Embedded Benchmarking and Expert Authoring for Ontology Mapping and Alignment Generation

Daniel SONNTAG
German Research Center for Artificial Intelligence, Stuhlsatzenhausweg 3, D-66123 Saarbruecken, Germany

Abstract. This paper presents an embedded benchmarking method for interactive ontology matching and alignment generation by user-friendly and editable HTML test cases. It can be used to compute incremental precision, recall, and F-measure values for supervised benchmarks in special situations, e.g., sequential composition of matchers where user feedback is required. The method demonstrates an extendable evaluation method for exploring a (composite) matcher's performance when embedded into a semantic search architecture using new datasets, new alignment input, or new individual matching algorithm as expert authoring environment. Expert users are involved by generating test cases, supervising initial alignments and parameters to the matching process and by combining matchers into global matching methods and HTML test cases.

Keywords. Ontology Matching, Benchmark Test, Evaluation

Introduction

Ontology mapping finds correspondences between two or more ontologies; the correspondences may stand for equivalence as well as other relations, such as consequence, subsumption, or disjointness between ontology entities. While implementations of mapping algorithms exist and have been evaluated in some test scenarios (see *http://oaei.ontologymatching.org/*), ontology mapping for domain-specific applications remains a challenge. This is due to the fact that in particular the performance of automatic methods depends on dataset-specific factors which may heavily affect the total execution time as well as the mapping quality. Ontology matching results, called alignments, can thus express the relations between the ontologies under consideration with various degrees of accuracy, precision and recall.

The objective of state of the art ontology mapping research therefore includes the development of scalable methods (e.g., by combining very efficient string-based methods with more complex structural methods), and tools for supporting users to tackle the interoperability problem between distributed knowledge sources (e.g., editors for iterative, semi-automatic mapping with advanced incremental visualisations [3]). In addition, cognitive support frameworks for ontology mapping really involve users [2], or try to model a natural language dialogue for interactive semantic mediation [4].

In this paper, we describe a component which is part of a semantic mediator to interactively mediate between a query interpretation created on a application layer and semantic background services based on ontologies in the use cases (i.e., medical image retrieval). More precisely, we implement a tool for domain experts that uses HTML based test cases for benchmarking interactive ontology mapping and alignment generation. Interactivity with the user plays a major role: The mapping between two ontologies will be provided in a semi-automatic way. The test cases' input and output are in convenient, editable HTML format. A characteristic feature of past approaches to ontology mapping is that they attempt to calculate the set of relations (e.g., equality relations) between all classes in a single shot. This means a single algorithm is applied for calculating the similarity between two ontologies and the resulting alignment is difficult to be supervised and edited. We exemplify the HTML benchmarking method by a medical use case example. (More information about the use case can be found in the THESEUS application scenario MEDICO—Towards Scalable Semantic Image Search in Medicine.[1])

Benchmarking Strategy

We develop tools that support the mapping engineer in an interactive way; interactive HTML test documents are employed for interactive semantic mediation for basically three purposes:

1. Interactive benchmark tests on similarity matching algorithms and pipelines;
2. Interactive definition of relevant user feedback;
3. Interactive alignment generation for further use.

(1) Every ontology application benefits from assessing the quality of similarity matching algorithms and pipelines. Although the provided algorithms can be used in multiple applications, the performance may significantly differ among the different scenarios and ontologies. Interactive benchmarks help to gain insight into the application domain and the expected results of the matching processes. (2) In order to optimise the semi-automatic computation of useful alignments, the experts have to explore which user feedback results in best performance enhancement of a matching process. In addition, the quantity of user feedback needed for a certain performance can be explored. In the cases we regard, the provided inputs are initial aligments to, e.g., structural matching algorithms, or selection criteria of threshold parameters. (A straightforward extension is to additionally control the algorithm parameters.) (3) The result of the alignment process should be reusable by other applications: once a qualitative alignment has been established (either manually, semi-automatically, or automatically), it has to be stored in an appropriate format for sharing and reuse among other processes and applications. A central aim is to manipulate (or rate) automatically generated alignments in form of expert authoring. For this purpose, alignments have to be stored in an application-independent format. High quality (expert authored) supervised alignments are important descriptions of an application domain and an important source of evidence for subsequent Semantic Web-related integrations towards a new application domain (such as medical image retrieval). Hence, they have to be easily accessible and editable.

[1] See *http://theseus-programm.de/scenarios/en/medico*.

The proposed benchmarking strategy is complementary to early ontology merging/aligning systems (including the well-known PROMPT and Chimaera) which adopt semi-automatic matching techniques through incremental interaction with users, or recent visual ontology matching tools (e.g., COMA++[2], OntoStudio[3]). This is due to the fact that the aforementioned interactive alignment suites can be used to produce alignment input for our HTML benchmarking tests, either as supervised test cases or initial alignment for specific algorithms. The exchange format is explained in the next section.

The HTML test cases described here will help define rules for semantic mediation since the expert user is involved in the specification of the test cases and provide suggestions on mappings to verify the semantic translations between ontologies. This makes the process of creating and validating mappings interactive and personalised to experts or expert groups. The initial alignments generated with the help of expert users can be used to implement the end user scenario (Interactive Semantic Mediation), whereby the initial aligments can be authored and refined by interactively selecting, clarifying, and disambiguating relationships between ontologies.

Architecture for Embedded Benchmarking and Authoring

Figure 1 outlines the three tier architecture consisting of an application layer, a query model/semantic search layer, and a dynamic knowledge base layer. The latter two are explained in more detail in the following list:

- *Query Model/Semantic Search Layer*:
  To provide access to the knowledge base, a subset of SPARQL[4] can be used (a popular standard used to access RDF and OWL data); the semantic RDF store serves assertions on elements (images and image annotations, i.e., relationships such as *is_part_of*, *has_disease_annotation*, or *has_anatomy*) in the medical datasets. A set of correspondence relations sought after in the process of interactive semantic mediation between concepts signs S_C thus takes $\subseteq S_C(Radlex\_terms \cup ICD9\_terms)$ as input. (The knowledge base consists of images (pixel data) and semantic annotations on these images based on Radlex[5] and ICD-9[6] terms. Relation sign alignment is not addressed here.)
- *Dynamic Knowledge Base Layer*:
  When developing new architectures that support semantic image search and scalable searching solutions, new medical data repositories are to be added dynamically. Basically, this means for industrial applications that additional information and knowledge has to matched to the existing image data repositories based on the semantic Radlex and ICD-9 annotations.

[2] *http://dbs.uni-leipzig.de/Research/coma.html/*
[3] *http://www.ontoprise.de/de/en/home/products/ontostudio.html/*
[4] *http://www.w3.org/TR/rdf-sparql-query/*
[5] *http://www.radlex.org/*
[6] *http://icd9cm.chrisendres.com/*

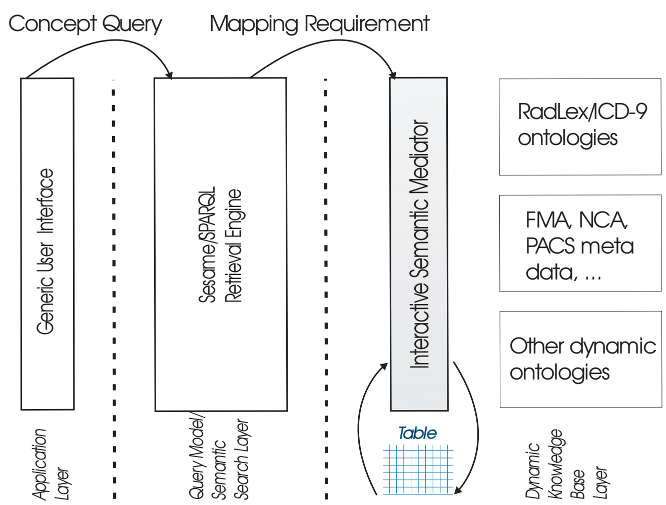


Figure 1. Three tier semantic search architecture accessing different ontologies to be integrated. The knowledge layer hosts the access ontologies and the interactive semantic mediator which is responsible for inducing an appropriate (partial) alignment between two heterogeneous information services, e.g., different ontologies.

Processing Alignments. In the context of offline alignment generation, a simple evidence generator is merely for testing purposes since the unprocessed similarities of a single measure hardly ever produce a reliable alignment. In the research community, better performing methods have been invented (cf. [1] for example). In the context of iterative alignment generation, however, simple evidence alignments provide a framework for showing the benefit of interactive alignment generation and embedded benchmarking. We hypothesise that this is especially true for partial alignments which we will consider in the context of benchmarking matcher combinations. In any case, alignments have to be stored as intermediate results of a matching process in a declarative fashion. We used the alignment format provided by Phaselib (see table 1). For example:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:j.0="http://km.opendfki.de/PHASE#"
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
<j.0:ProposedMappingsList>
<rdf:li>
<j.0:ProposedMappingEqual>
<j.0:FromCategory rdf:resource="ns_o1#term_A"/>
<j.0:ToCategory rdf:resource="ns_o2#term_B"/>
<j.0:Confidence>0.95</j.0:Confidence>
</j.0:ProposedMappingEqual>
</rdf:li>
          ...
</j.0:ProposedMappingsList>
</rdf:RDF>
```

This alignment format is automatically generated from the editable HTML input cases (explained in next paragraph). Vice versa, an HTML input file can be created from this format. This means that additional visual mapping tools (such as COMA++) can be employed in a complementary way in order to provide supervised mapping instances.

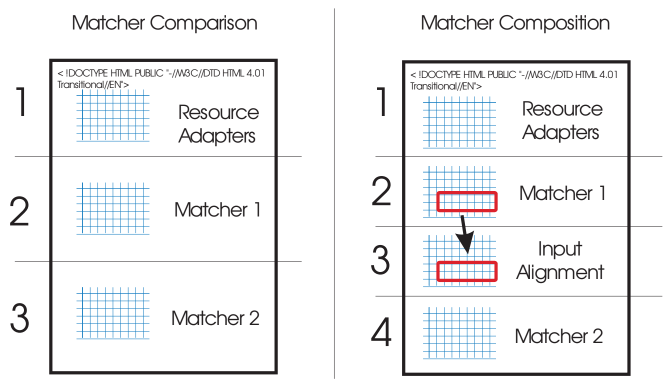
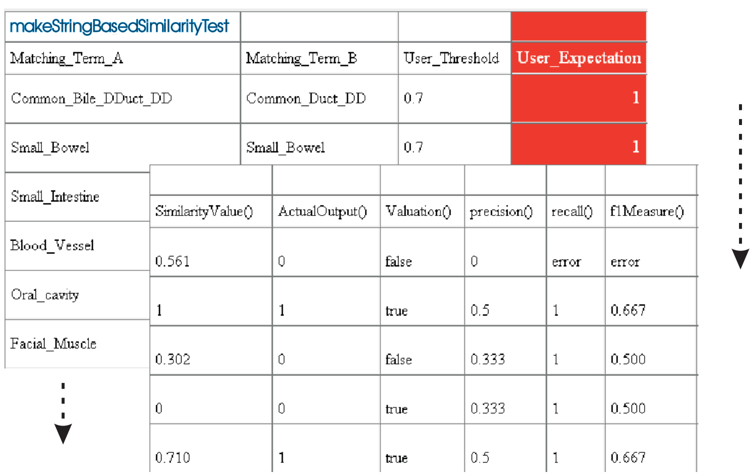


Figure 2. Global methods for comparing and combining local methods by top-down scripting: (Left) Three HTML tables are filled in for matcher comparison; (Right) Four tables are generated for a sequential composition of two matchers. All tables are editable which includes the automatically generated alignment input tables (i.e., Matcher 1) for input alignment based algorithms (i.e., Matcher 2).



makeStringBasedSimilarityTest			
Matching_Term_A	Matching_Term_B	User_Threshold	User_Expectation
Common_Bile_DDuct_DD	Common_Duct_DD	0.7	1
Small_Bowel	Small_Bowel	0.7	1
Small_Intestine			
Blood_Vessel			
Oral_cavity			
Facial_Muscle			

SimilarityValue()	ActualOutput()	Valuation()	precision()	recall()	f1Measure()
0.561	0	false	0	error	error
1	1	true	0.5	1	0.667
0.302	0	false	0.333	1	0.500
0	0	true	0.333	1	0.500
0.710	1	true	0.5	1	0.667

Figure 3. Example table. After running the test cases, the input tables rows (first four columns including a threshold value for acceptance and the highlighted user expectation for supervised tests) are completed by a computed similarity value and incremental precision, recall, and F-Measure values. The table header (makeStringBasedSimilarityTest) specifies the so-called DoFixture (FIT) identifying a table as an ontology resource, matcher, or alignment. (See poster for the complete HTML test.)

Authoring Benchmark Test The basic idea behind the HTML test cases is that you can load and exchange ontologies to match, write supervised benchmark cases, and sequentially combine matchers (which also may require input alignments). The software explained in table 1 enables us to do this in a single editable HTML file. Figure 2 and 3 show how the matcher comparison/composition works, and how the input and output HTML tables look like, respectively. Especially, the matcher composition shows that the second matching phase benefits from pre-compiling the first alignment which is also represented in an automatically generated HTML table to be authored by an expert (e.g., by selecting valid mappings).

Table 1. Java libraries used for the HTML test cases and the implemention of ontology matching pipelines

Open Source Software	Description
Fit (*http://fit.c2.com/*)	FIT is a framework for integrated testing. A test case is written in tables in an HTML file, so that no programming skills are needed for editing. A programmer writes a program which uses the test case's input and checks whether or not the expected output is computed. In our testing cases, we use the main FIT library.
FitNesse (*http://www.bandxi.com/fitnesse/*)	We make use of this extended version of FIT as a testing tool. This extension can be used as a plugin in Eclipse (whereby the procedure of writing the test cases and the respective programs remains the same as in FIT).
Phaselibs (*http://phaselibs.opendfki.de*)	First, this programming platform supports custom combinations of algorithms. Second, it is entirely written in Java which allows us to directly integrate the API with the expert authoring environment based on FIT. Third, the API supports individual modules and libraries for ontology adapters, similarity measures (e.g., string based, instance based, or graph based), and alignment generators.

Conclusion

We described a tool for ontology mapping and alignment generation which includes input alignments and benchmark testing material into a single convenient HTML document. In this way, we increase the transparency and usability of an incremental ontology matching process. The method should be particularly useful in cases where, in response to industry requirements, a collection of reference test sets is not available.

Acknowledgements. This research has been supported by the THESEUS Programme funded by the German Federal Ministry of Economics and Technology (01MQ07016). We would like to thank Pinar Wennerberg, Siemens AG, for providing the medical test cases and useful comments. The responsibility for this publication lies with the author.

References

[1] Marc Ehrig, Steffen Staab, and York Sure. Bootstrapping ontology alignment methods with APFEL. In *WWW '05: Special interest tracks and posters of the 14th international conference on World Wide Web*, pages 1148–1149, New York, NY, USA, 2005. ACM.

[2] Sean M. Falconer, Natalya Noy, and Margaret-Anne D. Storey. Towards understanding the needs of cognitive support for ontology mapping. In Pavel Shvaiko, Jérôme Euzenat, Natalya Fridman Noy, Heiner Stuckenschmidt, V. Richard Benjamins, and Michael Uschold, editors, *Ontology Matching*, volume 225 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2006.

[3] George G. Robertson, Mary P. Czerwinski, and John E. Churchill. Visualization of mappings between schemas. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 431–439, New York, NY, USA, 2005. ACM.

[4] Daniel Sonntag. Towards dialogue-based interactive semantic mediation in the medical domain. In Pavel Shvaiko, Jérôme Euzenat, Fausto Giunchiglia, and Heiner Stuckenschmidt, editors, *Proceedings of the Third International Workshop on Ontology Matching (OM-2008)*, CEUR Workshop Proceedings (to appear). CEUR-WS.org, 2008.