Generic Multimodal Ontologies for Human-Agent Interaction

A. Braşoveanu, A. Manolescu, M.N. Spînu

Adrian Braşoveanu
Lucian Blaga Univeristy of Sibiu, Romania
E-mail: adrian.brasoveanu@gmail.com

Adriana Manolescu
Agora University, Oradea and R&D Agora Ltd.
Cercetare Dezvoltare Agora Oradea, Romania
E-mail: adrianamanolescu@gmail.com

Marian Nicu Spînu
Aurel Vlaicu University of Arad,
Faculty of Exact Sciences
Department of Mathematics-Informatics
Romania, 310330 Arad, 2 Elena Drăgoi

Abstract: Watching the evolution of the Semantic Web (SW) from its inception to these days we can easily observe that the main task the developers face while building it is to encode the human knowledge into ontologies and the human reasoning into dedicated reasoning engines. Now, the SW needs to have efficient mechanisms to access information by both humans and artificial agents. The most important tools in this context are ontologies. The last years have been dedicated to solving the infrastructure problems related to ontologies: ontology management, ontology matching, ontology adoption, but as time goes by and these problems are better understood the research interests in this area will surely shift towards the way in which agents will use them to communicate between them and with humans. Despite the fact that interface agents could be bilingual, it would be more efficient, safe and swift that they should use the same language to communicate with humans and with their peers. Since anthropocentric systems entail nowadays multimodal interfaces, it seems suitable to build multimodal ontologies. Generic ontologies are needed when dealing with uncertainty. Multimodal ontologies should be designed taking into account our way of thinking (mind maps, visual thinking, feedback, logic, emotions, etc.) and also the processes in which they would be involved (multimodal fusion and integration, error reduction, natural language processing, multimodal fission, etc.). By doing this it would be easier for us (and also fun) to use ontologies, but in the same time the communication with agents (and also agent to agent talk) would be enhanced. This is just one of our conclusions related to why building generic multimodal ontologies is very important for future semantic web applications.
Keywords: multimodal ontology, ontology matching, interface agents, Semantic Web, human-agent interaction

1 Introduction

The Knowledge Society (KS) is a society where information is the primary resource which can be consumed by both humans and machines. If we want to build such a society in a proper way we need different kinds of infrastructure: hardware, software, organizational, etc. SW and

agents represent only a small part of the large infrastructure needed in order to build the true KS.

SW ([1], [2], [3], and [4]) is one of those disruptive technologies which tend to be talked about years before their coming of age. One of the visions presented in [1] was that of agents replacing humans for simple everyday tasks like buying tickets for a concert or making appointments to the doctor. The main reason why this vision hasn't yet come to life is one that is now well understood and also explained in the article's revision [2]: encoding the human knowledge into ontologies and the human reasoning into dedicated reasoning engines is not an easy task. This process requires trans-disciplinary knowledge, dedicated tools and repositories, and advanced techniques from mathematics, logics and software. It is in fact an extremely difficult procedure which relies entirely on the cooperation between hundreds or thousands of organizations and different standards. Since the standardization processes take a long time even in these days and the time of adoption for new technologies is sometimes around 2-3 years at least, we should not be surprised that it will take a while until the SW reaches the critical mass.

Ontologies represent the key to a successful communication between human and agents if they are done right. We are only beginning to understand the implications of using the ontologies for the great tasks we assigned for them, but some problems like ontology management (versioning, change, tools and standards), ontology matching (finding correspondences between different ontologies) and the adoption of ontologies on large scale by developers and users proved to be quite challenging. Ontology dynamics is definitely a field on which we should keep an eye on. According to [30] there is still no clear winner in the process of ontology matching (in other words: a standard or a methodology with clear rules to match almost everything automatically or semi-automatically since sometimes humans will need to check the results). Therefore we should not be surprised, when reading a journal or conference proceeding, that most of the articles refer to these tasks rather than to the desired using of ontologies which is to give agents a way of understanding our world and reason about it. It is the way things should be: in order to build a functional system we always need to have its parts figured out. We should however not lose sight of the system we need to build and this is one of the purposes of this paper: to look at the current state of the art in several fields of study and see if we are heading in the right direction. In this context we will especially examine some problems related to the multimodal communication between human and agent and try to see how they are solved by using ontologies.

2   Rationale and Approach: Why Complicate Things and Use Generic Multimodal Ontologies?

First we need to clear one question: what is an ontology? Some answers to this (and also some examples of how to use ontologies) can be found in [12], [15], [16], [17], [22], [23] and [31]. The classic definition proposed by Gruber tells us that an ontology is "explicit specification of a conceptualization" [12]. This definition is examined and extended by many papers, most recently by Guarino, Oberle and Staab in [16] which also focuses on the importance of "shared explicit specifications" because without commiting to ontologies every agent would understand something else (they also take the opportunity to revise the semiotic triangle). Ontologies are us, Mika's thesis [23] is a simple yet powerful statement. It tells us that since we are the ones who design the ontologies they will only express what we want them to express and will sometimes be useless without the context in which they have been created.

The main problem when designing ontologies is to carefully choose the concepts within a domain and the relationships between them in such a way as the ontology to be well founded because "any ontology will always be less complete and less formal than it would be desirable in

theory" [16]. In the light of this statement it should become quite clear why we sometimes need to use generic ontologies: there is simply no other way to address the problem of uncertainty when developing ontologies than genericity.
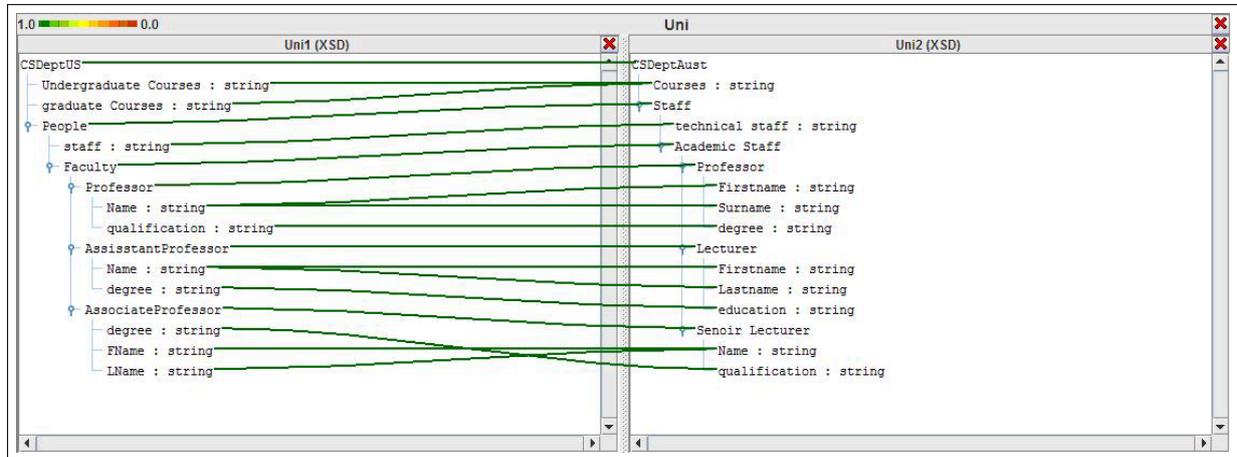

Figure 1: One of the most popular programs for ontology matching: COMA++, developed at the University of Leipzig. In this screenshot we can see how we can establish some correspondences between two ontologies representing a Computer Science Department

Nowadays there are probably thousands of ontologies in use, but if the SW will ever look like Berners-Lee's visions then ontologies will be common place for every designer, developer or user. Usually an ontology only addresses the problems from a narrow field of knowledge (domain ontology) so it is not uncommon that applications may use many ontologies for different purposes. In some of these cases it is useful to also use upper level ontologies which are general ontologies that represent concepts that are the same across all domains. A unique upper level ontology which should encompass all the human knowledge is not feasible and will never be built because of practical reasons (each society has its concepts, every field of knowledge has a certain language to protect itself, etc.), but upper level ontologies are used for mediation mainly in the idea that universal agreement between different ontologies will be/is possible. In other cases in order to use different ontologies the applications will use ontology matching schemes like those discussed in [10]. Since ontologies are the building blocks of SW, any application from this area must use them, even if that means adding layers of complexity because of the matching process, APIs, uncertainty. For everybody working in the IT industry these days it should be clear that the medium in which we work is becoming more and more like OHDUE (Open Heterogenous Dynamic Uncertain Environment) [8] and ontologies are part of this medium. These issues are addressed in articles and books like [10], [19], [30] (ontology matching), [26] (automatic generation of ontology APIs), [8] (OHDUE, agents). Because the field of ontology engineering is becoming more popular we should not be surprised that we will also hear a lot about the ontology driven software engineering. Ontology Driven Information Systems (ODIS) [36] is just one of the recent examples which fell into this category.

Given all these complications that appear when designing and working with ontologies it is interesting to ask a new question: why would we want to complicate our life even more by using multimodal ontologies? It is not enough that the ontology management or ontology matching problems still pose so many challenges? Are these new breed of ontologies even feasible?

Certainly from a user's perspective multiple modalities to enter input into a system (touch, voice, mouse, pen, etc.) can only mean increased usability (do we need to remember how touch screens became the norm in the mobile phones industry after iPhone was launched?), while from
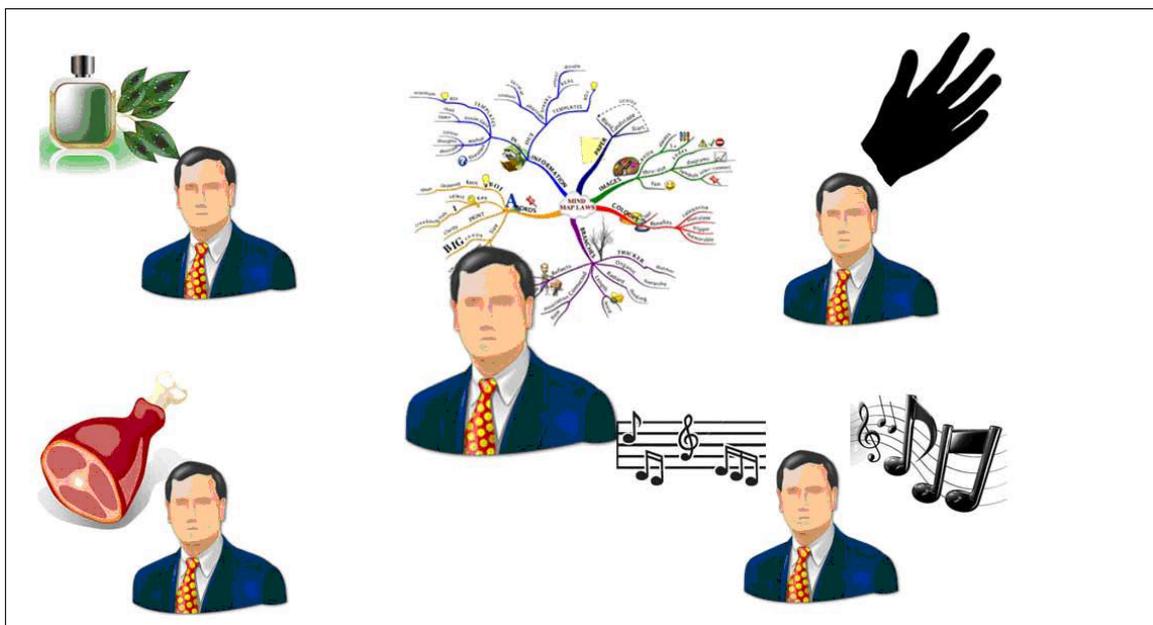
Figure 2: The multimodal communication dream: to use all the five senses (smell, sight, touch, taste, sound) during the process of communication.

a developer's perspective this means that software gets even more complicated than it is now. This is the right moment for such a development since for the multiple streams of data that come with multimodal communication we need distributed systems. Since multi-core processors are now luckily the norm in desktop computing we should have no problem (at least not hardware) dealing with the huge flux of data. In the past 40 years scientists have developed different mechanisms for getting audio, video and touch input, but the integration of all five senses in the communication between man and machine remains a dream. It is enough however to use one sense in different ways (for example for seeing we have images, text, video) to be able to speak about multimodal communication. In this respect different research groups (most notably [29]) started to develop also multimodal ontologies, but most of them took the approach of developing different ontologies for text, images, video or voice and then use ontology alignment to match them (multimodal integration through ontology matching [29]). A multimodal ontology gets us all the benefits of having such different ontologies. Like all things in life, multimodal ontologies do not come without bad parts (even harder to design, mantain and match), but they are definitely closer to our way of thinking. Is this a sufficient reason to try it? It might not be, but it is not the only one. The usage of multimodal ontologies will allow us to give a more natural, even realistic, feeling during communication between agents and humans, enhanced usability, the possibility to model mechanisms that are closer to the way we understand the world (diagrams, mind maps, feedback, brainstorming, slides, visual thinking, and others). It should be clear that it's not just art for art's sake, but rather art for a better life in the future.

3 Generic Multimodal Ontologies for Human-Agent Interaction

The process of multimodal ontology modeling is still open to exploratory research because ontologies are not everywhere. Without ontologies for all possible fields, and tools to match these ontologies it is debatable whether we will achieve an efficient semantic web, but rather the illusion of a semantic web maintained by few successful applications in certain areas (like social

networking, language translation or medicine). Since multimodal communication is difficult to process it is clear that in the first phase of any research regarding this subject, the communication between agents and humans will not be efficient. The question we need to ask ourselves in this situation is: If it is not efficient why should we bother at all to try something like this? The answer is simple and is typical for exploratory research: It takes time to find the best way to integrate multiple streams of data in an efficient manner and it also takes time to develop efficient ontology matching processes for such tasks. The role of exploratory research is to discover niches. The task of creating efficient mechanisms is one best suited for incremental research. Since this area of research is relatively new there is enough room for exploratory research and for breakthroughs.

Generic ontologies are rarely used by developers. Most of the articles present different ontologies and clearly state that they do not use generic ontologies because the problem's domain was well understood. Generic ontologies are best suited for modelling as we can see from [17], and [13]. It is easier to say you have an ontology with few concepts and not define all of them when doing modelling. The task of defining all the concepts and relationships between them is one that remains to the ontology engineer or to the developer. When dealing with models that are related to multimodal communication it makes sense to use generic multimodal ontologies. It also makes sense to use a generic ontology whenever dealing with uncertainty as suggested by [8] [28].

The agents of tomorrow will be built taking into account recent findings like the requirements-driven self-reconfiguration [6], multi-party, multi-issue, multi-strategy negotiation [35], natural language [18], and controlled natural language [32]. If we are to follow Berners-Lee vision from [1] we absolutely need to integrate such findings into our work. In fact according to [18] ontologies are the "common ground for virtual humans". Their architecture suggests using multimodal communication, but this is not clearly stated in the article since the ontology is not multimodal. If we look at [6] and [35] we can envision agents that dynamically change their strategies according to the environment and the context of conversations. This requires designing flexible ontologies, another reason to make them generic.

The agents must use ontologies if they are to understand something from this world. They also need to share them and commit to them if we want them to be able to talk between them. The multimodal ontology helps in some of the phases of multimodal communication: fusion and integration (getting the input from different channels), natural language processing, disambiguation, error reduction and fission (preparing the output). When designing a multimodal ontology one must also take into account the problems related to designing multimodal systems as described in [25], and also the medium in which these agents will evolve because an agent that needs to evolve in the urban computing environment [34] will have different needs than an agent that just surfs the web. The focus of research is usually on multimodal fusion, but a recent survey [9] shows that the interest in multimedia fission is increasing. Designing a multimodal ontology thus requires taking into account all these findings because the agent must be able to give us a response not only to understand our requirements. Probably one of the big challenges ahead is to annotate the multimodal content in real-time. This is particularly hard to do for video content, but not impossible, as [27] suggests. M3O (Multimedia Metadata Ontology) allows us to annotate the multimedia content from a page to retrieve it easier. If such ontologies will be improved then the road to the visions from [1] will be shorter.

4   Related Work

The current state of the art in multimodal HCI is presented in [7] and [20]. One of the conclusions from [7] leaves further space for improvements: "most researchers process each channel (visual, audio) independently, and multimodal fusion is still in its infancy". The same can be

rendered as true for the multimodal ontologies too. Since [7] is more recent we will use it as a basis for further investigation in this field.

Since there are only few interesting articles related to multimodal ontologies every year, we have selected a few of them to be used as basis for future research.

When searching for definitions related to ontologies and trends in the field of ontology development /matching some of the best research groups in the world are the ones from Trento (LOA and University of Trento), and Koblenz-Landau. Many of the articles cited in this paper come from some of the members of the Trento group: [6], [10], [16], [17], [30]. These are related to definitions of ontology, ontology matching, and modelling with ontologies. We have also used articles from the Koblenz-Landau group: [6], [26], [27] related to definitions, automatic generations of ontology APIs and M3O.

One interesting idea is that of multimodal context-aware interaction presented by Cearreta and his team in [5]. If we have to model emotions there might be no other solution than to use multimodal ontologies combined with special reasoners. Another article related to our subject is [29]. Their approach of using different ontologies for text and images and then use ontology matching can definitely be improved on the long term. They clearly state that for the moment multimodal ontology do not offer fast communication, but that in time speed might be improved. Also [24], [32] and [33] study the relationships between Natural Language Processing (NLP) and SW. The work of these research groups must be studied. One of them [32] is from Southampton, one of the workplaces of Timothy Berners-Lee.

When it comes to generic ontologies and tools for working with ontologies, one of the best research groups that needs to be followed is Stanford's [11], [28]. Their work on biomedicine ontologies and Protégé is fundamental.

5   Conclusions and Future Work

The SW tools are now an important part of the IT industry, the main clients coming from the fields of biomedicine, aeronautics, automotive, government and local administrations, and media. This sudden interest might be related to the success of social media [14], [21] and means that developers are starting to tap into the potential promises of the field. Even so there is a lot of work to be done regarding multimodal ontologies. The reason is one that was mentioned several times during this paper: the task of designing such ontologies is still difficult. As we do not have yet universal methods for ontology matching we do not have a clear methodology of designing multimodal ontologies (regardless of the fact that they are generic or not).

The main advantages of using generic multimodal ontologies should be better understood now: they offer us a modality to design the process of communication with agents as close to our way of thinking as possible and also play a very important role in several phases of the multimodal communication (multimodal fusion and integration, disambiguation, NLP, error reduction, multimodal fission, etc.). The main disadvantage will probably be efficiency for the next years, but given the exploratory nature of the research this is normal.

The future work of our group will consider implementing new mechanisms for linking the generic multimodal ontologies and affective interfaces with recent research in Semantic Web and

HCI in a 3 years interval (during the PhD studies of the first author). The objectives are to be fulfilled involving European teams of researchers interested in this kind of projects.

Bibliography

[1] T. Berners-Lee, J. Hendler, O. Lassila. *The Semantic Web.* Scientific American, May 2001, 34-43.

[2] N. Shadbolt, W. Hall, T. Berners-Lee. *The Semantic Web revisited.* IEEE Intelligent Systems, pages 96- 101, May/June 2006.

[3] T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt, D.J. Weitzner. *A Framework for Web Science.* Foundations and Trends in Web Science, 1 (1), pages 1-130, 2006.

[4] C. Bizer, T. Heath, T. Berners-Lee. *Linked Data - The Story So Far.* International Journal on Semantic Web and Information Systems, Volume 5, Issue 3.

[5] I. Cearreta, J. M. Lopez, N. Garay-Vitoria. *Modelling multimodal context-aware affective interaction.* Proceedings of the Doctoral Consortium of the Second international conference on ACII'07. Lisbon, Portugal, 57-64, 2007.

[6] F. Dalpiaz, P. Giorgini, J. Mylopoulos. *An Architecture for Requirements-driven Self-Reconfiguration.* Proc. of the 21st Int. Conf. on Advanced Information Systems Engineering, LNCS 5565, Springer, 246- 260, http://www.disi.unitn.it/ pgiorgio/papers/caise09-b.pdf, 2009.

[7] B. Dumas, D. Lalanne, S. Oviatt. *Multimodal Interfaces: A Survey of Principles, Models and Frameworks.* In D. Lalame, J. Kohlas, editors, Human Machine Interaction Research Results of the MMI Program, Springer, 3-27, 2009.

[8] I. Dzitac, B.E. Barbat. Artificial Intelligence + Distributed Systems = Agents. *International Journal of Computers, Communications & Control*, ISSN 1841-9836, 4(1):17-26, 2009.

[9] D.W. Embley, A. Zitzelberger. Theoretical Foundations for Enabling a Web of Knowledge. Retrieved from: http://dithers.cs.byu.edu/tango/papers/formalWoK.pdf, 2009.

[10] J. Euzenat, P. Shvaiko. *Ontology Matching.* Springer, 2007

[11] A. Ghazvinian, N. F. Noy, C. Jonquet, N. H. Shah, M. A. Musen. *What Four Million Mappings Can Tell You about Two Hundred Ontologies.* International Semantic Web Conference 2009: 229-242

[12] T. R. Gruber. *A Translation Approach to Portable Ontologies.* Knowledge Acquisition, 5(2):199- 220, 1993.

[13] M. Gruninger. Designing and Evaluating Generic Ontologies. In 'ECAI96's workshop on Ontological Engineering'.

[14] T. Gruber. *Collective knowledge systems: Where the social web meets the semantic web.* Journal of Web Semantics, 6(1):4-13, 2008.

[15] N. Guarino. *The Ontological Level: Revisiting 30 Years of Knowledge Representation.* In A. Borgida, V. Chaudhri, P. Giorgini, E. Yu (eds.), Conceptual Modelling: Foundations and Applications, Springer Verlag 2009: 52-67.

[16] N. Guarino, D. Oberle, S. Staab. *What is an Ontology?* In S. Staab and R. Studer (eds.), Handbook on Ontologies, Second Edition. International handbooks on information systems. Springer Verlag: 1-17, 2009.

[17] G. Guizzardi, T. Halpin. *Ontological foundations for conceptual modeling.* Applied Ontology 3, 1- 12, 2008.

[18] A. Hartholt, T. Russ, D. Traum, E. Hovy, S. Robinson. *A common ground for virtual humans: Using an ontology in a natural language oriented virtual human architecture.* In: Language Resources and Evaluation Conference (LREC). (May 2008)

[19] W. Hu, Y. Qu. Falcon-AO: *A practical ontology matching system.* Web Semantics: Science, Services and Agents on the World Wide Web 6 (2008) 237-239.

[20] A. Jaimez, N. Sebe. *Multimodal human-computer interaction: A survey.* Computer Vision and Image Understanding, Volume 108, Issues 1-2, October-November 2007, 116-134, Special Issue on Vision for Human-Computer Interaction, 2007.

[21] F. Limpens, F.Gandon, and M. Buffa. *Linking folksonomies and ontologies for supporting knowledge sharing: a state of the art.* Technical report, EU Project, ISICIL, 2009.

[22] D. Lonsdale, D. W. Embley, Y. Ding, L. Xu, M. Hepp. *Reusing Ontologies and Language Components for Ontology Generation, accepted for publication in Data and Knowledge Engineering.* Retrieved from: http://www.heppnetz.de/files/dke2008.pdf, 2010.

[23] P. Mika. *Social Networks and The Semantic Web,* Springer, 2007

[24] J. Niekrasz and M. Purver. *A multimodal discourse ontology for meeting understanding.* In The 2nd Joint Workshop on Multimodal Interaction and Related, 2005.

[25] L. Nigay, J. Coutaz. *A design space for multimodal systems: Concurrent processing and data fusion.* ACM Conf. Human Factors in Computing Systems (CHI), 1993.

[26] F. S. Parreiras, C. Saathoff, T. Walter, T. Franz, S. Staab. APIs a gogo: Automatic Generation of Ontology APIs. icsc, 342-348, 2009 IEEE International Conference on Semantic Computing, 2009

[27] C. Saathoff, A. Scherp. M3O: *The Multimedia Metadata Ontology.* Proceedings of the Workshop on Semantic Multimedia Database Technologies, 10th International Workshop of the Multimedia Metadata Community (SeMuDaTe 2009), Graz, Austria, 2009.

[28] Abraham Sebastian, Natalya Fridman Noy, Tania Tudorache, and Mark A. Musen. *A generic ontology for collaborative ontology-development workflows.* In Aldo Gangemi and Jérôme Euzenat, editors, EKAW, volume 5268 of Lecture Notes in Computer Science, 318-328. Springer, 2008.

[29] A.A.A. Shareha, M. Rajeswari, D. Ramachandram. *Multimodal Integration (Image and Text) Using Ontology Alignment.* American Journal of Applied Sciences 6 (6): 1217-1224, 2009.

[30] P. Shvaiko, J. Euzenat. Ten challenges for ontology matching. In Proceedings of the 7th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE), pages 1164-1182, Monterrey (MX), 2008.

[31] W. V. Siricharoen. *Ontology Modeling and Object Modeling in Software Engineering.* International Journal of Software Engineering and Its Applications, Vol. 3, No. 1, January, 2009, 43-59, 2009.

[32] P. Smart, J. Bao, D. Braines, N. Shadbolt. *Development of a Controlled Natural Language Interface for Semantic MediaWiki.* In: Proceedings of the Workshop on Controlled Natural Language, Springer- Verlag, Heidelberg, Germany.

[33] D. Sonntag, M. Romanelli. *A multimodal result ontology for integrated semantic web dialogue applications.* In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006), Genova, Italy, May 24-26.

[34] A. Tenschert, M. Assel, A. Cheptsov, G. Gallizo, E. Della Valle, I. Celino. *Parallelization and Distribution Techniques for Ontology Matching in Urban Computing Environments.* OM 2009

[35] D. Traum, S. Marsella, J. Gratch, J. Lee, and A. Hartholt.. *Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents.* In Proc. of Intelligent Virtual Agents Conference IVA-2008.

[36] M. Uschold. *Ontology-Driven Information Systems: Past, Present and Future.* In Proceedings of the 5thInternational Conference on Formal Ontology in Information Systems (FOIS2008), Saarbrücken, Germany, (Oct 31st - Nov 3rd), 2008.