how to make a correspondence between different terms. In this research we used ontology alignment as a process to create a new ontology (common ontology) using a common set of terms by overlapping the common parts of the source ontologies. Using this approach it is possible to share different conceptualizations, different terminologies, and different meanings between different systems. We believe that ontology alignment is one of the best approaches to solve the problem of data and semantic heterogeneity.

Acknowledgment

We would like to acknowledge the support of the Erasmus Mundus EuroAsia program (2010-2012) for the research foundation of this research, and also to

---

10 Javaserver Pages is a technology provides a simplified, fast way to create dynamic web content. http://www.oracle.com/technetwork/java/javaee/jsp/index.html.

11 Jena provides a collection of tools and Java libraries to help user to develop semantic web. http://jena.apache.org/.

acknowledge Universidade do Minho and Universitas Gadjah Mada for the collaboration.

REFERENCES

[1] A. Malucelli and others, “Ontology-based services for agents interoperability,” *Faculdade de Engenharia, Universidade do Porto, Porto*, 2006.
[2] Y. Xue, “Ontological View-driven Semantic Integration in Open Environments,” The University of Western Ontario, 2010.
[3] J. Euzenat and P. Shvaiko, *Ontology matching*. Springer-Verlag New York Inc, 2007.
[4] A. Gangemi, D. Pisanelli, and G. Steve, “Ontology integration: Experiences with medical terminologies,” in *Formal ontology in information systems*, 1998, vol. 46, pp. 98–004.
[5] N. F. Noy, “Semantic integration: a survey of ontology-based approaches,” *ACM Sigmod Record*, vol. 33, no. 4, pp. 65–70, 2004.
[6] L. Stojanovic, N. Stojanovic, and J. Ma, “An approach for combining ontology learning and semantic tagging in the ontology development process: eGovernment use case,” *Web Information Systems Engineering–WISE 2007*, pp. 249–260, 2007.
[7] H. S. Pinto and D. N. Peralta, “Combining ontology engineering subprocesses to build a time ontology,” in *Proceedings of the 2nd international conference on Knowledge capture*, 2003, pp. 88–95.
[8] A. Pease, I. Niles, and J. Li, “The suggested upper merged ontology: A large ontology for the semantic web and its applications,” in *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*, 2002, vol. 28.
[9] D. Calvanese, G. De Giacomo, and M. Lenzerini, “A framework for ontology integration,” in *The Emerging Semantic Web—Selected Papers from the First Semantic Web Working Symposium*, 2002, pp. 201–214.
[10] H. S. Pinto and J. P. Martins, “A methodology for ontology integration,” in *Proceedings of the 1st international conference on Knowledge capture*, 2001, pp. 131–138.
[11] J. P. M. A. Silva and others, “Automatic and intelligent integration of manufacture standardized specifications to support product life cycle-an ontology based methodology,” 2009.
[12] H. S. Pinto, A. Gómez-Pérez, and J. P. Martins, “Some issues on ontology integration,” 1999.
[13] J. Euzenat, “Semantic precision and recall for ontology alignment evaluation,” in *Proc. 20th International Joint Conference on Artificial Intelligence (IJCAI)*, 2007, pp. 348–353.
[14] B. Chen, H. Tan, and P. Lambrix, “Structure-based filtering for ontology alignment,” in *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2006. WETICE’06. 15th IEEE International Workshops on*, 2006, pp. 364–369.
[15] N. Choi, I. Y. Song, and H. Han, “A survey on ontology mapping,” *ACM Sigmod Record*, vol. 35, no. 3, pp. 34–41, 2006.
[16] N. F. Noy, “Ontology mapping,” *Handbook on ontologies*, pp. 573–590, 2009.
[17] Y. Kalfoglou and M. Schorlemmer, “Information-flow-based ontology mapping,” *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE*, pp. 1132–1151, 2002.
[18] T. Gruber, “What is an Ontology,” *Encyclopedia of Database Systems*, vol. 1, 2008.