Automatic Generation of Glosses in the OntoLearn System

Alessandro Cucchiarelli*, Roberto Navigli†, Francesca Neri*, Paola Velardi‡

* DIIGA, Università Politecnica delle Marche, via Brecce Bianche 12, 160131 Ancona, Italy
{cucchiarelli,neri}@diiga.univpm.it

†Dipartimento di Informatica, Università di Roma “La Sapienza”, via Salaria 113, 00196 Roma, Italy
(navigli,velardi) @di.uniroma1.it

Abstract
OntoLearn is a system for automatic acquisition of specialized ontologies from domain corpora, based on a syntactic pattern matching technique for word sense disambiguation, called structural semantic interconnection (SSI). We use SSI to extract from corpora complex domain concepts and create a specialized version of WordNet. In order to facilitate the task of domain specialists who inspects and evaluate the newly acquired domain ontology, we defined a method to automatically generate glosses for the learned concepts. Glosses provide an informal description, in natural language, of the formal specifications of a concept, facilitating a per-concept evaluation of the ontology by domain specialists, who are usually unfamiliar with the formal language used to describe a computational ontology. The proposed evaluation framework has been tested in a financial domain.

1. Introduction
Automatic methods for ontology learning and population have been proposed in recent literature (ECAI, 2002; KCAP, 2003; Navigli & Velardi, 2004), but a correlated, and equally important, issue is the evaluation of such automatically generated ontologies, not only to the end of comparing the different approaches, but also to verify whether an automatic process may actually compete with the typically human process of converging on an agreed conceptualization of a given domain. Ontology construction, apart from the technical aspects of a knowledge representation task (i.e. choice of representation languages, consistency and correctness with respect to axioms, etc.), is a consensus building process, one that implies long and often harsh discussions among the specialists of a given domain. Can an automatic method simulate this process? Can we provide domain specialists with a mean to measure the adequacy of a specific set of concepts as a model of a given domain? Often, specialists are unable to evaluate the formal content (Hovy, 2001) of a computational ontology (e.g. the denotational theory, the formal notation, the knowledge representation system capabilities like property inheritance, consistency, etc.). Evaluation of the formal content is rather tackled by computational scientists, or by automatic verification systems. The role of the specialists is instead to compare their intuition of a domain with the description of this domain, as provided by the ontology concepts.

To facilitate per-concept evaluation, we have devised a method for automatic gloss generation, as an extension of the OntoLearn ontology population system (Navigli & Velardi, 2004; Navigli, Velardi & Gangemi, 2003). Glosses provide a description, in natural language, of the formal specification automatically assigned to the learned concepts. An expert can easily compare his intuition with this natural language description of system’s choices. In this paper, after a sketchy description of the Ontolearn system, we describe in detail the gloss generation algorithm and we provide an evaluation in an Economy domain.

2. A summary of OntoLearn
OntoLearn is an ontology population method based on text mining and machine learning techniques. OntoLearn starts with an existing general purpose ontology (we use WordNet, though other choices would be possible) and a set of documents in a given domain, and produces a domain extended and trimmed version of the initial ontology. Concept learning is achieved in the following phases:

1) Terminology Extraction: A list of domain terms is extracted from a set of documents which are judged representative of a given domain. Terms are extracted using natural language processing and statistical techniques. Contrastive corpora and glossaries in different domains are used to prune terminology that is not domain-specific. Domain terms are selected also on the basis of an entropy-based measure that simulates specialist consensus on concepts choice: the probability distribution of a “good” domain term must be uniform across the individual documents of the domain corpus.

2) Semantic interpretation of terms: Semantic interpretation is based on a principle, compositional interpretation, and on an novel algorithm, called structural semantic interconnections (SSI). Compositional interpretation signifies that the meaning of a complex term can be derived compositionally from its components1, e.g. the meaning of business plan is derived first, by associating the appropriate concept identifier, with reference to the initial top ontology, to the component terms (i.e. sense 2 of business and sense 1 of plan in WordNet), and then, by identifying the semantic relations holding among the involved concepts (e.g.

\[
\text{plan}#1 \rightarrow \text{topic} \rightarrow \text{business}#2.
\]

1 Compositional interpretation works well (see also the evaluation section) in domains which are not overly technical, like tourism, economy, sport, politics. In other domains like medicine, or computer science, other strategies must be adopted, like glossary parsing. This is an in-progress research.
3) Extension and trimming the initial ontology: Once the terms have been semantically interpreted, they are organized in sub-trees, and appended under the appropriate node of the initial ontology, e.g. business_plan#1-kind_of_plan#1. Furthermore, certain nodes of the initial ontology are pruned to create a domain-view of the ontology. The final ontology is output in OWL language.

SSI is a word sense disambiguation algorithm used to determine the correct concept for each term component. The algorithm is based on building a graph representation for alternative senses of each term, and then selecting the appropriate senses on the basis of the detected interconnections among graphs. Relevant interconnection types are described by a context free grammar G.

Details on the SSI algorithm are in (Navigli & Velardi, 2004), along with a performance evaluation performed on a Tourism domain. In this paper, we provide a more accurate description of the semantic relation annotation task, since this is an enhanced feature of OntoLearn.

2.1. Annotation with semantic relations.

In order to complete the interpretation process, OntoLearn attempts to determine the semantic relations that hold between the components of a complex concept. In order to do this, it was first necessary to select an inventory of semantic relations. We examined several proposals, like EuroWordnet (Vossen, 1999), DOLCE (Masolo et al., 2002), FrameNet (Ruppenhofer Fillmore & Baker, 2002) and others. As also remarked in (Hovy, 2001), no systematic methods are available in literature to compare the different sets of relations. Since our objective was to define an automatic method for semantic relation extraction, our final choice was to use a reduced set of FrameNet relations, which seemed general enough to cover our application domains (tourism, economy, computer networks). The choice of FrameNet is motivated by the availability of a sufficiently large set of annotated examples of conceptual relations, that we used to train an available machine learning algorithm, TiMBL (Daelemans et al., 2002). The relations used are: Material, Purpose, Use, Topic, Product, Constituent Parts, Attribute. Examples for each relation are the following:

merge#1-purpose-merger#1
meeting#1-use-meeting#1
bond#1-const-part-bond#1
com-payer#1-product-company#1
net#1-attribute-net#1
 takeover#1-topic-takeover#1
sand#1-material-sand#1

We represented training instances as pairs of concepts annotated with the appropriate conceptual relation, e.g.:

[(computer#1,maker#3),Product]

Each concept is in turn represented by a feature-vector where attributes are the concept’s hyperonyms in WordNet, e.g.:

(computer#1,maker#3)→{(computer#1,machine#1,device#1,instrumentality#3),(maker#3,business#1,enterprise#2,organization#1)}

2.2. Evaluation of the Semantic Annotation Task

To test the semantic relation annotation task, we used a learning set (including selected annotated examples from FrameNet (FN), Tourism (Tour), and Economy (Econ)), and a test set with a distribution of examples shown in Table 1.

Table 1: Distribution of examples in the learning and test set

| Sem_Rel | Learning Set | Test Set |
|---------|--------------|----------|
| Tour    | Econ         |          |
| FN      | Tour         | Econ     |
| Tot     |              |          |

We then tested the learner on test sets for individual domains and to the entire test set, leading to the results shown in Table 2a,b and c.

Table 2: (a)Performance on Tourism, (b)Performance on Economy (c)Performance on the complete test set

| Sem_Rel | Learning Set | Test Set |
|---------|--------------|----------|
| Tour    | Econ         |
| FN      | Tour         |
| Tot     |              |
| Precision MACRO | 0.958 | 0.875 |
| Recall MACRO    | 0.283 | 0.636 |
| F1 MACRO      | 0.437 | 0.737 |
| Precision micro | 0.900 | 0.857 |
| Recall micro   | 0.087 | 0.635 |
| F1 micro       | 0.158 | 0.721 |

Notice that the relation Attribute is generated whenever the term associated to one of the concepts is an adjective. Therefore, this semantic relation is not included in the evaluation experiment, since it would artificially increase performances. We then tested the learner on test sets for individual domains and to the entire test set, leading to the results shown in Table 2a,b and c.

Table 2: (a)Performance on Tourism, (b)Performance on Economy (c)Performance on the complete test set

Notice that the global performance (Table 2c) is improved by the fact that examples of annotated relations in FrameNet are very repetitive (the same patterns – often the same words pairs - tend to appear for the same
relations. This perhaps justifies the higher performance obtained on a similar task in (Gildea & Jurafsky, 2002).

The performance measures shown in the above Tables are those adopted in TREC competitions. The parameter $d$ is a confidence factor defined in the TiMBL algorithm. This parameter can be used to increase system’s robustness in the following way: whenever the confidence associated by TiMBL to the classification of a new instance is lower than a given threshold, we output a “generic” conceptual relation, named Relatedness. We experimentally fixed the threshold for $d$ around 30% (central columns of Tables 2a, b and c).

3. The gloss generation algorithm

The Ontolearn system has been tested and evaluated in the context of the European project Harmonise IST-2000-29329, on interoperability among Tourism enterprises.

The lesson that we learned during the Harmonise EC project was that the domain specialists, tourism operators in our case, can hardly evaluate the formal aspects of a computational ontology. When presented with the domain extended and trimmed version of WordNet (OntoLearn’s phase 3 in Section 2), they were only able to express a generic judgment on each node of the hierarchy, based on the concept label. These judgments were used to evaluate the terminology extraction task.

To help human evaluation on a per-concept basis, we decided to enhance Ontolearn with a gloss generation algorithm. The idea is to generate glosses in a way which closely reflects the key aspects of the concept learning process, i.e. semantic disambiguation and annotation with a conceptual relation.

The gloss generation algorithm is based on the definition of a grammar with distinct generation rules for each type of semantic relation.

Let $S_{ij} \rightarrow S_{i} \cdot S_{j}$ be the complex concept associated to a complex term $w_{i}w_{k}$ (e.g. jazz festival, or long-term debt), and let:

- $<H>$ = the syntactic head of $w_{i}w_{k}$ (e.g. festival, debt)
- $<M>$ = the syntactic modifier of $w_{i}w_{k}$ (e.g. jazz, long-term)
- $<GNC>$ = the gloss of the new complex concept $S_{ik}$
- $<HYP>$ = the selected sense of $<H>$ (e.g. respectively, festival#1 and debt#1)
- $<MSGHYP>$ = the main sentence of the WordNet gloss of $<HYP>$
- $<MSGM>$ = the main sentence of the WordNet gloss of the selected sense for $<M>$.

Here we provide two examples of rules for generating GCNs:

- If $\text{sem rel}=\text{Topic}$, $<\text{GNC}> = \text{a kind of } <\text{HYP}>$, $<\text{MSGHYP}>$, $\text{relating to the } <\text{M}>$, $<\text{MSGM}>$.
  e.g.: GNC(festival): a kind of festival, a day or period of time set aside for feasting and celebration, relating to the jazz, a style of dance music popular in the 1920.

- If $\text{sem rel}=\text{Attribute}$, $<\text{GNC}> = \text{a kind of } <\text{HYP}>$, $<\text{MSGHYP}>$, $<\text{MSGM}>$.
  e.g.: GNC(long term debt) = a kind of debt, the state of owing something (especially money), relating to or extending over a relatively long time.

At the time the gloss generation feature was added to OntoLearn, the Harmonise tourism project was over, therefore we switched to another application domain, Economy, for which we could rely on the judgment of a specialist. The specialist was an economist of our University and was unaware of the method used to generate glosses; he was presented with a list of 48 concept-gloss pairs and asked to fill in an evaluation form (see Appendix) as follows: vote 1 means “unsatisfactory definition”, vote 2 means “the definition is helpful”, vote 3 means “the definition is fully acceptable”. Whenever he was not fully happy with a definition, the specialist was asked to provide a brief explanation. 23 concepts received a “3”, 15 a “2” and 10 a “1”, leading to an average of 2.27. We subsequently added a comment to the form (the field Diagnose in Appendix), in which we explain the cause of the errors (for entries marked with 1 or 2).

The following conclusions can be drawn from this experiment:

- Overall, the domain specialist fully accepted the system’s choices in 48% of the cases, and was reasonably satisfied in 79% of the cases.
- Only in one case the compositional criterion was found not acceptable (concept #44 in the Appendix).
- In most cases, the glosses associated to the generic Relatedness relation were found acceptable. Over 9 occurrences, the correspondent glosses were marked three times with 3 (e.g. concept #30), three times with two, three times with 1.

In some case, OntoLearn produces disambiguation errors, but since sense distinctions are very fine-grained in WordNet, the generated definition may still be marked as acceptable (e.g. #1) or fully acceptable. In 8 cases (in 4 of which the wrong concept was the same), OntoLearn’s disambiguation errors produce perceivable (by the specialist) errors in concept interpretation.

There are cases in which WordNet simply does not include the right senses (e.g. #4) for the component concepts. In this case, the source of error is not OntoLearn, but the inadequacy of the initial ontology.

Overall, the result of the experiment was very encouraging. The Ontolearn compositional approach is general enough to produce acceptable conceptualizations at least for mid-technical domains. Obviously, the proposed evaluation technique only in part attacks the complex issue of evaluating automatically generated ontologies, however it provides a reasonable basis to support a per-concept analysis. The implementation of this evaluation strategy is tuned for the Ontolearn algorithm, but indeed generating natural language sentences from formal language statements is a common practice.

Acknowledgements

We thank Dr. Iacobucci who gave up his precious time to evaluate our glosses.

4. References

Daelemans,W., Zavrel,J., Van den Sloot,K. & Van den Bosch,A (2002). TiMBL: Tilburg Memory Based Learner. Version 4.3 Reference Guide. Tilburg University.
**APPENDIX: Excerpt of the per-concept evaluation form**

| Concept # | Term:                  | Synt:   | Rel<w1,w2>:          | Gloss:                                                                                     | Specialist vote: | Comment by Specialist: | Diagnosis:                                      |
|-----------|------------------------|---------|----------------------|-------------------------------------------------------------------------------------------|------------------|------------------------|------------------------------------------------|
| 1         | Stock price            | N-N     |                      | a kind of price, the property of having material worth, relating to the stock, the capital raised by a corporation through the issue of shares entitling holders to partial ownership. | 3                | none                   | OntoLearn disambiguation error (a better definition is available in WordNet) |
| 2         | Pension fund           | N-N     |                      | a kind of fund, a reserve of money set aside for some purpose, for pension, a regular payment to a person that is intended to allow them to subsist without working. | 3                | none                   | none                                           |
| 3         | Capital-gain tax       | N-N     |                      | a kind of tax, charge against a citizen’s person or property or activity for the support of government, relating to the capital-gain, the amount by which the selling price of an asset exceeds the purchase price. | 3                | none                   | none                                           |
| 4         | Futures price          | N-N     |                      | a kind of price, the amount of money needed to purchase something, relating to the futures, bulk commodities bought or sold at an agreed price for delivery at a specified future date. | 2                | Comment by Specialist: | Definition of ‘future’ overly specific. In the economic contexts, not referred only to ‘commodities’ |
| 23        | Computer maker         | N-N     |                      | a kind of maker, a business engaged in manufacturing some product, which produces computer, a machine for performing calculations automatically. | 3                | none                   | none                                           |
| 26        | Bond market            | N-N     |                      | a kind of market, the world of commercial activity where goods and services are bought and sold, consisting of bond, a certificate of debt that is issued by a government or corporation in order to raise money. | 3                | none                   | none                                           |
| 30        | Takeover attempt       | N-N     |                      | a kind of attempt, earnest and conscientious activity intended to do or accomplish something, concerning the takeover, a change by sale or merger in the controlling interest of a corporation. | 3                | none                   | none                                           |
| 44        | Initial offering       | Agg-N   |                      | a kind of offering, something offered (as a proposal or bid), occurring at the beginning. | 1                | Comment by Specialist: | Fully wrong; meaning cannot be derived from individual words |

**References:**

- ECAL(2002). *Ontology Learning Tools Workshop* http://www-sop.inria.fr/acacia/WORKSHOPS/ECAL 2002-OLT/accepted-papers.html
- Gildea,D & Jurafsky,D.(2002). Automatic labeling of semantic roles. *Computational Linguistics* (28)3, 225–288.
- Hovy,E (2001). Comparing Sets of Semantic relations in Ontologies. In R. Geen, C.A. Bean and S. Myaeng Semantic relations. Kluwer.
- KCAP.(2003). *Knowledge mark-up and Semantic Annotation* workshop, http://km.aifb.uni-karlsruhe.de/ws/semanannot2003/papers.html
- Masolo,C. Borgo,S. Gangemi,A. Guarino,N. Oltramari,A. & Schneider,L.(2002) Sweetening Ontologies with DOLCE. Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web.
- Navigli,R & Velardi,P.(2004) Learning Domain Ontologies from Document Warehouses and Dedicated Web Sites. *Computational Linguistics* (50)2.
- Navigli,R. Velardi,P. & Gangemi,A.(2003). Corpus Driven Ontology Learning: a Method and its Application to Automated Terminology Translation. *IEEE Intelligent Systems* (18)1: 22–31.
- Ruppenhofer,J. Fillmore,C.J. & Baker,C.F.(2002) Collocaational Information in the FrameNet Database. In Brausch,A. and Povlsen,C (eds.), Proceedings of the Tenth Euralex International Congress. Copenhagen, Denmark. Vol. I: 359–369, 2002
- Vossen,P.(1999). *EuroWordNet: General Document. Version 3 Final.* http://www.hum.uva.nl/~ewn