Automatic Interchange of Knowledge between Business Ontologies

Alma Delia Cuevas Rasgado[1] , Adolfo Guzmán Arenas[2]

[1] Instituto Tecnológico de Oaxaca Av. Ing Victor Bravo Ahuja 125 Calzada Tecnológico C.P. 68036 Oaxaca, México almadeliacuevas@gmail.com

[2] Centro de Investigación en Computación Av. Juan de Dios Batiz casi esq. Miguel Othon de Mendizabal Unidad Profesional Adolfo Lopez Mateos, Edificio CIC. Col. Nueva Industrial Vallejo, Delegacion Gustavo A. Madero, C. P. 07738, Mexico D. F. a.guzman@acm.org

Abstract. A person adds new knowledge to his/her mind, taking into account new information, additional details, better precision, synonyms, homonyms, redundancies, apparent contradictions, and inconsistencies between what he/she knows and new knowledge that he/she acquires. This way, he/she incrementally acquires information keeping at all times it consistent. This information can be perfectly represented by Ontologies. In contrast to human approach, algorithms of Ontologies fusion lack these features, merely being computer-aided editors where a person solves the details and inconsistencies. This article presents a method of Ontology Merging (OM), its algorithm and implementation to fuse or join two Ontologies (obtained from Web documents) in an automatic fashion (without human intervention), producing a third ontology, and taking into account the inconsistencies, contradictions, and redundancies between both Ontologies, thus delivering a result close to reality. OM produces better results, when they are compared against fusions manually carried out. The repeated use of OM allows acquisition of much more information about the same topic.

Keywords: Ontology, Artificial Intelligence, Knowledge Representation, Semantic Web, Ontology Fusion.

1 Introduction

A person acquires information across his/her life by adding new knowledge (concepts, relations, typical values…) to the information he/she has in his/her mind (formally saying, in his/her ontology or knowledge structure), identifying redundancies, new information, small and large contradictions, synonymous and antonymous among others things. Nowadays, computers could do the same process (joining knowledge which comes from two different Ontologies) through an editor that makes preliminary alignment of concepts, and lets to a person the final decision. It is so called a computer-aided fusion. The problem to solve is how to mechanize that fusion.

This article presents an algorithm of Ontology Merging (OM) and its implementation to automatically fuse two Ontologies, obtaining a third one, and taking into account inconsistencies, synonymous, precision rates, contradictions and redundancies between them, in such manner that the obtained results are close to reality. The resulting ontology, actual knowledge ontology, can become quite useful if it is the fusion of much general and specific Ontology. The article summing up the Ph. D. thesis [5].

The rest of introduction describes some information management tasks and previous work. §2 contains some preliminary definitions related to OM. Some results produced by OM are shown in §3. §4 outlines the future work on how to enrich OM. In another paper we plan to describe a system (called OM*) that takes texts documents and converts them [16] to Ontologies, after which OM will join them to generate larger and larger Ontologies containing detailed knowledge about a given topic area. This large ontology (yet to be built) can be used, for instance, to answer difficult questions like in [2]. OM resembles CYC [18], in that both pursue to produce a large common knowledge ontology. CYC envisions to manually build such large ontology, while OM can do it mechanically.

1.1 Information Management

Internet contains huge amounts of information in billions documents located in Web sites, texts documents, doorway services, music blogs, photographic maps, etc. When we access them through search engines (Google, Cite Seer,…), only a small portion of the available information is recovered, because the search is performed in a syntactic form by labels, keywords, and phrases that is, by lexicographic comparisons. Moreover, the answer is a large list of documents that do not always contain the information sought. In addition, the desired information should be deduced or extracted manually processing each document (that is, by *reading* them) by a person.

If a large structure of knowledge about a given topic could be found in the Web (as an ontology, for instance; cf. §2.1), then an alternative form to obtain a desired complex information would be to query (by an “intelligent” query) such ontology. Sure this will be less painful than the actual procedure. To achieve this, two tools should be constructed: one is to smartly join small Ontologies into larger ones; another one is to state intelligent queries to a complex or large ontology.

This article is focused on the first of these tools: an automatic knowledge fuser. This fusion should consider not only the syntax of the word and phrases (contained in the description of the concepts forming the Ontologies to be fused), but also their semantics too (the neighboring words or concepts, synonyms, homonyms, and so on).

1.2 Related works

In contrast to the manual generation of Ontologies [18], OM performs such generation without the user intervention, obtaining pieces of knowledge (small Ontologies) and joining them carefully (verifying inconsistencies, joining synonymous, etc.).

Currently, ontology joiners are either manual or computer-aided as shown in the following. Also, OM is not only tuned up to a special or specific knowledge areas, but also can be used to merge Ontologies in any knowledge area (perhaps after modifying its initial knowledge basis; see §2.3).

The encyclopedias such as Wikipedia or Encyclopedia Britannica contain knowledge held in written documents, inserted and related to each other by hand. Inconsistency and contradiction among documents are controlled by restricting the publishers (inserters) of the documents and by a "final authority" (the Editor). In contrast to this, Ontologies can be produced electronically by repeated use of OM. Relations (among the *concepts* of ontology) are represented and merged by OM (it also resolves inconsistencies and contradictions).

Known methods of Ontologies fusion are computer-aided, but not fully automated processes. PROMPT [8], Chimaera [14], OntoMerge [6] and IF-Map require that a user solve problems presented during the fusion. Others such as FCA-Merge [19] use Formal Concept Analysis for the representation of their Ontologies, forcing them to be mutually consistent. But the majority of Ontologies in Web present inconsistencies when compared to other Ontologies. A recent fuser is HCONE-merge [13], which uses the semantic data base WordNet [7] as intermediary information for the fusion, requiring less user support that is an important advantage in computer-aided ontology fusion.

If each element of knowledge in Internet might have translated (located, placed) into a (piece inside of) ontology, then knowledge structuring would be more efficient for computer search. Consequently, OM will find its work easier too. For example, a comprehensive ontology about Albert Einstein's life is obtained from 50 biographies. Now these extensive descriptions have to be built manually. It is our hope that with OM's help these large Ontologies could be automatically built.

2 OM elements

The behavior of OM is better perceived through examples. To this end, a document collection has been recovered from Web sites. From this, the used Ontologies have been manually generated, but their fusion has been entirely constructed by OM.

2.1 Ontology definition

In Computer Science, Ontology is a data structure, a notation used to share and reuse knowledge between Artificial Intelligence systems. Thus, ontology is a set of definitions, classes, relations, functions, and other objects from the same context [5].

There is no a mathematically rigorous theory, which formally defines ontology. However, some intents can be found in [12]. From the Logics viewpoint, ontology is a pair:

(1) $$O = (Ç, R),$$

where:
Ç is a set of nodes (representing concepts), some of which are relations.
R is a set of restrictions, of the form (r; c_1; c_2;…; c_k) between the relation r and concepts c_1 until c_k (lower c is used to refer to each concepts of set Ç, while a semicolon separates members in the restrictions). For example, (cut; scissor; sheet), (print; printer; document; ink). In these examples, the concepts are relations too (cut and print). The restrictions are not limited to have two members besides the relation. Therefore, ontology is a hypergraph having as Ç the set of nodes and as R the set of hyper-relations.

2.2 Ontology fusion using OM

OM is a totally automatic algorithm that can join Ontologies with inconsistencies [4]. In the process two Ontologies A and B are fused to form a third ontology C. In general, we have:

$$C = A \cup \{c_C , r_B \mid c_C , r_B = ext(r_A, r_B) \ \forall \ c_A \in A \} \quad \textbf{(1)}$$

The resulting ontology C is the original ontology A plus some concepts and relations of B that the function *ext* produces.
In (1):
c_A is a concept in ontology A, r_A are the relations of c_A presented in A; r_B are relations of c_B that exist in B; c_B is the most similar concept (*cms*) to c_A in B; while $c_c \in C$ as explains in the following.
$\cup$ means ontology joining (different to the set union operation).
ext(r_A, r_B) is the algorithm that completes the relations r_B, which are not in A with those c_C (which are in B), and which do not contradict knowledge from A. That is, for each node $c_A \in$ A, all its relations are retained in C, and only *some* relations of c_B (*cms* to c_A in B) are added to C, as well as their "target" concept c_C. For instance, if the restriction (r_B c_B c_C) = (religion; Juárez; catholic) is in B, and it does not contradict knowledge from A, then restriction (r_A c_B c_C) = (religion; Benito Juárez; catholic) is added to r_A in C; here we assume that c_A = Benito Juárez and its most similar concept in B is r_B = Juárez. This will become clearer in the following examples. The algorithm *ext* is explained in §3.2 to §3.6.

2.3 Initial knowledge used by OM

OM is supported by some initial (built-in) knowledge bases and resources that help it to detect contradictions, to find synonymous, etc. These resources are:

1.- *Hierarchies of concepts*. A hierarchy is a tree of concepts where each node is a concept or a set; if it is a set, then its subsets are a partition of it. The hierarchy represents taxonomy of related terms. It is used to compute the concepts *confusion* [10]. We exploit it to detect synonyms and "false inconsistencies" due to degree of detail.

2.- *Articles and linking words* such as (*in*, *the*, *to*, *this*, *and*, *or*, etc.) that are ignored in the name or description of the concept.
3.- *Words that change or reject concepts in the relation's name* such as: *except*, *without*. For example, *Poppy without Petiole*. This means that the concept *Petiole does not form part of* the concept *Poppy*.

3 How OM fuses Ontologies

In this section, the fundamentals of OM are explained.

3.1 General description

To fuse Ontologies A and B into a resulting ontology C, OM performs the following steps:

1. Copy ontology A into C.
2. Starting from root concept c_{Root} in C:
3. Look at B for its *cms* c_B (using the COM algorithm); see in the following.
4. If there is a *cms* in B, new relations of the *cms* in B can be added to C as well as new concepts as follows:
   A. Redundant relations are verified and rejected (not copied to C; explanation and example in [5]);
   B. Subsets become partitions (§3.3), if appropriate;
   C. Synonyms are verified and properly fused (§3.2);
   D. Partitions from B that are not in A are added to C [5] when suitable;
   E. Homonyms are detected and handled correctly [5];
   F. Some inconsistencies are detected and solved, using the confusion theory; see in the following.
5. If there is not a *cms*, then take the next concept c_C depth-first and go back to step 3.

If in step 4 inconsistencies are not solved, then the relation prevailing in A is conserved in C (the conflicting relation in B is discarded); see also [5].

OM is supported by two important recent developments:

- The comparer of distinct Ontologies COM that considers a concept c_A in ontology A and looks for the most similar concept c_B in ontology B. COM pays attention both to the definition of c_A (in words or word phrases) as well as its relations to other nodes in A, and the same for the candidate c_B in B [11].
- Confusion Theory [10], that obtains the *confusion* (a number between 0 and 1) when concept *r* is used instead of concept *s.* We use it to properly handle redundancies (for instance, "*he was born in Mexico*" versus "*he was born in Guelatao*").

Finally, if the inconsistency can not be solved, then OM prefers the knowledge of A.

3.2 Identification and merging of synonyms

In this example (Fig. 1), a company sells oil (ontology A [20]), while ontology B represents an agent that requests information about A. The resulting ontology C will allow the base knowledge of the company to understand these requests. Synonymy is detected through the word definition `Ingredient` (item, ingredient) in A. Therefore, the `Item` partition will not be copied into the resulting ontology C. The same process of identification of synonym concepts is applied to identification of synonym relations.

3.3 Promoting subset to partitions

In ontology A [22] (Fig. 1), COM finds that the concept's `Hotel Finca Santa Marta` *cms* in B is the concept `Finca Santa Marta` [23], which has the partition *Hotel Amenities.* The members of this partition are precisely the same as the parts of (part of) `Hotel Finca Santa Marta` in A. Since that condition holds, OM selects for the resulting ontology C the relation partition *Hotel Amenities,* which has a more precise meaning than the corresponding knowledge in A.

4 Results

Table 1 presents some results. A and B, Ontologies to be fused have been manually constructed from different Internet documents. Each pair of Ontologies to be joined describes the same topic. For example, both Ontologies about `Turtles` have built from documents found in different Web sites. Each ontology describes `Turtle`. The Ontologies thus obtained have been completely joined by OM. The results produced by OM have been verified (compared) against a result obtained by manual fusion of the former Ontologies A and B. The OM's results in general are quite good.

The first column of Table 1 presents tested Ontologies and the time that OM took to fusion them. The slowest fusion is on "One hundred years of loneliness" because being a fiction novel, it has very rich semantics as well as the relations between concepts. Thus, OM carefully verifies elements of each relation following the algorithms presented in this article. The second column shows the sum of number of relations of A and B that have been fused. The third column marks the sum of number of fused concepts. In the same column, we can observe the result of the manual fusion compared to automatic fusion by OM. In some cases the results were different. The fourth one shows the numeric error calculate as follows: *numer of relations and concepts wrongs copied on C/total number of relations and concepts copied manually on C* and in the last one, the efficiency of OM, computed as follows: *number of relations and concepts correctly copied on C/total number of relations and concepts copied manually on C*.

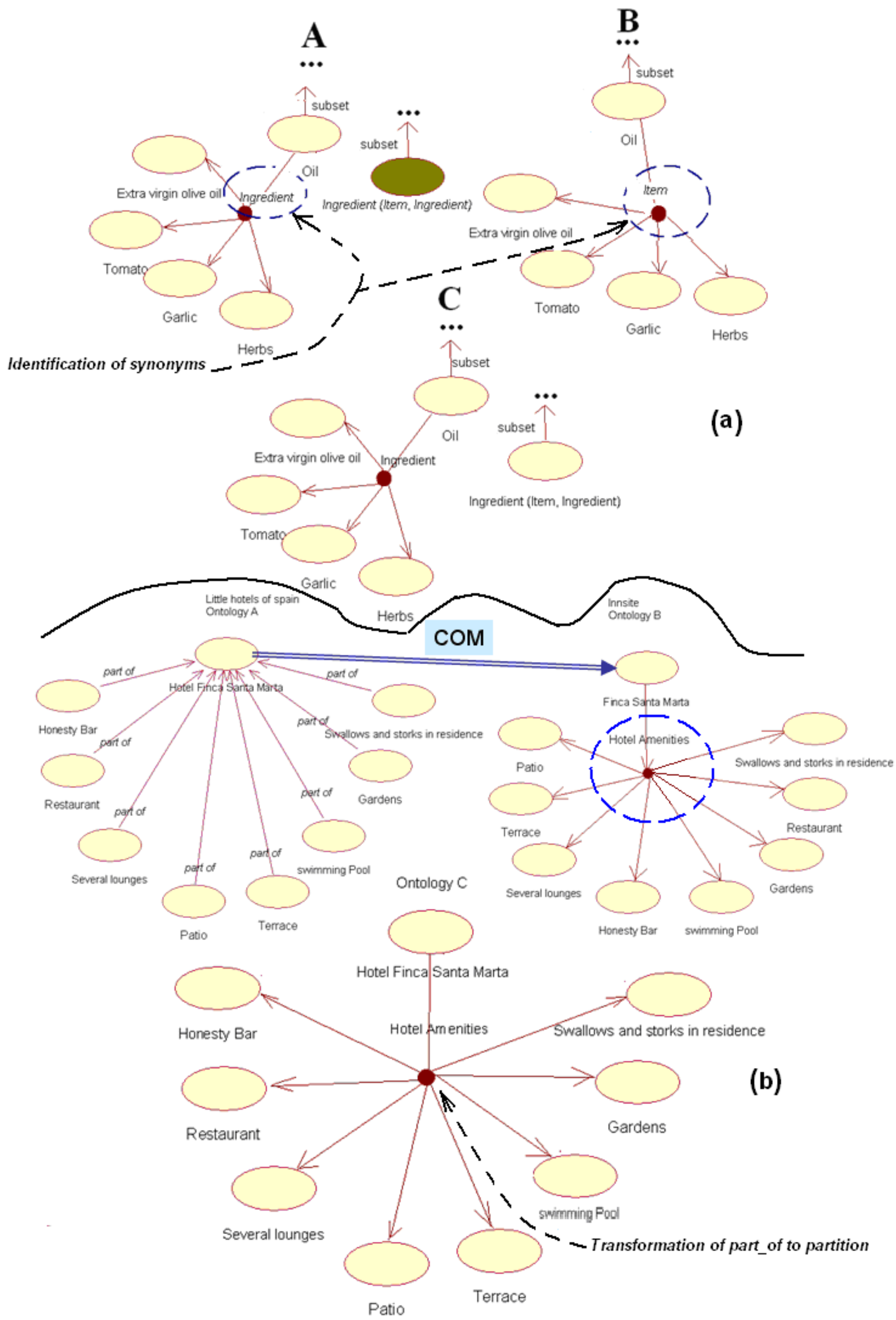

Fig. 1. (a) In ontologies A and B, Ingredient and Item concepts are identified as synonyms. C is the result of fusion A and B. (b) In B, partition Hotel Amenities of Hotel Finca Santa can be added into resulting ontology C, replacing the former part of structure.

5 Discussion

Existing ontology fusion methods (except HCONE-Merge [13], which uses WordNet to find the meanings of concepts to be aligned) share two features: (A) the aligning and fusion are done by (syntactically) comparing names and label of concepts and their neighborhood; and (B) they resort to human intervention for final acceptance of their suggestions. Thus, they are computer-aided fusers.

Our method called herein OM uses the definition of the concepts, their neighborhood, and their characteristics (relations, that is, restrictions in which they participate). These restrictions are verified by a recursive process, since each of them can be also a concept or point to a concept (in the sense that (religion `Juárez catholic`) points to concept `catholic`), which suffers the same OM verification before fusion. This recursive process can be interpreted as a semantic search in the concepts (called semantic analysis as well). That is, all possible knowledge in A and B about the concept to be merged into resulting C is taken into account. Thus, this version of "semantic analysis" has more possibilities, as OM shows, than the usual syntactic analysis or matching. Note also that usually the fusion is checked manually against a hand-made computer result. Automation of this procedure is not straightforward, since the "right" fusion is very subjective.

Table 1. Results of using OM in some examples from real cases.

Source Ontologies A and B	**Relations in C**	**Concepts in C**	**Error**	**Effic.**
Solar System (4 sec.)	45 relations of B [28] were added and fused correctly into 56 of A [27], obtaining a total of 59 relations on C	60 concepts of B were added and fused correctly into 79 of A, obtaining a total of 125 concepts on C	0	100
Neurotransmission and Schizophrenia (2 sec.)	79 relations of Neurotransmission (A) [24] were added and fusion correctly into 51 of Schizophrenia (B) [26] obtaining a total of 127 relations on C. The manual method gave 129 (2 of 129 were not copied)	56 concepts of Neurotransmission (A) were added and fused correctly into 26 of Schizophrenia obtaining a total of 77 nodes. The manual method gave 79 (2 of 79 concepts were missed)	0.019	98
100 Years of loneliness (10 minutes)	283 relations of B [30] were added and correctly fused into 231 of A [29] obtaining a total of 420. The manual method gave 432 (12 of 432 were not copied).	126 concepts of B were added and correctly fused into 90 of A obtaining a total of 141. The manual method gave 149 (8 of 149 concepts were missed).	0.034	96.5

6 Future work

OM could be completed (let us call OM* this complete version) with a pair of additional tools:

- The parser or converter of texts documents into Ontologies [16]. It can be called a "pre-processor" of OM to (automatically) produce the Ontologies that OM fuses (work in progress).
- The question-answering program, mentioned in §5.1, which will allow to exploit or to use a practical purpose of knowledge that OM* join (work in progress).

7 Conclusions

Taking the advantage of OM, we can fuse two Ontologies automatically. The progress made can be gauged from the quality of the obtained results (Table 1). OM detects and solves some inconsistencies, detects synonymous, homonymous, redundant information, and different degrees of detail or precision.

Missing topics are: a) an analyzer that converts documents from natural language into ontology, and b) a question-answerer (using the resulting ontology of OM) to answer difficult (i.e., "intelligent" or "tough") questions, which we are considering as a part of the future work.

Acknowledgments. The authors recognize the partial support of CONACYT and SNI. Prof. Sergei Levachkine (CIC-IPN) provided useful suggestions.

References

1. BanerJee, S., and Pedersen, T. . Extended Gloss Overlaps as Measure of Semantic Relatedness. In Proc. of *IJCA-03*, pp. 805-810. México. (2003).
2. Botello, A. Infiriendo Relaciones Entre Bases of Datos Autónomas. CIC-IPN. Tesis Doctoral en proceso of desarrollo. (2007).
3. Cuevas, A., and Guzmán, A. Improving the Search for the Most Similar Concept in other Ontology. In proc. *XVIII Congreso Nacional y IV Congreso Internacional of Informática y Computación*. Torreón Coah. México. (2005).
4. Cuevas, A., and Guzman, A. A Language and Algorithm for Automatic Merging of Ontologies. *Handbook of Ontologies for Business Interactions*, ed. Peter Rittgen, 383-406 USA.: Idea Group Inc, Publishers. Hershey, PA, USA.
5. Cuevas, A. Unión of ontologías usando propiedades semánticas. Ph. D. thesis. CIC-IPN. México. (2006). http://148.204.20.100:8080/bibliodigital/ShowObject.jsp?idobject=34274&idrepositorio=2&type=recipiente
6. Dou, D., McDermott, D., and Qi, P. Ontology Translation by Ontology Merging and Automated Reasoning. In Proc. *EKAW Workshop on Ontologies for Multi-Agent Systems*. (2002).

7. Fellbaum, C. WordNet An Electronic Lexical Database. Library of Congress Cataloging in Publication Data. (1999).
8. Fridman, N., and Musen, M. PROMPT: Algoritm and Tool for Automated Ontology Merging and Alignment. In *Proc. Seventeenth National Conference on Artificial Intelligence*. pp 450-455, Austin, TX, USA. (2000).
9. Gruber, T. Toward principles for the design of ontologies used knowledge sharing. Originally in N. Guarino & R. Poli, (Eds.), *International Workshop onf Formal Ontology*, Padova, Italy. (1993).
10. Guzmán, A., and Levachkine, S. Hierarchies Measuring Qualitative Variables. *Lecture Notes in Computer Science* LNCS 2945 [Computational Linguistics and Intelligent Text Processing], Springer-Verlag. 262-274. ISSN 0372-9743. (2004).
11. Guzmán, A., and Olivares, J. Finding the Most Similar Concepts in two Different Ontologies. *Lecture Notes in Artificial Intelligence* LNAI 2972, Springer-Verlag. 129-138. ISSN 0302-9743. (2004).
12. Kalfoglou, Y., and Schorlemmer, M. Information-Flow-based Ontology Mapping. Proceedings of the 1st *International Conference on Ontologies, Databases and Application of Semantics* (ODBASE'02), Irvine, CA, USA. (2002)
13. Kotis, K., and Vouros, G., Stergiou, K. Towards Automatic of Domain Ontologies: The HCONE-merge approach. *Elsevier's Journal of Web Semantic (JWS),* vol. 4:1, pp 60-79. Available on line and (ScienceDirect): http://authors.elsevier.com/sd/article/S1570826805000259 (2006).
14. McGuinness, D., Fikes, R., Rice, J., and Wilder, S. The Chimaera Ontology Environment Knowledge. In *Proceedings of the Eighth International Conference on Conceptual Structures Logical, Linguistic, and Computational Issues (ICCS 2000)*. Darmstadt, Germany. (2000).
15. Montes of Oca, V., Torres, M., Levachkine, S. and Moreno, M. Spatial Data Description by Means of Knowledge-Based System. *Lecture Notes in Computer Science*, Vol. 4225, Springer-Verlag (2006).
16. Nery, P. Parser for the conversion of text documents into ontologies. Thesis in progress. CIC-IPN. México. (2007).
17. Puscasu, G., Ramírez Barco P, *et al*. On the identification of temporal clauses. *LNAI 4293*, 911-921 MICAI 06. (2006).
18. Reed, S. L., and Lenat, D. Mapping Ontologies into Cyc. In proceeding of *AAAI Workshop on Ontologies and the Semantic Web*, Edmonton, Canada. (2002).
19. Stumme, G., Maedche, A. Ontology Merging for Federated ontologies on the semantic web. In: E. Franconi, K. Barker, D. Calvanese (Eds.): Proc. *Intl. Workshop on Foundations of Models for Information Integration (FMII'01)*, Viterbo, Italy, 2001. lNAI, Springer (2002).

Web Sites:

20. Oil http://www.rebost.com/empresadef.php3?id_idioma=3
21. Finca Santa Martha Hotel Ontology A http://www.littlehotelsofspain.co.uk/santamarta.php
22. Finca Santa Marta Hotel Ontology B http://www.innsite.com/inns/A004065.html
23. Loom http://www.isi.edu/isd/LOOM/LOOM-HOME.html
24. Neurotransmission es.wikipedia.org/wiki/Neurotransmisor
25. Ontology on wikipedia http://es.wikipedia.org/wiki/Ontolog%C3%ADa
26. Schizophrenia www.nimh.nih.gov/publicat/spSchizoph3517.cfm
27. Solar System Ontology A http://es.wikipedia.org/wiki/Sistema_Solar
28. Solar system Ontology B http://www.solarviews.com/span/solarsys.htm
29. 100 Years of loneliness Ontology A html.rincondelvago.com/cien-anos-de-soledad_gabriel-gar cia-marquez22.html
30. 100 Years of loneliness Ontology B http://monografías.com/trabajos10/ciso/ciso.shtml.