# CATO - A Lightweight Ontology Alignment Tool

Karin Koogan Breitman, Carolina Howard Felicíssimo, Marco Antonio Casanova

PUC-RIO – Pontifical Catholic University of Rio de Janeiro, Department of Informatics,
Rua Marquês de São Vicente 225, Rio de Janeiro, CEP 22453-900, RJ, Brasil
`{karin, cfelicissimo, casanova}@inf.puc-rio.br`

**Abstract.** Ontologies are becoming increasingly common in the World Wide Web as the building block for a future *Semantic Web*. In this Web, ontologies will be responsible for making the semantics of pages and applications explicit, thus allowing electronic agents to process and integrate resources automatically. The ability to integrate different ontologies meaningfully is thus critical to assure coordinated action in multi agent systems. In this paper, we propose a strategy and tool, CATO, that allow for totally automatic ontology alignment for the Semantic Web.

## 1   Introduction

Ontologies are rapidly becoming the *lingua franca* to express the semantics of information on the Web. As envisioned by Tim Berner's Lee [1], in the future, rather than sharing a few domain ontologies, crafted by knowledge engineers, e.g. WordNet [2] and CYC [3], every Web site and application in the Web will have its own ontology. There will be a "*great number of small ontological components consisting largely of pointers to each other*" [4].

His predictions seem to be true, as the number of tools for ontology edition, visualization and verification are drastically growing. The co-existence of a multitude of ontologies poses a further problem: semantic interoperability. In this paper, we focus on the ontology integration problem from a multi agent system perspective. The main contribution of the proposed strategy is to combine well known algorithmic solutions, such as natural language processing and tree comparison [16, 17], to the ontology integration problem.

Despite the existence of some strategies and supporting tools for ontology integration, most available techniques are either completely manual or semi-automatic, but all depend on user intervention to some degree. In the next section, we discuss some ontology integration techniques. In section 3, we introduce our alignment strategy. In section 4, we discuss the limitations of our strategy. Our conclusions are presented in section 5.

## 2   Related Work

Semantic interoperability among ontologies has been in the research agenda of knowledge engineers for a while now. A few approaches to help deal with the ontology integration problem have been proposed. The most prominent ones are: merging

[20], alignment [20, 21], mapping [21] and integration[1] [22]. The GLUE system [23] makes use of multiple learning strategies to help find mappings between two ontologies. IPROMPT provides guidance to the ontology merge process by describing the sequence of steps and helping identify possible inconsistencies and potential problems. AnchorPROMPT [21], an ontology alignment tool, automatically identifies semantically similar terms. It uses a set of anchors (pairs of terms) as input and treats the ontology as a directed graph. The Chimaera environment [36] provides a tool that merges ontologies based on their structural relationships. Instead of investigating terms that are directly related to one another, Chimaera uses the super and subclass relationships that hold in concept hierarchy to find possible matches. Their implementation is based in Ontolingua editor [24].

## 3   Ontology alignment with CATO

In this section, we outline the ontology alignment strategy that CATO implements. CATO takes as input any two ontologies written in W3C recommended standard OWL. An online version of CATO is publicly available at the following address: http://cato.les.inf.puc-rio.br/. It was fully implemented in JAVA and uses a specific API (Application Programming Interface) that deals with ontologies, JENA [25]. The listings in this paper were all generated by CATO.

### 3.1 Proposed strategy

The philosophy underlying our strategy is purely syntactical. We perform both lexical and structural comparisons in order to determine if concepts in different ontologies should be considered semantically compatible. We use a refinement approach, broken into three successive steps, illustrated in Figure 1.

Our assumption is that the use of lexically equivalent terms implies the same semantics, if the ontologies in question are in the same domain of discourse. For pairs of ontologies in different domains, lexical equivalence does not provide guarantee that the concepts share the same meaning.

To solve this problem, our strategy proposes to use structural comparison. Concepts that were once identified as lexically equivalent are now structurally investigated. Making use of the intrinsic structure of ontologies, a hierarchy of concepts connected by subsumption relationships [7], we now isolate and compare concept sub-trees. Investigation on the ancestors (super-concepts) and descendants (sub-concepts) will provide the necessary additional information needed to verify whether the pair of lexically equivalent concepts can actually be assumed to be semantically compatible.

---

[1] Please note that we use the term ontology integration as an abstraction that encapsulates all different treatments, including Pinto et all ontology integration approach.
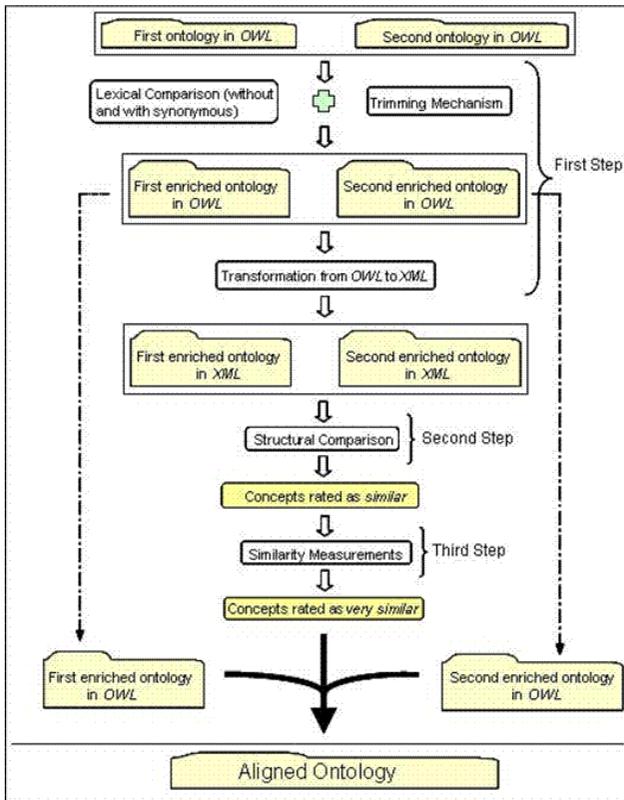
**Fig. 1.** CATO ontology alignment strategy

### 3.3.1 First Step: Lexical Comparison

The goal of this step is to identify lexically equivalent concepts. We assume the last are also semantically equivalent in the domain of discourse under consideration, an assumption which is not always warranted.

Each concept label in the first ontology is compared to every concept label present in the second one, using lexical similarity as the criteria. Besides using the label itself, synonyms are also used. The use of synonyms enriches the comparison process because it provides more refined information. As a result of the first stage of the proposed strategy, the original ontologies are enriched with links that relate concepts identified as lexically equivalent.

### 3.3.2 Second Step: Structural Comparison Using *TreeDiff*

Comparison at this stage is based on the subsumption relationship that holds among ontology concepts. Ontology properties and restrictions are not taken into consideration. Our approach is thus more restricted than the one proposed in [21], that analy-

ses the ontologies as graphs, taking into consideration both taxonomic and non taxo-
nomic relationships among concepts.

Because we only consider lexical and structural relationships in our analysis, we
are able to make use of well-known tree comparison algorithms. We are currently
using the *TreeDiff* [16] implementation available at [29]. Our choice was based on its
ability to identify structural similarities between trees in reasonable time.
The third and last step is based on similarity measurements. Concepts are rated as
very similar or little similar based on pre-defined similarity thresholds. We only align
concepts that were both classified as lexically equivalent in the second step, and thus
rated very similar. Thus the similarity measurement is the deciding factor responsible
for fine tuning our strategy. We adapted the similarity measurement strategies pro-
posed in [29, 30].
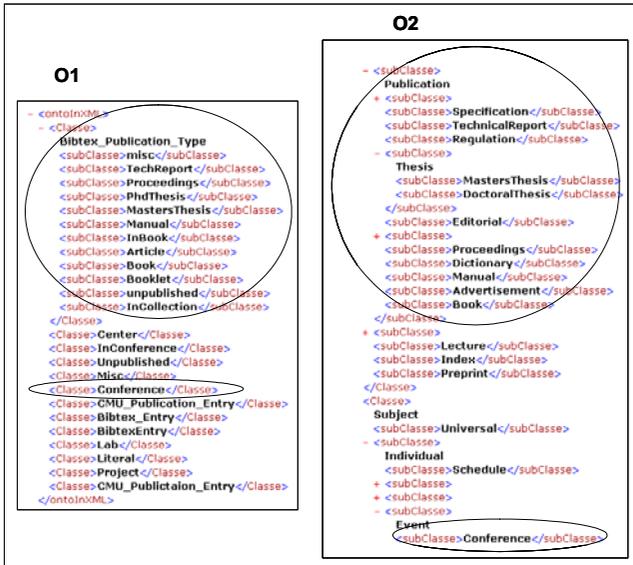


**Fig. 2.** Compared Ontologies

### 3.3.3. Third Step: Fine Adjustments based on Similarity Measurements
The third and last step is based on similarity measurements. Concepts are rated as
very similar or little similar, based on pre-defined similarity thresholds. We only align
concepts that were both classified as lexically equivalent in the second step, and thus
rated very similar. Thus the similarity measurement is the deciding factor responsible
for fine tuning our strategy. We adapted the similarity measurement strategies pro-
posed in [29, 30]. Table I illustrates the output of the similarity measurements for the
example illustrated in Figure 2. The output of this final step is a single ontology, that
provides a common understanding for the semantics represented by the two input
ontologies.

**Table 1.** Similarity Percentages for concepts in the equivalence group illustrated in Figure 2

```
Similarities Level:
Bibtex_Publication_Type -> Publication   *** Similarity Level: 23.076923%
Conference -> Conference   *** Similarity Level: 100.0%
Proceedings -> Proceedings   *** Similarity Level: 100.0%
Book -> Book   *** Similarity Level: 100.0%
Manual -> Manual   *** Similarity Level: 100.0%
```

## 4. Discussion

In order to guarantee the desired response time and discard user intervention, some commitments had to be made. To guarantee reasonable performance, we limited our approach to lexical and structural comparisons. Much richer analysis could be performed if additional information was used, e.g. restrictions (slots) as it is done in both the Chimaera and Prompt approaches [6, 21].

For the sake of efficiency, we are only taking into consideration syntactical information, i.e., lexical and structural equivalence, in the proposed alignment strategy. However, this limitation of the strategy can be overcome by the adaptation of the second step to take into consideration other ontology primitives, such as properties (the strategy could work with graphs instead of trees) and axioms.

## 5. Conclusions

In this paper, we discussed the implementation of a software component responsible for the automatic taxonomical alignment of ontologies. Our strategy is based on the application of well known software engineering strategies, such as lexical analysis, tree comparison and the use of similarity measurements, to the problem of ontology alignment. Motivated by the requirements of multi agent systems, we proposed an ontology alignment strategy and tool that produces an intermediate ontological representation that makes it possible for software agents searching for information to share common understanding over information available on the Web [31, 32 and 33].

## References

1.  Berners-Lee, T.; Lassila, O. Hendler, J.: The Semantic Web. Scientific American, May 2001. Available at: <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html/>. Accessed on November, 2004.
2.  Fellbaum, C.; ed: WordNet: An electronic Lexical Database. Cambridge, MA . MIT Press, 1998.
3.  Guha, R. V., D. B. Lenat, K. Pittman, D. Pratt, and M. Shepherd: Cyc: A Midterm Report. Communications of the ACM Vol.33 , No. 8 - August, 1990.
4.  Hendler, J.: Agents and the Semantic Web. IEEE Intelligent Systems. March/April, pp.30-37, 2001.
5.  Bechhofer, S., Ian Horrocks, Carole Goble, Robert Stevens: OilEd: a Reason-able Ontology Editor for the Semantic Web. Proceedings of KI2001, Joint German/Austrian conference on Artificial Intelligence, September 19-21, Vienna. Springer-Verlag LNAI Vol. 2174, pp. 396-408, 2001.
6.  McGuiness, D.; Fikes, R..; Rice, J.; Wilder, S.: An Environment for Merging and Testing Large Ontologies. Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR-2000), Brekenridge, Colorado, April 12-15, San Francisco: Morgan Kaufmann, pp. 483-493, 2002.
7.  Maedche, A.: Ontology Learning for the Sematic Web. Kluwer Academic Publishers, 2002.

8.   Fensel, D.; Wahlster, W.; Berners-Lee, T.; editors: Spinning the Semantic Web. MIT Press, Cambridge Massachusetts, 2003.
9.   Goméz-Peréz, A.; Fernandéz-Lopéz, Corcho, O.: Ontology Engeneering. Springer Verlag, 2004.
10.  Ushold, M; Gruninger, M.: Ontologies: Principles, Methods and Applications. Knowledge Engineering Review. Vol 11 No.2 - 1996.
11.  Guarino, N.: Formal Ontology and information systems. In Proceedings of the FOIS'98 – Formal Ontology in Information Systems, Trento – 1998.
12.  Noy, N.; McGuiness, D.: Ontology Development 101 – A guide to creating your first ontology. KSL Technical Report,  Standford University, 2001.
13.  Booch, G.; Rumbaugh, J.; Jacobson, I.: The Unified Modeling Language user guide. Addison Wesley - 1999.
14.  Yu, E.: Towards Modelling and Reasoning Support for Early-Phase Requirements Engineering. Proceedings of the Third International Symposium on Requirements Engineering - RE97. IEEE Computer Society Press, pp.226-235, 1997.
15.  Sowa, J. F.: Knowledge Representation: Logical, Philosophical and Computational Foundations. Brooks/Cole Books, Pacific Grove, CA, 2000.
16.  Wang, J.: An Algorithm for Finding the Largest Approximately Common Substructures of Two Trees. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 20, Number 8, pp. 889-895, 1998.
17.  TAI, K.,C..: The tree-to-tree correction problem. Journal of the ACM, 26(3), pp. 422-433, 1979.
18.  M. Wooldridge, N. R. Jennings, and D. Kinny: A methodology for agent-oriented analysis and design. In O. Etzioni, J. P. Muller, and J. Bradshaw, editors, Agents '99: Proceedings of the Third International Conference on Autonomous Agents, Seattle, WA, May 1999.
19.  Williams, A.B.: Learning to Share Meaning in a Multi-Agent System. Journal of Autonomous Agents and Multi-Agent Systems, Vol. 8, No. 2, 165-193, March 2004.
20.  Noy, N. F., Musen, M. A.: SMART: Automated Support for Ontology Merging and Alignment. Workshop on Knowledge Acquisition, Modeling, and Management, Banff, Alberta, Canada, 1999.
21.  Noy, N. F., Musen, M. A.: The PROMPT Suite: Interactive Tools For Ontology Merging And Mapping. International Journal of Human-Computer Studies, 2003.
22.  Pinto, S.H.; Goméz-Peréz, A.; Martins, J.P.: Some Issues on Ontology Integration. In: Workshop on Ontologies and Problems Solving Methods: Lessons Learned and Future Trends. Proceedings of the Workshop on Ontologies and Problem Solving Methods: Lessons Learned and Future Trends (IJCAI99), 1999.
23.  Doan, A., et. al.: Learning to match ontologies on the Semantic Web. In: The VLDB Journal — The International Journal on Very Large Data Bases, Volume 12, Issue 4, 2003. ISSN: 1066-8888. pp. 303-319, 2003.
24.  Farquhar, A. Fikes, R.; Rice, J.: The Ontolingua Server a Tool for Collaborative Ontology Construction. Proceedings of the Tenth Knowledge Acquisition for Knowledge Base Systems Workshop, Banff, Canada, 1996.
25.  Jena, the Semantic Web Framework, Available at: <http://jena.sourceforge.net/>. Accessed on November, 2004.
26.  CMU RI Publications. Available at: <http://www.daml.ri.cmu.edu/ont/homework/cmu-ri-publications-ont.daml/>. Accessed on November, 2004.
27.  Agent Transaction Language for Advertising Services. Available at: <http://www.daml.ri.cmu.edu/>. Accessed on November, 2004.
28.  Mondeca SA, A Semantic Knowledge Company. Available at: <http://www.mondeca.com/>. Accessed on November, 2004.
29.  Bergmann, U.: "Evolução de Cenários Através de um Mecanismo de Rastreamento Baseado em Transformações". PhD Thesis of the Department of Informatics of PUC-Rio, 2002.
30.  Alexander Maedche and Steffen Staab: Comparing Ontologies Similarity Measures and a Comparison Study. Institute AIFB, University of Karlsruhe, Internal Report, 2001.
31.  Williams, A.B., Padmanabhan, A., Blake, M.B.: Local Consensus Ontologies for B2B-Oriented Service Discovery. Second International Joint Conference on Autonomous Agents and Multi-Agent Systems, Melbourne, Australia, July 14-18, 2003.
32.  Haendchen, F., A.; Staa, A.v.; Lucena, C.J.P: A Component-Based Model for Building Reliable Multi-Agent Systems. In Proceedings of 28th SEW - NASA/IEEE Software Engineering Workshop, Greenbelt, MD, IEEE Computer Society Press, Los Alamitos, CA, 2003.
33.  Breitman, K.K., Haendchen, A.F., Staa, A., Haeusler, H.: Using Ontologies to Formalize Services Specifications in Multi-Agent Systems -  Third NASA - Goddard/ IEEE Workshop FAABS III - Formal Approaches to Agent-Based Systems - Greenbelt, MA - April, 2004.
34.  Nuseibeh, B.; Easterbrook, S.; Russo, A.: Leverage Inconsistency in Software Development Computer. - Vol 33 No. 4 - April 2000 - pp. 24-29, 2000.
35.   Easterbrook, S.; Chechik, M. - 2nd International Workshop on Living with Inconsistency – Summary, IEEE, 2001.
36.  D. McGuinness, R. Fikes, J. Rice, and S. Wilder: The Chimaera Ontology Environment. In Proceedings of the 17th National Conference on Artificial Intelligence (AAAI), 2000.