

Leveraging Knowledge from Different Communities Using Ontologies

Herlina Jayadianti^{1,2}, Carlos Sousa Pinto¹, Lukito Edi Nugroho²,
Paulus Insap Santosa², and Wahyu Widayat²

¹ Departamento de Sistemas da Informacao,
Universidade do Minho, Campus de Azurem,
Guimaraes, Portugal

² Electrical Engineering and Information Technology,
Gadjah Mada University, Yogyakarta, Indonesia

³ Economic Development, Gadjah Mada University, Yogyakarta,
Indonesia

herlinajayadianti@gmail.com, csp@dsi.uminho.pt,
{Lukito, Insap}@mti.ugm.ac.id, wahyu@mep.ugm.ac.id

Abstract. The purpose of this paper is to provide research based understanding of leveraging knowledge and managing knowledge within and across several communities using the poverty domain as a case study. We hypothesize that leveraging knowledge with a good taxonomy and a good integration process are good approaches to organize and share knowledge. Problems appear when a group of people in different communities share data and collaborate using different perceptions, different concepts, different terms (terminologies), and different semantics to represent the same reality. In this paper we present an approach to solve this problem. We will generate a common set of terms based on the terms of several different storage devices, used by different communities, in order to make data retrieval independent of the different perceptions and terminologies used by those communities. We use ontologies to represent the particular knowledge of each community and discuss the use of mapping and integration techniques to find correspondences between the concepts used in those ontologies.

Keywords: Leveraging Knowledge, Knowledge Management, Common Ontology, Perception, Terminology, Ontology.

1 Introduction – Leveraging Knowledge

Information technology has lead many institutions or communities to imagine a new world of leverage knowledge. Internet made it possible for professionals allowing them to draw on the latest thinking of their peers no matter where they are located. As a result many communities are rethinking how works gets done, linking people to electronic media so they can leverage each other's knowledge. Knowledge is different from information and sharing it requires a different set of concepts and tools. Four characteristics of knowledge distinguish it from information: [1]–[3]

- Knowing is a human act; whereas information is an object that can be filed, stored and moved around.
- Knowledge is created in present moment; whereas information is fully made and can sit in storage. To share knowledge we need to think about the current situation.
- Knowledge belongs to communities.
- Knowledge circulates through communities in many ways.

Leveraging knowledge involves a unique combination of human and information systems [3], [4]. Leveraging knowledge also allows us to tap into many innovations to access diverse knowledge bases and integrate them to create new competencies with multi-dimensional concepts [5], [6]. Ironically to leverage knowledge we need to focus on the community that owns it and the people who use it, not the knowledge itself. Six implications for leveraging knowledge are: [3]

- Different communities. Focus on knowledge important to both the business and the people. People naturally seek help, share insights and build knowledge in areas they care about.
- Sharing information. The ways to share knowledge should be as multidimensional as knowledge itself. Most corporate knowledge sharing efforts revolve around tools.
- Let the community decide what to share and how to share it. Knowledge needs to have an “owner” who cares. It is tempting to create organization-wide systems for sharing knowledge.
- Community support structure. Communities are held together by people who care about the community.
- Use the community’s terms for organizing knowledge.
- Integrate sharing knowledge into the natural work flow.

Knowledge and learning are the only capabilities that can provide sustained competitive advantage. ‘Knowledge’ is the content of learning. ‘Learning’ is the process of gaining new knowledge, so that the firm is constantly accumulating and assimilating knowledge and this becomes the basis for creating and improving organizational routines [7]. Knowledge is a critical resource that warrants much more attention. If we are serious about managing knowledge, then we need to embrace the concepts associated with knowledge management [3], [6], [8]. Since knowledge is the sense we make of information, then the way information is organized is also a sense making device. A good taxonomy should be intuitive for those who use it.

To be “intuitive” it needs to tell the story of the key distinctions of the field, reflecting the natural way discipline members think about the field [3]. There are great temptations to make all systems for organizing knowledge the same, such as formatting information – to make it easily transferred, and having the same metadata – to make it easy searched, indexed and used in different context. However beyond that, the system for organizing information should be the community’s. If a community of people sharing knowledge spans several disciplines, then such thing of terms and structures should be the *common* among those communities [3].

Having some *common* ground, among those communities, either within an application area or for some high-level general concepts, this could alleviate the problem of integrating knowledge [9]–[12]. Based on the presented reasons, we believe that ontologies with common terms and common concepts are very important in a knowledge sharing process. In this paper we describe an approach of leveraging knowledge using a common set of terms derived from several different ontologies. This paper is organized as follows: (1) Introduction; (2) Knowledge and Common ontology; (3) Implementation of the solution; (4) Conclusions.

2 Knowledge and Common Ontology

Figure 1 shows the relation between knowledge and ontologies.

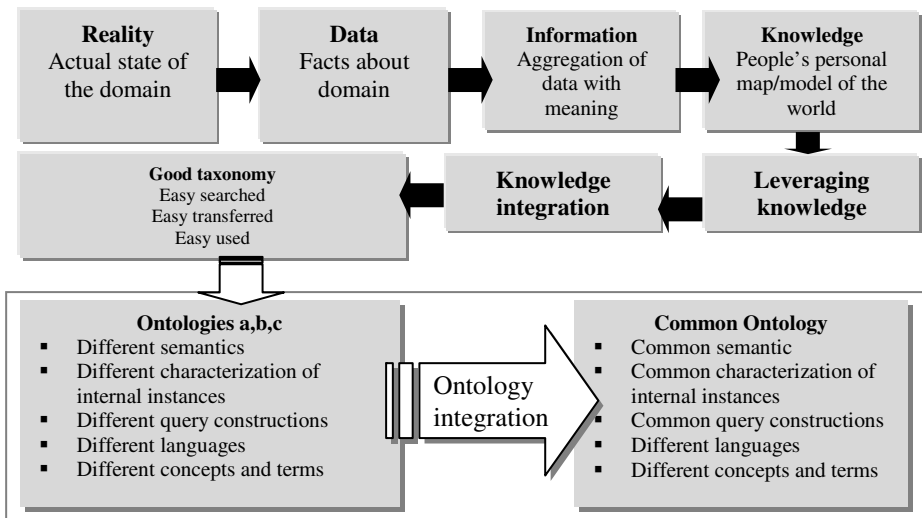


Fig. 1. Managing Knowledge, and Leveraging Knowledge with Ontologies

At the level “Reality” we represent the actual state of a particular domain. At this level we can find lots of data. Data are facts in the context of a domain of discourse. At the next level, establishing relationships between data, it is possible to derive information and expand it beyond the limits of understanding of each person. Knowledge is obtained by adding experience, reflection and reasoning to information. If different information is discussed by people, it is easy to understand what is inside their minds, either by arguments or communication, but what happens if those differences exist at the machine level? We need to combine information so that machines can “think” and understand the concepts we can find inside human brains. To do that, we can use ontologies to represent data and information of the several communities. Ontology is some formal description of a domain of discourse. However, ontology is not enough to make computers understand what is necessary. Scattered ontologies should then be incorporated and integrated into a new ontology,

a Common Ontology (See Fig.2). Ontology integration is one way to solve the problem of data, information, and semantic heterogeneity. Semantic heterogeneity on naming includes problems with synonyms (same concept with different terms) and homonyms (same term with different meanings). Semantic heterogeneity occurs when the same reality is modeled by two or more different people or systems [13].

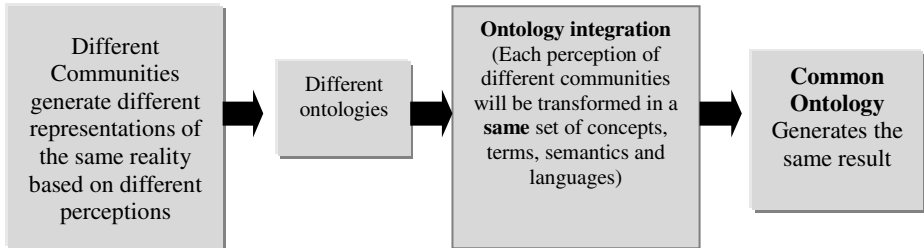


Fig. 2. Towards a Solution of Different Perceptions

The goal of ontology integration is to derive a more general domain ontology (Common Ontology) from other several ontologies in that domain. Every person has his/her own knowledge. They can justify everything based on their thoughts, perceptions and conceptualizations. Conceptualization is an abstraction of the external world inside an individual mind. It can be used to construct one or several concepts and also to interpret some reality in a conceptual way [14].

3 Implementing the Solution

Ontology integration is one way to solve the problem of semantic heterogeneity and it can be done using several approaches. For example, merging, matching or mapping. The integration of ontologies creates a new ontology by reusing other available ontologies through assembling [15]–[17], extending [18], or specializing operations [19]. In integration processes the source ontologies and the resultant ontology can have different amounts of information [14]. Ontology process integration implies several steps. According to Noy [20] there are some specific challenges in the ontology integration process:

- Finding similarities and differences between ontologies in an automatic and semi-automatic way;
- Defining mappings between ontologies;
- Developing an ontology integration architecture;
- Composing mappings across different ontologies;
- Representing uncertainty and imprecision in mappings.

Particularly, in ontology integration, some tasks should be performed to eliminate differences and conflicts between those ontologies [20]. Ontology integration is used to find similarities and differences between ontologies. The goal of ontology

integration is to derive a more general domain ontology (Common Ontology) from several other ontologies in the same domain, into a consistent unit. The domain of both the integrated and the resulting ontologies is the same.

3.1 How to Get Common Terms – Common Ontology (CO)

Groups of people in different communities will probably have a different way of view to the same reality. Different view of each set of users is then called user view (UV) and can be implemented using an ontology. Common Ontology is expected to overcome the differences that exist in the several user views (UVs). CO will contain common terms which will then be equated with each term in the UVs. Common term is a common word recognized and used with the same meaning by different communities. To get the CO terms we use WordNet¹, Thesaurus² and Swoogle³ (See Table 1). Wordnet is a large lexical database or electronic dictionary for English [21], [22]. 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 [23]. Thesaurus is a reference work that lists words grouped together according to similarity of meaning. Swoogle is the first Web search engine dedicated to online semantic data. Its development was partially supported by DARPA and NFS (National Science Foundation).

Table 1. Equivalences for some terms related to poverty from different applications

Search string	Synonym		
	Wordnet 2.1 (Noun)		Swoogle (Terms)
Hospital	Infirmary, medical institution	Clinic, emergency room, health service, hospice, infirmary, rest home.	Hospital, hospital,
Clinic	Medical institution, Session, Medical building, health facility, facility	Emergency room, hospice, infirmary, nursing home, rest home.	Clinic, Clinical, ClinicalTreatment

There are two senses for the term hospital in Wordnet (version 2.1).

Sense 1. *hospital, infirmary -- (a health facility where patients receive treatment)*
=> *Medical building, health facility, healthcare facility -- (building where medicine is practiced)*

Sense 2. *hospital -- (a medical institution where sick or injured people are given medical or surgical care)*
=> *Medical institution -- (an institution created for the practice of medicine)*

¹ <http://wordnet.princeton.edu/>

² <http://thesaurus.com/>

³ <http://swoogle.umbc.edu/>

Swoogle was the first search engine dedicated to online semantic data. Its development was partially supported by DARPA and NFS (National Science Foundation).

There are three senses for the term clinic in Wordnet (version 2.1).

Sense 1. *clinic* -- (a medical establishment run by a group of medical specialists)
=> *Medical institution* -- (an institution created for the practice of medicine)

Sense 2. *clinic* -- (meeting for diagnosis of problems and instruction or remedial work in a particular activity)
=> *Session* -- (a meeting for execution of a group's functions)

Sense 3. *clinic* -- (a healthcare facility for outpatient care)
=> *Medical building, health facility, healthcare facility* -- (building where medicine is practiced)

3.2 Implementation

Let's consider the referred case study of poverty and two ontologies or user views both representing that domain. By using WordNet, Thesaurus and Swoogle we can find synonyms and similarities that were chosen from ontology UV1 and Ontology UV2. The next step is to find the number of each synonym with Google and Swoogle. The results provided by Google and Swoogle are different due to the number of documents that are available in each system. Google provides more documents than Swoogle. Currently, Swoogle only indexes some metadata about Semantic Web documents⁴.

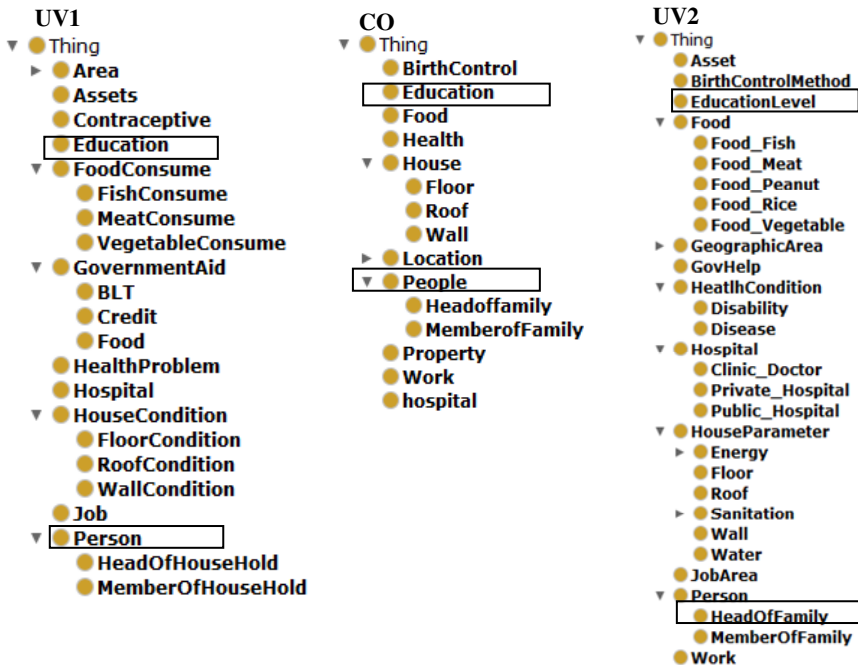


Fig. 3. Terms in ontology UV1, UV2 and CO

⁴ http://swoogle.umbc.edu/index.php?option=com_swoogle_manual&manual=faq

Based on data from Swoogle and Google, then we selected the term with highest references. We assume that the number of references reflects a widely and commonly usage of the term by users. We use a common term as a term in CO. For example: term Person is more specific than People (See Fig.3). We use Wordnet, Swoogle and Google not only for comparing the number of result to get common terms but also to find common ObjectProperties. For example: ObjectProperties hasFloorMadeFrom (CO) is more general than hasHouseFloorMadeFrom (UV2) and hasLargestFloorAreaMadeFrom (UV1) (See Fig. 4). Poverty is not the focus of our research. We just use that case as a real scenario that allows us to demonstrate our approach. We combine different existing terminologies 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 communities in charge of dealing with poverty data, generating problems due to differences in the criteria used to make their surveys, even considering that the semantics of these different criteria are the same. For example, let’s consider the two communities, BKKBN⁵ and BPS⁶, that are responsible for collecting data on poverty. Each community has a different system and use different sets of terms to describe the same domain and

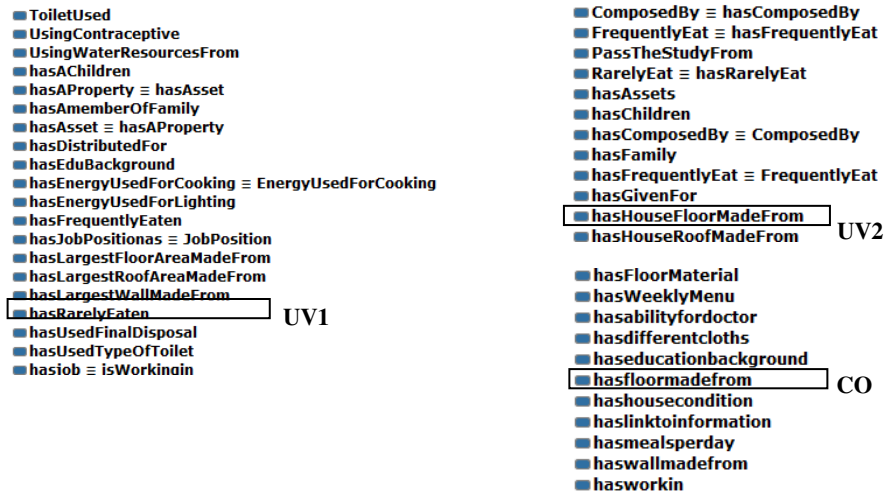


Fig. 4. ObjectProperties in ontology UV1, UV2 and CO

⁵ 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. www.bkkbn.go.id

⁶ Badan Pusat Statistik (BPS) or Central Berau of Statistic is a non departmental government institution directly responsible to the President of Indonesia. www.bps.go.id

different criteria to classify people as poor or not. To be similar (\cong) or not equal (\neq) 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. Currently, both communities are working separately to collect and manage data on poverty. Each community sends data to the government based on its perception.

3.3 Testing Queries

SPARQL⁷ is a query for Resource Description Framework (RDF)⁸. SPARQL can be used to express queries across diverse data sources whether the data is stored natively as RDF or viewed as RDF via middleware. SPARQL contains capabilities for querying and also supports extensible values for testing and constraining queries [24]. SPARQL commands from our work shows below.

```
Prefix :<http://www.semanticweb.org/CO.owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?People ?Food ?Floor ?Location
WHERE { ?People :hasWeeklyMenu?Food. ?People :hasFloorMaterial?Floor.
?People :islivinginvillage?Location.
?Food :FoodName ?value1. ?Floor :Material ?value2. ?Location
:VillageName ?value3.
FILTER (?value1 = 'Vegetable' && ?value2 = 'Soil' && ?value3
='Widodomartani')
}
Prefix : <http://www.semanticweb.org/UV1.owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?Person ?Food ?Job ?Floor ?Area
WHERE {?Person :RarelyEat ?Food. ?Person :hasJobPositionAs ?Job. ?Person
:hasHouseFloorMadeFrom ?Floor. ?Person :islivinginvillage ?Area.
?Food :FoodName ?value1. ?Job :JobName ?value2. ?Floor :TypeOfFloor
?value3. ?Area :hasName ?value4.
FILTER (?value1 = 'Chicken' && ?value2 = 'Farmer' && ?value3 = 'Soil' &&
?value4 = 'Widodomartani')
}
```

ObjectProperties hasHouseFloorMadeFrom in UV1 is equivalent to ObjectProperties hasFloorMaterial in CO. ObjectProperties hasFloorMaterial is more common than ObjectProperties has HouseFloorMadeFrom. In this work, we found the same result these queries (See Fig.5).Our future work will include

⁷ <http://www.w3.org/TR/rdf-sparql-query/>

⁸ RDF is a standard model for data interchange on the web. <http://www.w3.org/RDF/>

SPARQL query:			SPARQL query:		
Prefix: <http://www.semanticweb.org/UV1.owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX owl: <http://www.w3.org/2002/07/owl#> PREFIX xsd: <http://www.w3.org/2001/XMLSchema#> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> SELECT ?Person ?Food ?Job ?Floor ?Area WHERE { ?Person :RarelyEat ?Food. ?Person :hasJobPosit ?Floor. ?Person :isLivinginVillage ?Area. ?Food :FoodName ?value1. ?Job :JobName ?value2. ?Floor :Ty FILTER (?value1 = 'Chicken' && ?value2 = 'Farmer' && ?value3 = ' }			Prefix: <http://www.semanticweb.org/CO.owl#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX owl: <http://www.w3.org/2002/07/owl#> PREFIX xsd: <http://www.w3.org/2001/XMLSchema#> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> SELECT ?Person ?Food ?Floor ?Location WHERE { ?Person :hasWeeklyMenu ?Food. ?Person :hasFloorMaterial ?Flo ?Food :FoodName ?value1. ?Floor :Material ?value2. ?Location :VillageNam FILTER (?value1 = 'Vegetable' && ?value2 = 'Soil' && ?value3 = 'Widodomarta }		
Person	Food	Job	People	Food	Floor
prs-SISWO_UTOMO	CHICKEN	FARMER	ASHARI	VEGETABLE	SOIL
prs-ASHARI	CHICKEN	FARMER	SISWO_UTOMO	VEGETABLE	SOIL
			TUKIYAH	VEGETABLE	SOIL

Fig. 5. The same result between ontology UV1 and CO

functionalities that will allow users ask queries using JSP⁹ (Java Server Pages) and JENA¹⁰ ontology API against OWL/RDF files. Through the ontology API, JENA provides a consistent programming interface for ontology applications.

4 Conclusions

In this research we try to leveraging knowledge by using an ontology integration as a process to create a new ontology (Common Ontology). Using this approach it is possible to share different conceptualizations, different terminologies, and different meanings between different systems. We believe that ontology integration 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 for the research foundation of this research, and also to acknowledge Universidade do Minho and Universitas Gadjah Mada for the collaboration.

References

- [1] McDermott, R.: Knowing in community. IHRIM, 19 (2000)
- [2] McDermott, R.: Knowing is a human act, Informatik-Informatique. Zeitschrift der Schweizerischen Informatikorganisationen 1, 7–9 (2002)
- [3] McDermott, R.: Why information technology inspired but cannot deliver knowledge management. Knowledge and Communities 41(4), 21–35 (2000)

⁹ Javaserer Pages is a technology provides a simplified, fast way to create dynamic web content.

<http://www.oracle.com/technetwork/java/javase/jsp/index.html>

¹⁰ <http://jena.apache.org/>

- [4] Lesser, E.L., Fontaine, M.A., Slusher, J.A.: *Knowledge and Communities*. Routledge (2000)
- [5] Hatteland, C.J.: Commentary on leveraging knowledge based resources: The role of contracts. *Journal of Business Research* 65(2), 162–163 (2012)
- [6] Mudambi, R., Swift, T.: Leveraging knowledge and competencies across space: the next frontier in international business. *Journal of International Management* 17(3), 186–189 (2011)
- [7] Carayannis, E.G.: Knowledge-driven creative destruction, or leveraging knowledge for competitive advantage strategic knowledge arbitrage and serendipity as real options drivers triggered by co-opetition, co-evolution and co-specialization. *Industry and Higher Education* 22(6), 343–353 (2008)
- [8] Quinn, J.B.: Strategic outsourcing: leveraging knowledge capabilities. *Sloan Management Review* 40(4), 9–21 (1999)
- [9] Euzenat, J., Shvaiko, P.: *Ontology matching*. Springer-Verlag New York Inc. (2007)
- [10] Gangemi, A., Pisanelli, D., Steve, G.: Ontology integration: Experiences with medical terminologies. In: *Formal Ontology in Information Systems*, vol. 46, pp. 98–94 (1998)
- [11] Noy, N.F.: Semantic integration: a survey of ontology-based approaches. *ACM Sigmod Record* 33(4), 65–70 (2004)
- [12] Wache, H., Voegele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., Hübner, S.: Ontology-based integration of information—a survey of existing approaches. In: *IJCAI 2001 Workshop: Ontologies and Information Sharing*, vol. 2001, pp. 108–117 (2001)
- [13] Bouzeghoub, M., Goble, C.A., Kashyap, V., Spaccapietra, S. (eds.): *ICSNW 2004*. LNCS, vol. 3226. Springer, Heidelberg (2004)
- [14] Xue, Y.: *Ontological View-driven Semantic Integration in Open Environments*. The University of Western Ontario (2010)
- [15] Pinto, H.S., Martins, J.P.: A methodology for ontology integration. In: *Proceedings of the 1st International Conference on Knowledge Capture*, pp. 131–138 (2001)
- [16] Kim, K.Y., Manley, D.G., Yang, H.: Ontology-based assembly design and information sharing for collaborative product development. *Computer-Aided Design* 38(12), 1233–1250 (2006)
- [17] Farquhar, A., Fikes, R., Rice, J.: Tools for assembling modular ontologies in Ontolingua. In: *Proc. of Fourteenth American Association for Artificial Intelligence Conference (AAAI 1997)*, pp. 436–441 (1997)
- [18] Heflin, J., Hendler, J.: Dynamic ontologies on the web. In: *Proceedings of the National Conference on Artificial Intelligence*, pp. 443–449 (2000)
- [19] Zimmermann, A., Krotzsch, M., Euzenat, J., Hitzler, P.: Formalizing ontology alignment and its operations with category theory. *Frontiers in Artificial Intelligence and Applications* 150, 277 (2006)
- [20] Noy, N.F., Crubézy, M., Ferguson, R.W., Knublauch, H., Tu, S.W., Vendetti, J., Musen, M.A.: Protégé-2000: An Open-Source Ontology-Development and Knowledge-Acquisition Environment: AMIA 2003 Open Source Expo. In: *AMIA Annual Symposium Proceedings*, vol. 2003, p. 953 (2003)
- [21] Fellbaum, C.: *WordNet. Theory and Applications of Ontology: Computer Applications*, 231–243 (2010)
- [22] Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.J.: Introduction to wordnet: An on-line lexical database*. *International Journal of Lexicography* 3(4), 235–244 (1990)
- [23] Pedersen, T., Patwardhan, S., Michelizzi, J.: WordNet: Similarity: measuring the relatedness of concepts. In: *Demonstration Papers at HLT-NAACL 2004*, pp. 38–41 (2004)
- [24] Pérez, J., Arenas, M., Gutierrez, C.: Semantics and Complexity of SPARQL. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) *ISWC 2006*. LNCS, vol. 4273, pp. 30–43. Springer, Heidelberg (2006)