



NeOn-project.org

NeOn: Lifecycle Support for Networked Ontologies

Integrated Project (IST-2005-027595)

Priority: IST-2004-2.4.7 – “Semantic-based knowledge and content systems”

D3.4.1 Evaluation of Mapping

Deliverable Co-ordinator: Sandra Kohler

Deliverable Co-ordinating Institution: Intelligent Software Components (iSOCO)

Other Authors: Jérôme Euzenat (INRIA); Germán Herrero (ATOS); Caterina Caracciolo (FAO)

This deliverable evaluates existing ontology mapping algorithms. It describes the experiments and their analysis realized with the ontologies of the case studies: Ontologies of the fisheries domain, the invoice domain and the semantic nomenclature domain.

Document Identifier:	NEON//D3.4.1/v1.0	Date due:	February 28 th , 2009
Class Deliverable:	NEON EU-IST-2005-027595	Submission date:	February 28 th , 2009
Project start date:	March 1, 2006	Version:	v1.0
Project duration:	4 years	State:	Final
		Distribution:	Public

NeOn Consortium

This document is a part of the NeOn research project funded by the IST Programme of the Commission of the European Communities by the grant number IST-2005-027595. The following partners are involved in the project:

<p>Open University (OU) – Coordinator Knowledge Media Institute – KMi Berrill Building, Walton Hall Milton Keynes, MK7 6AA United Kingdom Contact person: Martin Dzbor, Enrico Motta E-mail address: {m.dzbor, e.motta} @open.ac.uk</p>	<p>Universität Karlsruhe – TH (UKARL) Institut für Angewandte Informatik und Formale Beschreibungsverfahren – AIFB Englerstrasse 11 D-76128 Karlsruhe, Germany Contact person: Peter Haase E-mail address: pha@aifb.uni-karlsruhe.de</p>
<p>Universidad Politécnica de Madrid (UPM) Campus de Montegancedo 28660 Boadilla del Monte Spain Contact person: Asunción Gómez Pérez E-mail address: asun@fi.upm.es</p>	<p>Software AG (SAG) Uhlandstrasse 12 64297 Darmstadt Germany Contact person: Walter Waterfeld E-mail address: walter.waterfeld@softwareag.com</p>
<p>Intelligent Software Components S.A. (ISOCO) Calle de Pedro de Valdivia 10 28006 Madrid Spain Contact person: Jesús Contreras E-mail address: jcontreras@isoco.com</p>	<p>Institut ‘Jožef Stefan’ (JSI) Jamova 39 SI-1000 Ljubljana Slovenia Contact person: Marko Grobelnik E-mail address: marko.grobelnik@ijs.si</p>
<p>Institut National de Recherche en Informatique et en Automatique (INRIA) ZIRST – 655 avenue de l’Europe Montbonnot Saint Martin 38334 Saint-Ismier France Contact person: Jérôme Euzenat E-mail address: jerome.euzenat@inrialpes.fr</p>	<p>University of Sheffield (USFD) Dept. of Computer Science Regent Court 211 Portobello street S14DP Sheffield United Kingdom Contact person: Hamish Cunningham E-mail address: hamish@dcs.shef.ac.uk</p>
<p>Universität Koblenz-Landau (UKO-LD) Universitätsstrasse 1 56070 Koblenz Germany Contact person: Steffen Staab E-mail address: staab@uni-koblenz.de</p>	<p>Consiglio Nazionale delle Ricerche (CNR) Institute of cognitive sciences and technologies Via S. Martino della Battaglia, 44 - 00185 Roma-Lazio, Italy Contact person: Aldo Gangemi E-mail address: aldo.gangemi@istc.cnr.it</p>
<p>Ontoprise GmbH. (ONTO) Amalienbadstr. 36 (Raumfabrik 29) 76227 Karlsruhe Germany Contact person: Jürgen Angele E-mail address: angele@ontoprise.de</p>	<p>Food and Agriculture Organization of the United Nations (FAO) Viale delle Terme di Caracalla 1 00100 Rome Italy Contact person: Marta Iglesias E-mail address: marta.iglesias@fao.org</p>
<p>Atos Origin S.A. (ATOS) Calle de Albarracín, 25 28037 Madrid Spain Contact person: Tomás Pariente Lobo E-mail address: tomas.pariantelobo@atosorigin.com</p>	<p>Laboratorios KIN, S.A. (KIN) C/Ciudad de Granada, 123 08018 Barcelona Spain Contact person: Antonio López E-mail address: alopez@kin.es</p>

Work package participants

The following partners have taken an active part in the work leading to the elaboration of this document, even if they might not have directly contributed to the writing of this document or its parts:

iSOCO

ATOS

INRIA

FAO

Change Log

Version	Date	Amended by	Changes
0.1	25-11-2008	Sandra Kohler	Table of Contents, Evaluation Setting
0.2	13-12-2008	Jérôme Euzenat	Including the part of FAO and INRIA (chapter 3 and 4)
0.3	30-01-2009	Sandra Kohler	Chapter 5.1, (Including AROMA)
0.4	12-02-2009	Germán Herrero	Including the experiments realized by ATOS (chapter 5.2)
0.5	13-02-2009	Sandra Kohler	Introduction, Conclusion
1.0	13-02-2009	Sandra Kohler	Final version send to reviewer
1.1	20-03-2009	Jérôme Euzenat	Changes proposed by reviewer (chapter 2.1, 4.1 and 4.3.)
1.2	27-03-2009	Sandra Kohler	Changes proposed by reviewer

Executive Summary

Ontology mapping - refers to the activity of finding the correspondences between two or more ontologies and storing/exploiting them (Ontology Aligning is a synonym for this activity, while Ontology Matching can be seen as the first step of this process) see [1] - can help to add additional context information to the ontologies, but doing it manually is a rather time-consuming process. In this deliverable we evaluate different ontology mapping algorithms by comparing them to reference alignments defined by domain experts. To classify the alignments we use the three different semantic relation types defined in the Neon Mapping Metamodel [2]: Equivalence, Containment and Overlap. For algorithms which calculate only equivalence relations, the alignments of the non-equivalence type are ignored.

The experiments are realized with different sets of ontologies developed or reused in the case studies of the NeOn project: the fisheries ontologies like Agrobio [3] (WP7), the semantic nomenclature ontologies like Galen [4] (WP8) and the invoice ontologies like the InvoicingBackboneOntology [5] (WP8). The experiments of the fisheries use case were realized during the 2008 OAEI (Ontology Alignment Evaluation Initiative) workshop.

Table of Contents

1	Introduction	6
2	Evaluation Setting	7
2.1	Scope of evaluation	7
2.1.1	<i>Ontology Mapping</i>	<i>7</i>
2.2	Used Technologies.....	8
2.2.1	<i>Alignment Plugin</i>	<i>8</i>
2.2.2	<i>OntoConto</i>	<i>8</i>
2.3	Expected Conclusions	8
3	Guidelines of the Ontology Alignment Evaluation Initiative	9
3.1	General Guidelines.....	9
3.1.1	<i>Typical exceptions.....</i>	<i>10</i>
3.2	Application to the realized experiment	11
4	Experiment of the fisheries case study (within OAEI 2008)	12
4.1	Used Ontologies	12
4.2	Used Algorithms	13
4.3	Description of realized experiments	14
4.4	Results	15
4.4.1	<i>Quantitative Results</i>	<i>15</i>
4.4.2	<i>Qualitative Results</i>	<i>16</i>
5	Experiment of the pharmaceutical case studies	17
5.1	Experiment with the Invoice Ontologies	17
5.1.1	<i>Used Ontologies.....</i>	<i>17</i>
5.1.2	<i>Used algorithms</i>	<i>19</i>
5.1.3	<i>Realized experiments.....</i>	<i>19</i>
5.1.4	<i>Results.....</i>	<i>21</i>
5.1.4.1	<i>Quantitative Results</i>	<i>21</i>
5.1.4.2	<i>Qualitative Results</i>	<i>24</i>
5.2	Experiments with the Semantic Nomenclature Ontologies	26
5.2.1	<i>Used ontologies.....</i>	<i>26</i>
5.2.2	<i>Used algorithms</i>	<i>28</i>
5.2.3	<i>Realized Experiments</i>	<i>29</i>
5.2.4	<i>Results.....</i>	<i>29</i>
5.2.4.1	<i>Quantitative Results</i>	<i>29</i>
5.2.4.2	<i>Qualitative Results</i>	<i>30</i>
6	Conclusions and ongoing work.....	32
7	References	33

List of tables

Table 1 Size of returned results and samples.....	14
Table 2 Participant results per datasets.....	15
Table 3 InvoicingBackboneOntology aligned with EdifactInvoiceMessageOntology	20
Table 4 InvoicingBackboneOntology aligned with PharmaInnovaInvoiceOntology	20
Table 5 InvoicingBackboneOntology aligned with UBLInvoicingOntology	20
Table 6 Results TaxoMap	21
Table 7 Results Semanticmapper	22
Table 8 Results OLA algorithm	23
Table 9 Results of AROMA algorithm	23
Table 10 Results TaxoMap	29
Table 11 Results OLA algorithm	30
Table 12 Results of AROMA algorithm	30

List of figures

Figure 1: Visualization of the TaxoMap alignment between IBO and PIIO.	22
Figure 2: Visualization of the mapping between the IBO and EIMO	24

1 Introduction

Ontology mapping refers to the activity of finding the correspondences between two or more ontologies and storing/exploiting them (see NeOn Glossary, [1]). The expression Ontology Aligning is used as a synonym. Ontology Mapping can help to add additional context information to the ontologies (see chapter 2.1.1), but doing it manually is a rather time-consuming process. In this deliverable we therefore evaluate existing ontology mapping algorithms which execute the mapping automatically. We aim to find out how good different mapping algorithms detect alignments between ontologies describing the same domain or different aspects of the domain. We also like to detect how mature the current available algorithms are, and which aspects might be improved. To evaluate the matchers¹ we test them by comparing their results with reference alignments defined by domain experts. Most of the matchers we are using are available within the NeOn Alignment Plugin [6]. It is important to mention that we are not going to evaluate the plugin itself. Nevertheless it provides great help in respect to the easy access to the algorithms and works very reliable.

Ontology mapping is an important task in the work with networked ontologies. Alignments permit the coupling of different ontologies to obtain a more complete description of a domain. Using alignments we can create a net of content without the need of creating bigger and bigger ontologies. Mapping the ontologies we obtain the concepts or attributes where to plug them with each other. Importing different ontologies created for different aspects we can obtain a big coverage of a concrete domain or the intersection between different domains.

Another aspect we want to clarify in this deliverable is how autarkic ontology mapping algorithms can obtain reliable alignments and how much post-work for the user is involved in the process. Even if the user has to rework the results, reducing the set of possible alignments will still be a helpful and time saving task.

The deliverable is structured as follows: In the second chapter we describe the evaluation setting, defining the scope and the purpose of the evaluation. In the third chapter we explain the guidelines of the Ontology Alignment Evaluation Initiative (OAEI) which are applied in the experiments realized with the ontologies of the fisheries case study described in chapter 4 (these experiments were realized during the 2008 OAEI workshop in Karlsruhe²). Chapter 5 demonstrates the experiments realized in the pharmaceutical case study both with the ontologies of the semantic nomenclature and the invoice domain. In chapter 6 we finish with the conclusions and the ongoing work.

¹ We use “matcher” as synonym for “ontology mapping algorithm”.

² <http://om2008.ontologymatching.org>

2 Evaluation Setting

In this chapter we outline the scope of the evaluation, see section 2.1 and the tools and algorithms we are going to use for the experiments, 2.2. Moreover we define the conclusions we expect to be able to draw, 2.3.

2.1 Scope of evaluation

We are evaluating existing ontology mapping algorithms. The experiments of the fisheries case study were realized during the 2008 OAEI tracks. The algorithms used for these experiments are mainly not available in the Alignment Plugin (see 2.2.1); they are described in chapter 4.1. For the experiments of the pharmaceutical case studies we used algorithms available in the online version of the Alignment Plugin; the algorithms used for both the experiments of the semantic nomenclature case study and the invoice case study can be found in chapter 5.2.2 respectively 5.1.2.

It is important to clarify, that we use the Alignment Plugin as a helpful tool, but we do not evaluate it in this deliverable. The plugin permits the use of different algorithms in an easy manner, but it is not responsible for the results of the alignments. In this deliverable we evaluate the mapping algorithms by comparing their results with reference alignments defined by experts, calculating precision and recall.

2.1.1 Ontology Mapping

As described in D3.3.2 [6] the alignment of different ontologies can help to add context, due to the fact that knowledge can vary in nature and expression. Ontologies are developed in a special environment, having a concrete purpose. This means that ontologies will be re-used in settings that are not those that have led to their design. To clear their specific context, matching to other ontologies of the same domain helps.

Another scope is the set of ontologies describing different aspects of the same domain. The experiments of the invoice ontologies describe this scenario. Matching ontologies of this kind can help to find the links where to connect them with each other to describe the whole domain more completely.

Ontology mapping helps to build up a net of ontologies. The mapping shows the points where the different ontologies can be coupled with each other. Like this the re-use is facilitated and domains can be described more detailed, covering all aspects without the need to create bigger and bigger ontologies.

2.2 Used Technologies

The experiments of the pharmaceutical case studies³ were realized with the help of two NeOn plugins the Alignment Plugin [6] and OntoConto [7]. We used the Alignment Plugin to create the alignments, to save them and to post-process them. OntoConto on the other hand we used for the visualization of the alignments.

2.2.1 Alignment Plugin

The Alignment Plugin is a tool providing an unique interface to allow the user to work with different kind of matchers without having to know the details and without having to use different programs and interfaces for each of them. The Alignment Plugin has both an offline and an online mode [6]:

- In the offline mode, the Alignment API is integrated within the plugin, and thus, allows the plugin to perform all functions implemented in this API.
- In the online mode, the NeOn Alignment Plugin provides functions related to managing and accessing alignments and algorithms on a server, in addition to the functions offered by the offline mode.

The experiments in this work were realized with algorithms available in the online mode, because they take more parameters into account and therefore return more sophisticated alignments.

2.2.2 OntoConto

OntoConto [7] is a visualization and editing tool for ontology mapping. It visualizes the results returned by the algorithms offered by the Alignment Plugin. Visualization of alignments opens an easier and more intuitive access to the mapping, helping to get a first overview of the returned alignments.

2.3 Expected Conclusions

In this deliverable we will evaluate the existing ontology mapping algorithms. In this context we want to find out which algorithms return the best results and for which domains. In case of the invoice ontologies for example we have a set of each other complementing ontologies which describe different aspects of the same domain. We want to find out if automatic ontology mapping works for specific situations like this as well. In case of bad results, we try to find out why the algorithms do not return reasonable alignments. Moreover we want to detect how big the post-processing work for the user is and finally we will recommend concrete matchers.

³ The experiments with the fisheries ontologies were realized during the 2008 OAEI track, they did not necessarily made use of these plugins.

3 Guidelines of the Ontology Alignment Evaluation Initiative

The Ontology Alignment Evaluation Initiative⁴ (OAEI) is a coordinated international initiative that organizes the evaluation of the increasing number of ontology matching systems. The main goal of the Ontology Alignment Evaluation Initiative is to compare systems and algorithms on the same basis and to allow anyone for drawing conclusions about the best matching strategies. The ambition of the OAEI is that from such evaluations, tool developers can learn and improve their systems. The OAEI campaign provides the evaluation of matching systems on consensus test cases.

Two first events were organized in 2004: (i) the Information Interpretation and Integration Conference (I3CON) held at the NIST Performance Metrics for Intelligent Systems (PerMIS) workshop and (ii) the Ontology Alignment Contest held at the Evaluation of Ontology-based Tools (EON) workshop of the annual International Semantic Web Conference (ISWC). Then, unique OAEI campaigns occurred in 2005 at the workshop on Integrating Ontologies held in conjunction with the International Conference on Knowledge Capture (K-Cap), in 2006 at the first Ontology Matching workshop collocated with ISWC, and in 2007 at the second Ontology Matching workshop collocated with ISWC+ASWC. Finally, in 2008, OAEI results were presented at the third Ontology Matching workshop collocated with ISWC, in Karlsruhe, Germany⁵.

3.1 General Guidelines

The principles of OAEI evaluations are very simple [8]:

- Organisers provide pairs of ontologies, usually in OWL;
- Participants return the alignments resulting from matching these ontologies with a particular algorithm and a particular configuration (hence we give them the name of the system);
- Organisers evaluate these results with regard to reference alignments, usually the evaluation measures are precision and recall adapted to alignments. The reference alignment can be hidden from the participants or disclosed.

In practice the evaluation is run in three phases:

1. Preparatory phase

Ontologies to be matched and (where applicable) alignments are provided. This gives potential participants the occasion to send observations, bug corrections, remarks and other test cases to the organizers. The goal of this preparatory period is to ensure that the delivered tests make sense to the participants.

⁴ <http://oaei.ontologymatching.org>

⁵ <http://om2008.ontologymatching.org>

2. Execution phase

During the execution phase, participants use their systems to automatically match the ontologies from the test cases. Participants have been asked to use one algorithm and the same set of parameters for all tests in all tracks. It is fair to select the set of parameters that provide the best results (for the tests where results are known). Beside parameters, the input of the algorithms must be the two ontologies to be matched and any general purpose resource available to everyone, i.e., no resource especially designed for the test. In particular, the participants should not use the data (ontologies and reference alignments) from other test sets to help their algorithms.

In most cases, ontologies are described in OWL-DL and serialized in the RDF/XML format. The expected alignments are provided in the Alignment format expressed in RDF/XML. Participants also provide the papers that are published hereafter and a link to their systems and their configuration parameters.

3. Evaluation phase

The organizers evaluate the alignments provided by the participants and return comparisons on these results. In order to ensure that it is possible to process automatically the provided results, the participants are requested to provide (preliminary) results. In the case of blind tests only the organizers do the evaluation with regard to the withheld reference alignments.

The standard evaluation measures are precision and recall computed against the reference alignments. For the matter of aggregation of the measures we use weighted harmonic means (weights being the size of the true positives). This clearly helps in the case of empty alignments. Another technique that has been used is the computation of precision/recall graphs so it is advised that participants provide their results with a weight to each correspondence they find. New measures addressing some limitations of precision and recall have also been used for testing purposes as well as measures compensating for the lack of complete reference alignments.

3.1.1 Typical exceptions

There are typical cases which do not correspond to the general guidelines. We present them below to complete the general guidelines even though none of these exceptions occurs in the experiments within this deliverable.

Non OWL ontologies

In case of ontologies expressed in a language different from OWL, the usual practice is to translate them into OWL. This may seem sometimes not following the semantics of the ontologies. However, OAEI is about *ontology* matching. Hence, we cannot expect tools participating to be aware of any kind of resources.

Lack of reference alignment

There are cases in which no reference alignment is available beforehand. In such cases, one way to evaluate the results is to poll the participant results together, to sample them, and to have human evaluators assessing the selected correspondences.

Then, precision and recall can be evaluated on these correspondences. This is described more detailed in chapter 4.3. The obtained precision and recall is called “relative” precision and recall because it is relative to the systems which participated and the opinion of the assessors.

This relativity makes that the results should be taken cautiously. However, they are often correlated with other results.

Different evaluation measures

There are many reasons to think that simple precision and recall are not the best measures for evaluating ontology alignments. Hence, several alternate measures have been proposed over the years: relaxed precision and recall [9], semantic precision and recall [10], etc.

In practice, these measures are usually correlated with precision and recall and are more complex to compute. Hence they are not very much used.

3.2 Application to the realized experiment

The fisheries case study was a typical type of experiments, realized during the OAEI workshop 2008. The ontologies are in OWL and the measures are relative precision and recall. The experiments are described in chapter 4.

4 Experiment of the fisheries case study (within OAEI 2008)

The Food and Agriculture Organization of the United Nations (FAO) collects large amounts of data about all areas related to food production and consumption, including statistical data, e.g., time series, and textual documents, e.g., scientific papers, white papers, project reports. For the effective storage and retrieval of these data sets, controlled vocabularies of various types (in particular, thesauri and metadata hierarchies) have extensively been used. Currently, this data is being converted into ontologies for the purpose of enabling connection between data sets otherwise isolated from one another. The FAO test case aims at exploring the possibilities of establishing alignments between some of the ontologies traditionally available. We chose a representative subset of them, that we describe below.

4.1 Used Ontologies

The FAO task involves the three following ontologies [3]:

- AGROVOC⁶ is a thesaurus about all matters of interest for FAO, it has been translated into an OWL ontology as a hierarchy of classes, where each class corresponds to an entry in the thesaurus. For technical reasons, each class is associated with an instance with the same name. Given the size and the coverage of AGROVOC, we selected only the branches of it that have some overlap with the other considered ontologies. Then we selected the fragments of AGROVOC about “organisms,” “vehicles” (including vessels), and “fishing gears.”
- ASFA⁷ is a thesaurus specifically dedicated to aquatic sciences and fisheries. In its OWL translation, descriptors and non-descriptors are modeled as classes, so the ontology does not contain any instance. The tree structure of ASFA is relatively flat, with most concepts not having subclasses, and a maximum depth of 4 levels. Concepts have associated annotations, each of which containing the English definition of the term.
- Two specific fisheries ontologies in OWL⁸, modelling coding systems for commodities and species, used as metadata for statistical time series. These ontologies have a fairly simple class structure, e.g., the species ontologies has one top class and four subclasses, but a large number of instances. They contain instances in up to 3 languages (English, French and Spanish).

Based on these ontologies, participants were asked to establish alignments between:

1. AGROVOC and ASFA (from now on called agrasfa),
2. AGROVOC and fisheries ontology about biological species (called agrobio),
3. The two ontologies about biological species and commodities (called fishbio).

⁶ http://www.fao.org/aims/ag_intro.htm

⁷ <http://www.fao.org/fishery/asfa/8>

⁸ <http://www.fao.org/aims/neon.jsp>

The assignment has been reduced to these three tasks in order to reduce the workload of participants and evaluators. We found them representative enough of the whole task. In the following we will refer to the agrasfa subtrack, the agrobio subtrack and the fishbio subtrack when we talk about these different alignments tasks.

Given the structure of the ontologies described above, the expectation about the resulting alignments was that the alignment between AGROVOC and ASFA (agrasfa) would be at the class level, since both model entries of the thesaurus as classes. Analogously, both the alignment between AGROVOC and biological species (agrobio), and the alignment between the two fisheries ontologies (fishbio) is expected to be at the instance level. However, no strict instructions were given to participants about the exact type of alignment expected, as one of the goals of the experiment was to find how automatic systems can deal with a real-life situation, when the ontologies given are designed according to different models and have little or no documentation.

The equivalence correspondences requested for the agrasfa and agrobio subtracks are plausible, given the similar nature of the two resources (thesauri used for human indexing, with some overlap in the domain covered).

In the case of the fishbio subtrack this is not true, as the two ontologies involved are about two domains that are disjoint, although related, i.e., commodities and fish species. The relation between the two domains is that a specific species (or more than one) are the primary source of the goods sold, i.e. the commodity. Their relation then is not an equivalence relation but can rather be seen, in OWL terminology, as an object property with domain and range sitting in different ontologies. The intent of the subtrack fishbio is then to explore the possibility of using the machinery available for inferring equivalence correspondence to non conventional cases.

4.2 Used Algorithms

Participants⁹ to the FAO test case were:

- Aroma is an extensional matcher using linguistic and statistical techniques developed by INRIA. Aroma uses association rule mining and thus returns subsumption relationships.
- ASMOV is a terminological, structural and extensional matcher developed by Infotech, USA. The result of individual matchers are aggregated by a weighted sum and ASMOV uses a semantic validator for feeding back the system with consequences and inconsistency detection which leads to adding and suppressing correspondences respectively.
- DSSim is a multi-matcher system using Dempster-Shafer theory for aggregating the results of each matcher. It is developed by Open University. Many of these matchers are based on the terminological part of the ontologies.
- Lily is a combination of three matchers which are dynamically chosen with regard to the ontology to match. In particular when ontologies are large, they are first partitioned, then passed on to different terminological and structural matchers. Finally, a semantic matcher is able to postprocess the obtained alignments to filter correspondences. Lily is developed at Tsinghua University.
- RiMOM is a multi strategy matcher which dynamically selects the techniques to be used in function of ontology characteristics. Most of the techniques are terminological or structural. When several strategies are used, their results are aggregated in a weighted average. RiMOM is also from Tsinghua University.

⁹ Participant in this context stands for the ontology mapping algorithms used by the participant of the OAEI workshop.

- SAMBO is mostly a thesaurus-assisted string matching system developed at Linköping University. Like the other systems it also aggregates several classical techniques, e.g., structural.

All these systems are described in details in the proceedings of the Ontology matching workshop [11].

4.3 Description of realized experiments

All participants but one, Aroma, returned equivalence correspondence only. The non-equivalence correspondences of Aroma were ignored. A sampled reference alignment (A_0) was obtained by randomly selecting a specific number of correspondences from each system and then pooling together.

This sample alignment was evaluated by FAO experts for correctness. This provided a partial reference alignment R_0 . We had two assessors: one specialized in thesauri and daily working with AGROVOC (assessing the alignments of the track agrasfa) and one specialized in fisheries data (assessing subtracks agrobio and fishbio). Given the differences between the ontologies, some transformations had to be made in order to present data to the assessors in a user-friendly manner. For example, in the case of AGROVOC, evaluators were given the English labels together with all available “used for” terms (according to the thesauri terminology familiar to the assessor).

Dataset	retrieved (A^*)	evaluated (A_0)	correct (R_0)	(A_0 / A^*)	(R_0 / A_0)
agrasfa	2588	506	226	.19	.45
agrobio	742	264	156	.36	.59
fishbio	1013	346	131	.26	.38
TOTAL	4343	1116	513	.26	.46

Table 1 Size of returned results and samples.

Table 1 summarizes the sample size per each data set. The second column (retrieved) contains the total number of distinct correspondences provided by all participants for each track. The third column (evaluated) reports the size of the sample in A_0 extracted for manual assessment. The fourth column (correct) reports the number of correspondences found correct by the assessors. The two last columns are the ratio of evaluated correspondences and the ratio of correct correspondences among them.

After manual evaluation, we realized that some participants did not use the correct URI in the agrasfa dataset, so some correspondences were considered as different even though they were actually the same. However, this happened only in very few cases. For each system, precision was computed on the basis of the subset of alignments that were manually assessed, i.e., $A^* \cap A_0$. Hence,

$$P_0(A, R_0) = P(A \cap A_0, R_0) = |A \cap R_0| / |A \cap A_0|$$

The same was considered for recall which was computed with respect to the total number of correct correspondences per subtrack, as assessed by the human assessors. Hence,

$$R_0(A, R_0) = R(A \cap A_0, R_0) = |A \cap R_0| / |R_0|$$

Recall is expected to be higher than actual recall because it is based only on correspondences that at least one system returned, leaving aside those that no system were able to return.

We call these two measures relative precision and recall because they are relative to the sample that has been extracted.

4.4 Results

In the following we show and analyse the quantitative and qualitative results of the experiments.

4.4.1 Quantitative Results

Table 2 summarizes the precision and (relative) recall values of all systems, by subtracks. The third column reports the total number of correspondences returned by each system per subtrack. All non-equivalence correspondences were discarded, but this only happened for one system (Aroma). The fourth column reports the number of alignments from the system that was evaluated, while the fifth column reports the number of correct alignments as judged by the assessors. Finally, the sixth and seventh columns reports the values of relative precision and recall computed as described in 4.3.

System	subtrack	retrieved $ A $	Evaluated $ A \cap A_0 $	Correct $ A \cap R_0 $	RPrecision $P_0(A, R_0)$	RRecall $R_0(A, R_0)$
Aroma	agrasfa	195	144	90	0.62	0.40
	agrobio	2	4	0		
	fishbio	11				
ASMOV	agrasfa	1				
	agrobio	0				
	fishbio	5				
DSSim	agrasfa	218	129	70	0.54	0.31
	agrobio	339	214	151	0.71	0.97
	fishbio	243	166	79	0.48	0.60
Lily	agrasfa	390	105	91	0.87	0.40
MapPSO	Agrobio* ¹⁰	6				
	Fishbio*	16				
RiMOM	agrasfa	743	194	159	0.81	0.70
	agrobio	395	219	149	0.68	0.95
	fishbio	738	217	118	0.54	0.90
SAMBO	agrasfa	389	176	121	0.69	0.53
SAMBOdtf	agrasfa	650	219	124	0.57	0.55

Table 2 Participant results per datasets.

¹⁰ The star next to a task marks the algorithms which matched properties.

One system (MapPSO) returned alignments of properties, which were discarded and therefore no evaluation is provided in the table. The results of ASMOV were also not evaluated because they were too small and their correspondences were not selected. Finally, the evaluation of Aroma is incomplete due to the non equivalence correspondence returned, that were discarded before pooling the results together to create the reference alignment.

4.4.2 Qualitative Results

The sampling method that has been used is certainly not perfect. In particular, it did not allow to evaluate two systems which returned few results (ASMOV and MapPSO). However, the results returned by these systems were not likely to provide good recall.

Moreover, the very concise instructions and the particular character of the test sets, clearly puzzled participants and their systems. As a consequence, the results may not be as good as if the systems were applied to polished tests with easily comparable data sets. This provides a honest insight of what these systems would do when confronted with these ontologies on the web. In that respects, the results are not bad.

From DSSim and RiMOM results, it seems that fishbio is the most difficult task in terms of precision and agrasfa the most difficult in terms of recall (for most of the systems). The fact that only two systems returned usable results for agrobio and fishbio makes comparison of systems very difficult at this stage. However, it seems that RiMOM is the one that provided the best results. RiMOM is especially interesting in this real-life case, as it performed well both when an alignment between classes and an alignment between instances is appropriate. Given the fact that in real-life situations it is rather common to have ontologies with a relatively simple class structure and a very large population of instances, this is encouraging.

5 Experiment of the pharmaceutical case studies

In this chapter we describe both the experiments realized with the ontologies of the invoice domain (5.1) and the experiments realized with the ontologies of the semantic nomenclature domain (5.2). We give an overview of the particular ontologies (chapters 5.1.1 and 5.2.1), the used algorithms (see 5.1.2 and 5.2.2), before we describe the experiments (5.1.3, 5.2.3) and the both quantitative and qualitative results (5.1.4 and 5.2.4).

5.1 Experiment with the Invoice Ontologies

The ontologies used in these experiments are based on the ontologies developed for the invoice part of the pharmaceutical case studies; see [5]. We modified them to improve some inconsistencies and to make them compatible with the NeOn Toolkit.

5.1.1 Used Ontologies

EdifactInvoiceMessageOntology (EIMO)

Brief Description

EIMO is based on the international standard language for electronic data interchange for administration, commerce and transport developed under the United Nations. The modularized part is corresponding to the Invoice Message subset of Edifact.

Ontology Details

URI: <http://www.isoco.com/ontologies/neon/EdifactInvoiceMessageOntology.owl>

Concepts: 163

ObjectProperties: 67

DatatypeProperties: 212

UBLInvoicingOntology (UBLIO) *Brief Description*

UBLIO is based on the Universal Business Language, it was developed using the Ontolog UBL ontology¹¹. This ontology was part of the old InvoiceOntology. To conserve the semantic of this standalone ontology some concepts of DolceUltralight and the InvoicingBackboneOntology were included.

Ontology Details

URI: <http://www.isoco.com/ontologies/neon/UBLInvoiceOntology.owl>

Concepts: 123 (wherefrom 11 of DolceUltralight and 14 of the InvoicingBackboneOntology)

¹¹ <http://ontolog.cim3.net/>

ObjectProperties: 111

DatatypeProperties: 3

PharmalInnovalInvoiceOntology (PIIO)

Brief Description

PIIO is based on the original XML files describing the PharmalInnova Invoice Model. We realized some changes to complete the ontology; adding more properties and restrictions. We combined moreover the concepts of emitted and received invoices to one single concept, because both describe the same real world object of an invoice.

Ontology Details

URI: <http://www.isoco.com/ontologies/neon/PharmalInnovalInvoiceOntology.owl>

Concepts: 31

ObjectProperties: 27

DatatypeProperties: 61

InvoicingBackboneOntology (IBO)

Brief Description

IBO is based on DolceUltralight, IOLight and the W3C-time ontology. Additional concepts, describing business processes are added to modularize the domain of the invoices.

Ontology Details

URI: <http://www.isoco.com/ontologies/neon/InvoicingBackboneOntology.owl>

Concepts: 408

ObjectProperties: 347

DatatypeProperties: 36

AggregatedInvoiceOntology (AIO)

Brief Description

AIO is a bundle of all four described ontologies. Besides the backbone build up by IBO it includes as well the part describing the Edifact Invoice Message, the part of UBL and the PharmalInnova Invoice Model.

Ontology Details

URI: <http://www.isoco.com/ontologies/neon/AggregatedInvoiceOntology.owl>

Concepts: 700

ObjectProperties: 571

DatatypeProperties: 296

For the experiments we matched PIIO, EIMO and UIO each with the InvoicingBackboneOntology and the other way round. The idea is to check how these four ontologies have to be combined with each other to obtain a more complete description of the invoice domain. As the name already shows the InvoicingBackboneOntology is the backbone where the three more specific ontologies have to be imported. Therefore we restrict our experiments to these three combinations.

5.1.2 Used algorithms

The algorithms we used are available within the online version of the Alignment Plugin:

- **TaxoMap** [12]: TaxoMap performs a linguistic similarity measure between labels of concepts. It distinguishes between equivalence, subclass and semantically related relationships.
- **Semanticmapper** [6]: The Semanticmapper is an algorithm capable of discovering mappings between ontology elements. The Semanticmapper has two main components: a linguistic component and a structural one.
- **OLA** [6]: OLA relies on a universal measure for comparing the entities of two ontologies that combines in a homogeneous way all the knowledge used in entity descriptions: it deals with external data types, internal structure of classes as given by their properties and constraints, external structure of classes as given by their relationships to other classes.
- **AROMA** [13]: Aroma is an extensional matcher using linguistic and statistical techniques developed by INRIA. Aroma uses association rule mining and thus returns subsumption relationships.

Apart of the experiments realized with these ontology mapping algorithms we realized experiments with an algorithm relating single concepts with each other: Scarlet. The results can be found in 5.1.4.2.

- **SCARLET** [14]: Scarlet compares two entities and checks for their relation: It searches in Watson to check for ontologies building the bridge between the concepts the user wants to relate and returns their semantic binding.

5.1.3 Realized experiments

We matched the EdifactInvoiceMessageOntology, UBLInvoicingOntology, and PharmaInnovaInvoiceOntology with the InvoicingBackboneOntology. First we defined the mappings manually to obtain the reference alignments. To assign the detected matches we used the three different types of semantic relations defined in the NeOn Mapping Metamodel [2]:

Equivalence: Equivalence states that the connected elements represent the same aspect of the real world according to some equivalence criteria. A very strong form of equivalence is equality, if the connected elements represent exactly the same real world object.

Containment: Containment states that the element in one ontology represents a more specific aspect of the world than the element in the other ontology. Depending on which of the elements is more specific, the containment relation is defined in the one or in the other direction.

Overlap: Overlap states that the connected elements represent different aspects of the world, but have an overlap in some respect. In particular, it states that some objects described by the element in the one ontology may also be described by the connected element in the other ontology

The following tables show the amount of alignments detected by the expert, we use them as reference alignment to calculate recall and precision for the ontology mapping algorithms applied on the different combination of ontologies. Table 3 gives the numbers for the mapping of the

InvoicingBackboneOntology with the EdifactInvoiceMessageOntology. Table 4 shows the obtained alignments for the mapping between the InvoicingBackboneOntology and the PharmaInnovalInvoiceOntology. Finally Table 5 gives the numbers for the mapping of the InvoicingBackboneOntology with the UBLInvoicingOntology.

IBO aligned with EIMO:

	Concepts	Properties
Equivalence	3	1
Containment	4	
Overlap	9	2

Table 3 InvoicingBackboneOntology aligned with EdifactInvoiceMessageOntology

IBO aligned with PIIO

	Concepts	Properties
Equivalence	2	3
Containment	3	1
Overlap	3	2

Table 4 InvoicingBackboneOntology aligned with PharmaInnovalInvoiceOntology

IBO versus UBLIO

	Concepts	Properties
Equivalence	26 ₁₂	1
Containment	4	1
Overlap	4	

Table 5 InvoicingBackboneOntology aligned with UBLInvoicingOntology

¹² The UBLInvoicingOntology uses 11 concepts of DolceUltraLight and 14 concepts of the InvoicingBackboneOntology to create the hierarchical backbone

5.1.4 Results

To compare the mappings assessed by the expert with the results of the algorithms we use Recall and Precision:

$$\text{Recall} = |R \cap P| / |R|$$

$$\text{Precision} = |R \cap P| / |P|$$

With R being the set of alignments detected by the expert and P the set of alignments found by the algorithm; consequently $R \cap P$ is the set of by the algorithm correctly assessed alignments.

5.1.4.1 Quantitative Results

In the following we show the quantitative results for the different algorithms. Table 6 shows the numbers for the TaxoMap algorithm. The results obtained for the alignment between UBLIO and IBO are very good. Especially the Recall is very high. On the other hand no correct maps for PIIO aligned with IBO and only one correct map between EIMO and IBO were obtained.

To give an example for the visualization of the alignments, we show an OntoConto screenshot of the mapping between the InvoicingBackboneOntology and the PharmaInnovaInvoiceOntology, see Figure 1¹³.

	Expert	Retrieved	Correct	Recall	Precision
UBLIO:IBO	23	48	18	0.78	0.375
IBO:UBLIO	23 ¹⁴	34	20	0.87	0.59
PIIO:IBO	8	8	0	0	0
IBO:PIIO	8	7	0	0	0
IBO:EIMO ¹⁵	16	12	1	0.06	0.08

Table 6 Results TaxoMap

¹³ It can only give an idea of the visualization, because the dynamic navigation through the aligned ontologies cannot be demonstrated here.

¹⁴ The matcher is not taking into account the imported ontologies. Therefore the matches between the concepts of Dolce Ultralight are not contained.

¹⁵ The mapping between EdifactInvoiceMessageOntology and the InvoicingBackboneOntology returned an error.

The following table (Table 8) shows the results for the OLA algorithm. The matching between the EdifactInvoiceMessageOntology and the InvoicingBackboneOntology did not work and is therefore missing. This algorithm shows the same behaviour: Good results for the UBLIO alignment with IBO, but bad results for PIIO with IBO.

OLA detects equivalence matches only. For each alignment the algorithm gives a ratio between 0 and 1. For the calculation of Recall and Precision we only evaluated the alignments with a ratio above 0.5. All together we found 98 matches between the InvoicingBackboneOntology and the PharmaInnovalInvoiceOntology, thereof 30 with a ratio above 0.5. The same amount of matches we got for the mapping the other way around. Between the InvoicingBackboneOntology and the UBLInvoicingOntology we found 239 alignments, thereof 91 with a ratio above 0.5.

	Expert	Received	Correct	Recall	Precision
PIIO:IBO	14	30	0	0	0
IBO:PIIO	14	30	0	0	0
UBLIO:IBO	22	91	20	0.9	0.2
IBO:UBLIO	22	91	20	0.9	0.2

Table 8 Results OLA algorithm

Table 9 shows the quantitative results for the AROMA algorithm. It shows the same behavior as the three algorithms before: Good results for the UBLInvoicingOntology, but bad results for the EdifactInvoiceMessageOntology and the PharmaInnovalInvoiceOntology.

We attach a screenshot to demonstrate the visualization of the alignment between the InvoicingBackboneOntology and the EdifactInvoiceMessageOntology, see Figure 2.

	Expert	Retrieved	Correct	Recall	Precision
PIIO:IBO	5	5	0	0	0
IBO:PIIO	5	5	0	0	0
UBLIO:IBO	27	24	24	0,89	1
IBO:UBLIO	27	24	24	0,89	1
EIMO:IBO	4	21	0	0	0
IBO:EIMO	4	20	0	0	0

Table 9 Results of AROMA algorithm

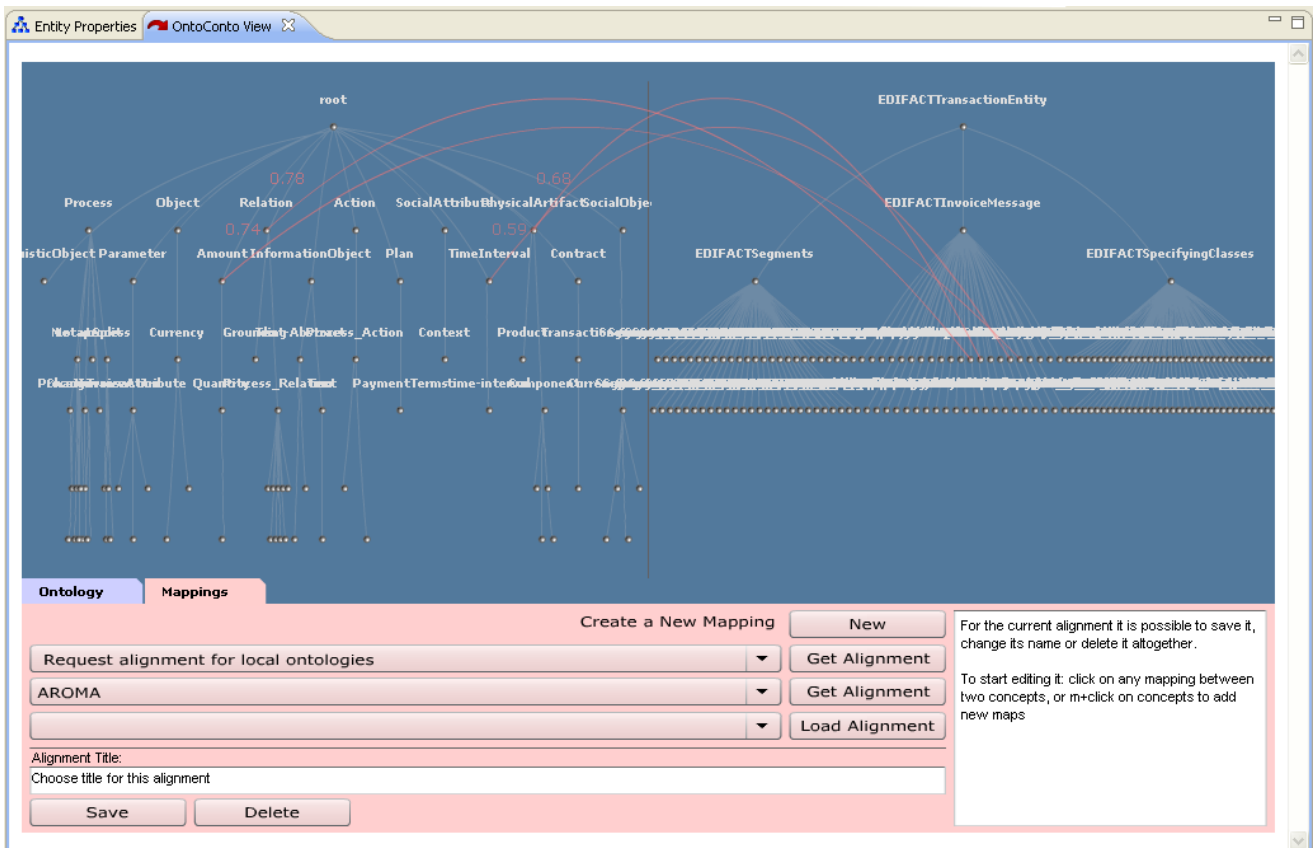


Figure 2: Visualization of the mapping between the IBO and EIMO

5.1.4.2 Qualitative Results

a) TaxoMap algorithm

As shown in Table 6 the results for the mapping between the UBLInvoicingOntology and the InvoicingBackboneOntology are high: Recall and Precision are very high, while the results of the other mappings are far worse. This is easily explained: The UBLInvoicingOntology contains some concepts which are also part of the InvoicingBackboneOntology. These concepts have exactly the same name. The TaxoMap algorithm only evaluates matches as “equivalent” if their labels are equal. An explanation for the bad behaviour of the algorithm for the other alignments is moreover the different use of the same words: The concept “InvoiceType” of the InvoicingBackboneOntology for example is assigned as subclass of “Invoice” of PharmalInnova, which is logically correct. But in this case we know that the “Invoice” concept stands for a specified type of invoice following the PharmalInnova Invoice Model. Therefore the expert evaluated it as subconcept of “InvoiceType”. Another example are the concepts “Term_of_payment” (PharmalInnova) and “PaymentTerms” (InvoicingBackbone). The algorithm describes the first as an “isA”-relation to the second, while the expert evaluated them as equal.

The results of the matching between the UBLInvoicingOntology with the InvoicingBackboneOntology versus the matching of the InvoicingBackboneOntology with the UBLInvoicingOntology suggest the conclusion that it is better to match the bigger ontology to the smaller one than vice versa. But missing the results for the mapping between the EdifactInvoiceMessageOntology and the InvoicingBackboneOntology and considering the bad results of the InvoicingBackboneOntology with the PharmalInnovaInvoiceOntology a concrete presumption cannot be stated.

b) Semanticmapper algorithm

Table 7 shows that the recall for all alignments is far better than the precision. This result is not surprising: The Semanticmapper algorithm proposes possible alignments. The user has to decide in a post-processing step which mapping is the most adequate one. Again it seems that comparing the smaller ontology with the bigger one is more efficient than doing it the other way round.

c) OLA algorithm

The OLA algorithm does only detect the alignments of concepts with exactly identical labels (Table 8). Hence the mapping between the InvoicingBackboneOntology and the UBLInvoicingOntology has a very high recall while the results of the matching between the InvoicingBackboneOntology and the PharmaInnovaInvoiceOntology detects no correct alignments.

d) AROMA algorithm

The AROMA algorithm shows the same effect as the other algorithms (Table 9): Concepts with the same name are detected very well, but other alignments are not found. AROMA has the best precision of all tested algorithms. This suggests the conclusion that AROMA is the best choice for ontologies which use the same labels describing the same concept: The occurrences are detected without producing any noise.

e) Conclusion

It was predictable that the algorithms are not going to return great results for this type of experiment: The used ontologies are describing different aspects of the invoice domain. All of them together describe the whole domain; they complement each other. Because of that some concepts have (almost) the same label, but do not describe exactly the same aspect. An expert understands that and does not match concepts of this type, while an algorithm cannot reason with this problem. Currently most of the algorithms are not able to evaluate the whole structure of the ontologies and in comparison to a human being they cannot use background information about the domain. But nevertheless the algorithms can give great hints, and facilitate the expert's work by proposing pre-selected matches.

f) Scarlet algorithm

Because of the before described nature of the used ontologies the number of correct alignments was not very high. To improve the results we were experimenting with the Scarlet algorithm [14]. Scarlet compares two entities and checks for their relation¹⁶. To do so, it searches in Watson to check for ontologies building the bridge between the concepts the user wants to relate and returns their semantic binding. The problem with this kind of approach for the ontologies used in this experiment is their quite infrequent domain. There is hardly any (free available) ontology existing describing invoices, payment conditions, etc. Therefore the algorithm did not return any results for the terms we were intending to relate (like for example "Invoice" and "Payment" or "Company" and "Invoice") - no ontologies were found to connect the concepts.

¹⁶ An online demo can be found here: <http://scarlet.open.ac.uk/index2.jsp>

5.2 Experiments with the Semantic Nomenclature Ontologies

5.2.1 Used ontologies

Semantic Nomenclature Reference Ontology

The Semantic Nomenclature Reference Ontology is the core of the ontology network used in the case study. This ontology has three main goals: first, act as a bridge between the different application ontologies and domain ontologies; the second goal is to implement one of the requirements like the disambiguate between the clinical drug/branded drug; the third functionality is to act as the application ontology for the Semantic Nomenclature prototype. The Reference Ontology is based on the main recommendations provided by the pharmaceutical community, and also using the semantic model of Snomed as background knowledge, mainly from the Pharmaceutical/Biological product term used in the terminology.

Ontology Details

URI: <http://212.170.156.131:10000/ontologies/ReferenceNomenclature.owl>

Concepts: 49, 7 top-hierarchy concepts – Classification, Clinical_Finding, Event, HealthCare_Entity, Pharmaceutical_Product, Procedure, Substance

Object Properties: 23

Datatype Properties: 12

Instances: 9

Digitalis Ontology

In this ontology, the knowledge of the database Digitalis schema is represented. The main concept is Pharmaceutical_Product that could be the point of link with the reference ontology. This link is possible via a mapping between Digitalis ontology and the Reference Ontology. Other classes represent the main concepts extracted from the tables of the DigitalisDB and the relations represented in their schema model are used to describe with more detail the information around the marketed product and its use.

Ontology Details

URI: <http://212.170.156.131:10000/ontologies/DigitalisOntology.owl>

Concepts: 14, *Active_Ingredient*, *Chemical_Association*, *Composition*, *Dosage*, *Ingredient*, *Ingredient_AI*, *INSALUD_Therapeutical_Subgroup*, *Laboratory*, *Reference_Price*, *OMS_Therapeutical_Subgroup*, *Therapeutical_Subgroup*, *Pharmaceutical_Product*, *Pharmaceutical_Form*, *Status*

Object Properties: 7

Datatype Properties: 45

Instances: External Ontology populated using R2O & ODEMapster

BOTPlus Ontology

The BOTPlus ontology gathers the knowledge represented in the schema of the BOTPlus database. The main concept is Pharmaceutical_Product, that could be one of the concepts that connect via mapping the BOTPlus ontology and the Pharmaceutical Reference Ontology. Moreover, the BOTPlus ontology captures more data than the marketed product information, like information about interactions, pathology, and active ingredients.

Ontology Details

URI: <http://212.170.156.131:10000/ontologies/BOTPlusOnto.owl>

Concepts: 40

Object Properties: 12

Datatype Properties: 76

Instances: External Ontology populated using R2O mappings & ODEMapster

ATC Ontology

Motivation

This ontology represents the ATC classification recommended by the WHO. The ontology has two root concepts: ATC_Code and Group_Code_Part. Despite this, the concept ATC_Classified_Product that represents all the pharmaceutical products classified through the ATC code, is implemented.

Ontology Details

URI: <http://212.170.156.131:10000/ontologies/ATCOntologyv2.owl>

Concepts: 122

Object Properties: 2

Datatype Properties: 3

Instances: 11399 Using R2O mappings & ODEMapster NeOn plugin

RxNorm Ontology

RxNorm is the NLM terminology of standard names and codes for clinical drugs. This terminology links clinical drugs, both branded and generic, to their active ingredients, drug components and related brand names. And also, this terminology connects to a dataset. RxNorm is one of a suite of designated standards used in U.S. Federal Government systems for the electronic exchange of clinical health information.

Ontology Details

URI: <http://212.170.156.131:10000/ontologies/rxnorm.owl>

Concepts: 11

Object Properties: 14

Datatype Properties: 0

UMLS Ontology

The purpose of the UMLS Semantic Network is to provide a consistent categorization of all concepts represented in the UMLS Metathesaurus and to provide a set of useful relationships between these concepts.

Ontology Details

URI: <http://swpatho.ag-nbi.de/owldata/umlssn.owl>

Concepts: 135

Object Properties: 54

Datatype Properties: 1

Instances: 13

Galen Ontology

The GALEN ontology is a result from the OpenGALEN Foundation¹⁷ (a non profit organisation). The main goal of the ontology is to provide terminology and classifications related with the anatomy, surgical deeds, diseases, and their modifiers used in the definitions of surgical procedures.

Ontology Details

URI: <http://www.co-ode.org/galen/full-galen.owl>

Concepts: 23141

Object Properties: 950

SPC Ontology

The Summary of Product Characteristics (SPC) is the basis of information for health professionals on how to use the medicinal product safely and effectively. As a result of an ontology learning from the SPC template, we obtained a new ontology to describe branded drugs, with the purpose of enriching the Semantic Nomenclature ontology network from a domain level perspective. The core of the ontology is the Medicinal_Product concept, with is described by the different properties detailed in the SPC template.

Ontology Details

URI: <http://212.170.156.131:10000/ontologies#SPCOntology.owl>

Concepts: 31

Object Properties: 12

Datatype Properties: 11

Instances: 4

5.2.2 Used algorithms

The algorithms we used are the same as in the invoicing case study, except the Semanticmapper algorithms which was excluded, because of the low precision being expected. These algorithms are available within the online version of the Alignment Plugin. The description of the algorithms can be found in section 5.2.2; here they are only listed:

- **TaxoMap**
- **OLA**
- **AROMA**

¹⁷ <http://www.opengalen.org/>

5.2.3 Realized Experiments

In the semantic nomenclature case study we evaluate the alignment between the application ontologies vs. the reference ontology and, the domain ontologies vs. the reference ontology. In [4] we introduced that we include some manual mapping axioms between the reference ontology and the domain/application ontologies. In this case, we can describe equivalence mappings: between clinical drugs concepts (domain vs. reference) and branded drugs concepts (application vs. reference). Moreover, we can define Containment mappings between Clinical_Drug (reference ontology) and the branded drugs (application ontologies).

For the evaluation, we have two different types of experiments: first, check if the different alignment methods detect the mappings we have defined manually; second, which other mappings are suggested for the rest of the objects involved in the ontologies. In both cases, we only expect mappings at the class level

5.2.4 Results

To compare the mappings assessed by the experts with the results of the algorithms we use Recall and Precision, using the same formula as in the invoicing case study:

$$\text{Recall} = |R \cap P| / |R|$$

$$\text{Precision} = |R \cap P| / |P|$$

With R being the set of alignments detected by the expert and P the set of alignments found by the algorithm; consequently $R \cap P$ is the set of by the algorithm correctly assessed alignments.

5.2.4.1 Quantitative Results

a) TaxoMap algorithm

	Recall	Precision	N. Concepts ¹⁸
Digitalis:NomenclatureReference	0.5	0.25	4:4
BOTPlus:NomenclatureReference	0.66	0.75	14:14
RxNorm:NomenclatureReference	0.5	0.5	2:2
SPCOntology:NomenclatureReference	0.6	0.8	10:10
ATC:NomenclatureReference	0	0	2:2
Galen:NomenclatureReference	0.66	0.375	32:32
UMLS:NomenclatureReference	0.05	0.15	20:20

Table 10 Results TaxoMap

¹⁸ This shows the number of alignments between different concepts of the source and the target ontology. 4:4 for examples shows, that there were four different concepts related with each other. 7:5 on the other hand shows that seven concepts of one ontology were mapped to 5 concepts of the second ontology. It only takes the concepts into account, not other objects of the ontologies like object properties or datatype properties.

b) OLA algorithm

	Recall	Precision	N. Concepts
Digitalis:NomenclatureReference	0.5	0.25	7:5
BOTPlus:NomenclatureReference	0.5	0.43	10:10
RxNorm:NomenclatureReference	0.33	0.29	5:5
SPCOntology:NomenclatureReference	0.15	0.45	16:16
ATC:NomenclatureReference	0	0	0:0
Galen:NomenclatureReference	0.05	0.11	42:47
UMLS:NomenclatureReference	---	---	---

Table 11 Results OLA algorithm

c) AROMA algorithm

	Recall	Precision	N. Concepts
Digitalis:NomenclatureReference	0.5	0.15	17:15
BOTPlus:NomenclatureReference	0.5	0.17	39:40
RxNorm:NomenclatureReference	0.25	0.23	12:10
SPCOntology:NomenclatureReference	0.33	0.31	32:29
ATC:NomenclatureReference	---	---	
Galen:NomenclatureReference	---	---	
UMLS:NomenclatureReference	---	---	

Table 12 Results of AROMA algorithm

We detected some problems for executing the AROMA algorithm between some ontologies: ATC ontology, Galen and UMLS against the NomenclatureReference. We did not obtain any result or feedback from the Alignment server. The same problem happened in the execution of the OLA algorithm between UMLS and NomenclatureReference.

5.2.4.2 Qualitative Results

a) TaxoMap algorithm

The TaxoMap algorithm discovers alignments at concept level. As is explained before, the nature of the ontologies used in the experiments is similar, but with some different objective, most of them are domain ontologies in the pharmaceutical sector, and two of them are application ontologies (Digitalis and Botplus). The number of the alignments discovered is not too high, depending on the size of the source and target ontologies, but the precision of the results is better than the precision of the other algorithms. The recall and precision results show that the quality of the mappings suggested by the algorithm is very accurate, and does not provide a large set of possible mappings that have to be post-processed. Even though the rates are not too high, this algorithm obtains the best results in the Recall metric. Moreover, the results provided by the alignment server using this method, reports

information about the type of mapping (isClose, isA, =) and this is very useful for the domain expert when checking if the suggested mapping are correct or not.

a) OLA algorithm

The number of mappings suggested by this algorithm is higher than the number of mappings returned by the other algorithms used in this evaluation. But, the rate of the recall and precision metrics drop and the level of satisfaction of the results is not too optimistic. In the semantic nomenclature case study, this algorithm offers the worst results in the evaluation. The main reason is that the algorithm tries to discover alignments between any type of object of the ontologies, based on search identical labels, providing a large number of possible mappings, but not the proper ones. This provokes that the metrics drop and the precision was not adequate.

b) AROMA algorithm

One of the results extracted from the AROMA alignment test was that all the results are equivalence mappings. Taking this into account, the results were not too satisfactory, because the algorithm discovers more alignments than TaxoMap, but the precision and recall of the results were worse. As the number of mappings proposed increased, the post-processing step (accept/reject mappings) involves more time. Another detail extracted from these tests was the fact that the mapping suggestions were not only done at the concept level, but also at the datatype property and the object property level.

As final note, the Semanticmapper algorithm was included in the experiment initially. But the results obtained were not satisfactory, because it was returned a large list of suggested mappings. This caused a very low precision.

We found out, that two of the evaluated algorithms do not only search for mappings at a concept level, but also in other types of ontology objects. In a next series of experiments the results could be classified depending on the type of object of the ontology which is aligned.

6 Conclusions and ongoing work

In this deliverable we evaluated several ontology mapping algorithms using ontologies of three different domains. In the following we give a short summary of the results:

- a) The fisheries ontologies: Only two algorithms could perform all three different alignments: DSSim and RiMOM. Considering all the difficulties it seems that RiMOM is the algorithm providing the best results. It performed well both when an alignment between classes and an alignment between instances was appropriate. Given the fact that in real-life situations it is rather common to have ontologies with a relatively simple class structure and a very large population of instances, this is encouraging.
- b) The invoice ontologies: The Aroma algorithm returned the best results both in respect to the recall and the precision. Nevertheless the results were only good for the mapping between the InvoicingBackboneOntology and the UBLInvoicingOntology. These two ontologies contain several concepts with the same label. We can summarize that all evaluated algorithms have problems to map ontologies describing different aspects of the same domain.
- c) The semantic nomenclature ontologies: For this domain the TaxoMap algorithm returned the best results. The ontologies describe very similar domains, but with different objectives. The number of the alignments discovered is therefore not too high, but the precision of the results is better than the precision of the other algorithms. The algorithm provides very precise alignments and the post-processing is therefore not time consuming. The algorithm reports information which type of semantic relation is found (Equivalence, Containment or Overlap), that provides useful information to the user and is therefore another plus for the user.

We can conclude that all algorithms work quite well if the used labels are similar, but all of them have problems detecting less obvious alignments. Therefore they perform more successful if they are applied to the same domain than applied to ontologies describing different aspects of a domain.

Another point to mention is the performing time. The ontologies we were using in the invoice domain are quite small compared to for example ontologies in the health-sector. But even with these smaller ones calculation time of the algorithms was quite long. Especially in the case of the Semanticmapper the analysis of the proposed mappings took a long time and lots of matches were detected. Therefore the handling (both the calculation and the post-processing) of big ontologies will be very time-consuming. This shows once more the need to create small highly expressive ontologies and the importance of the approach to modularize big ontologies in smaller easier treatable modules.

7 References

1. M. C. Suárez-Figueroa, G. Aguado de Cea, C. Buil, C. Caracciolo, M. Dzbor, A. Gómez-Pérez, G. Herrero, H. Lewen, E. Montiel-Ponsoda, and V. Presutti, NeOn Deliverable D5.3.1: NeOn Development Process and Ontology Life Cycle, 2007
2. P. Haase, R. Palma, J. Euzenat, and M. d'Aquin, NeOn Deliverable D1.1.1: Networked Ontology Model, 2006
3. C. Caracciolo, and A. Gangemi, NeOn Deliverable D7.2.2: Revised and Enhanced Fisheries Ontologies, 2007
4. G. Herrero, and T. Pariente, NeOn Deliverable D8.3.2: Semantic Nomenclature Ontology Network, 2009
5. J. M. Gómez-Pérez, C. Buil, G. Herrero, T. Pariente, A. Baena, J. Candini, and J.C. Dalmacio, NeOn Deliverable D8.3.1: Ontologies for the Pharmaceutical Case Studies, 2007
6. C. Le Duc, M. d'Aquin, J. Barrasa, J. David, J. Euzenat, R. Palma, R. Plaza, M. Sabou, and B. Villazón-Terrazas, NeOn Deliverable D3.3.2: Matching ontologies for context: The NeOn Alignment plug-in, 2008
7. B. Pajntar, D. Mladenić, M. Grobelnik, S. Kohler, and M. Dzbor, NeOn Deliverable D4.5.2: NeOn Toolkit plugin realizing revised and extended techniques for contextualized visualization of ontologies and ontology networks, 2008
8. J. Euzenat, M. Ehrig, and R. García Castro, Towards a methodology for evaluating alignment and matching algorithms Version 1.0., OAEI, 2005, <http://oaei.ontologymatching.org/doc/oaei-methods.1.pdf>.
9. M. Ehrig, and J. Euzenat, Relaxed precision and recall for ontology matching. In Proceedings of the K-Cap workshop on Integrating Ontologies, pages 25–32, Banff (CA), 2005.
10. J. Euzenat, Semantic precision and recall for ontology alignment evaluation, in: Proc. 20th International Joint Conference on Artificial Intelligence (IJCAI), Hyderabad (IN), pp248-253, 2007
11. P. Shvaiko, J. Euzenat, F. Giunchiglia, and H. Stuckenschmidt, Proc 3rd international workshop on ontology matching, Karlsruhe (DE), <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-431/>, 2008
12. F. Hamdi, H. Zargayouna, B. Safar, and C. Reynaud, TaxoMap in the OAEI 2008 alignment contest, http://www.dit.unitn.it/~p2p/OM-2008/oaei08_paper12.pdf, 2008
13. J. David, F. Guillet, and H. Briand, Association Rule Ontology Matching Approach, International Journal on Semantic Web and Information Systems, 2, (2007), no. 2, p 27-49, http://www.inrialpes.fr/exmo/people/jdavid/publies/JDavid_IJSWIS_2007.pdf, 2007
14. M. Sabou, M. d'Aquin, and E. Motta, SCARLET: SemantiC relAtion discoverY by harvesting onLinE onTologies, Demo at European Semantic Web Conference, Tenerife, Spain, http://people.kmi.open.ac.uk/marta/papers/eswcDemo_2008.pdf, 2008