HAL
open science

AROMA results for OAEI 2008

Jérôme David

To cite this version:

Jérôme David. AROMA results for OAEI 2008. Proc. 3rd ISWC workshop on ontology matching (OM), Oct 2008, Karlsruhe, Germany. pp.128-131. hal-00793537

HAL Id: hal-00793537

https://inria.hal.science/hal-00793537v1

Submitted on 22 Feb 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AROMA results for OAEI 2008

Jérôme David[1]

INRIA Rhône-Alpes, Montbonnot Saint-Martin, France
Jerome.David-at-inrialpes.fr

Abstract. This paper presents the results obtained by AROMA for its first participation to OAEI. AROMA is an hybrid, extensional and asymmetric ontology alignment method which makes use of the association paradigm and a statistical interstingness measure, the implication intensity.

1 Presentation of AROMA

1.1 State, purpose, general statement

AROMA is an hybrid, extensional and asymmetric matching approach designed to find out relations (equivalence and subsumption) between entities issued from two textual taxonomies (web directories or OWL ontologies). Our approach makes use of the association rule paradigm [Agrawal *et al.*, 1993], and a statistical interestingness measure used in this context. AROMA relies on the following assumption: *An entity A will be more specific than or equivalent to an entity B if the vocabulary (i.e. terms and also data) used to describe A, its descendants, and its instances tends to be included in that of B.*

1.2 Specific techniques used

AROMA is divided into three successive main stages: (1) The pre processing stage allows to represent each entity (classes and properties) by a set of terms, (2) the second stage consists of the discovery of association rules between entities, and finally (3) the post processing stage aims to clean and enhance the alignment.

The first stage constructs a set of relevant terms and/or datavalues for each entity. To do this, we extract the vocabulary of entities from their annotations and individual values with the help of single and binary term extractor applied to stemmed text.

The second stage of AROMA discovers the subsumption relations by using the association rule model and the implication intensity measure [Gras *et al.*, 2008]. In the context of AROMA, an association rule $a \rightarrow b$ represents a quasi-implication (i.e. an implication allowing some counter-examples) from the vocabulary of entity a into the vocabulary of the entity b. Such a rule could be interpreted as a subsumption relation from the antecedent entity toward the consequent one. For example, the binary rule $car \rightarrow vehicle$ could be interpret: "The concept car is more specific than the concept $vehicle$". The rule extraction algorithm takes advantage of the partial order structure provided by the subsumption relation, and a property of the implication intensity for pruning the search space.

The last stage concerns the post processing of the association rule set. It performs the following tasks:

- the deduction of equivalence relations,
- the suppression of cycles in the alignment graph,
- the suppression of the redundant correspondences,
- the selection of the best correspondence for each entity (the alignment is an injective function),
- the enhancement of the alignment by using a string similarity -based matcher (Jaro-Winkler similarity) and previously discovered correspondences.

For more details, the reader should refer to [David *et al.*, 2007; David, 2007].

1.3 Link to the system and parameters file

The version of AROMA used for OAEI 2008 is available at:
`http://www.inrialpes.fr/exmo/people/jdavid/oaei2008/AROMA_oaei2008.jar`.

The source code is available at :
`http://www.inrialpes.fr/exmo/people/jdavid/oaei2008/AROMAsrc_oaei2008.jar`

For align two ontologies use the following command line:

```
java -jar AROMA_oaei2008.jar onto1.owl onto2.owl
```

The resulting alignment is provided on the standard output in the alignment format.

1.4 Link to the set of provided alignments (in align format)

`http://www.inrialpes.fr/exmo/people/jdavid/oaei2008/results_AROMA_oaei2008.zip`

2 Results

We participated to benchmark, anatomy and fao tests. We used the same configuration of AROMA for all tests. We did not have scaling problems. We only comment benchmark results because we do not have the results on the other tests. We also discuss why we did not participate to directory and mldirectory tests.

2.1 benchmark

Since AROMA only relies on textual information, it obtains bad recall values when the alterations affect all text annotations both in the class/property descriptions and in their individual/property values. AROMA seems do be not influenced by structural alterations (222-247). On these tests, AROMA favors high precision values in comparison to recall values.

2.2 anatomy

On anatomy test, we do not use any particular knowledge about biomedical domain. Anatomy ontologies use their own annotation properties. We have made some adaptations in order to deal with these annotations.

In terms of precision and recall, AROMA performs worse than the label equality matcher. In particular, it obtains quite low recall value. Nevertheless, it discovers 30% of non-trivial correspondences that are not found by the term equality matcher.

Since AROMA takes benefits of the subsumption relation for pruning the search space, it runs quite fast. This pruning feature partially explains the low recall value obtained by AROMA on the anatomy test.

2.3 fao

We do not make any adaptation for this test. Fao ontologies (as anatomy) use their own, as a consequence, some textual data were not taken into account by AROMA.

On this test, AROMA also obtains low recall value and some results have not been evaluated due to the lack of returned correspondences.

2.4 directory

The two large directories (given previous years but not this year) are divided into very small sub directories. AROMA cannot align such very small directories because our method is based on a statistical measure and then it needs some large amount of textual data. However, AROMA discovers correspondences when it is applied to the complete directories. It would be interesting to reintroduce these large taxonomies for the next editions.

2.5 mldirectory

AROMA only relies on common textual data shared by ontologies to be align and it does not use multi-lingual resources. As a consequence, it does not work with this kind of tests.

3 General comments on AROMA

Even if results are quite good on benchmark track, alignments provided by AROMA have quite low recall values. This is partially due to the pruning strategy used. To overcome this drawback, we used a syntactical matcher in order to augment alignments. Even if it performs well on benchmarks ontologies, this matcher seems not to be very efficient on real cases (fao and anatomy).

4 Conclusion

For its first participation to OAEI, AROMA passed the benchmark, FAO and Anatomy tests. We do not have any scaling problem with these tests.

The results on benchmarks shows that AROMA is dependent on the amount of textual information available and it has bad results when both labels and comments are suppressed. However, AROMA is not very influenced by structural alterations. On anatomy track, AROMA has good runtimes but lacks in terms of recall. FAO track corroborates this drawback.

References

[Agrawal *et al.*, 1993] Rakesh Agrawal, Tomasz Imielinski, and Arun Swami. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pages 207–216. ACM Press, 1993.

[David *et al.*, 2007] Jérôme David, Fabrice Guillet, and Henri Briand. Association rule ontology matching approach. *International Journal on Semantic Web and Information Systems*, 3(2):27–49, 2007.

[David, 2007] Jérôme David. *AROMA : une méthode pour la découverte d'alignements orientés entre ontologies à partir de règles d'association*. PhD thesis, Université de Nantes, 2007.

[Gras *et al.*, 2008] Régis Gras, Einoshin Suzuki, Fabrice Guillet, and Filippo Spagnolo, editors. *Statistical Implicative Analysis, Theory and Applications*, volume 127 of *Studies in Computational Intelligence*. Springer, 2008.