Embedded Benchmarking and Expert Authoring for Ontology Mapping and Alignment Generation

Daniel Sonntag
German Research Center for Artificial Intelligence, Saarbrücken, Germany
daniel.sonntag@dfki.de

Introduction

We propose an extendable evaluation method for exploring ontology matching performances. The tests are embedded into a semantic search architecture and allow to build tests with 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.

Features

- Interactive ontology matching and alignment generation by user-friendly and editable HTML test cases;
- Ontology mapping for domain-specific applications;
- State-of-the-art ontology mapping research should includes the development of scalable methods by

  - combining methods; and

  - tools for supporting users to tackle the interoperability problem between distributed knowledge sources. ->

Convenient editors for iterative, semi-automatic mapping.

Embedded Benchmarking Tests in Search Architecture

Interactive Semantic Mediation for 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.

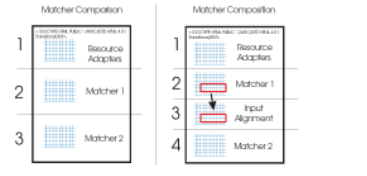


Figure 2. Global methods for comparing and combining local methods by top-down scripting: (Left) Three HTML tables are filled in for matcher comparison; (Right) Four tables are generated for a sequential composition of two matchers. All tables are editable which includes the automatically generated alignment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

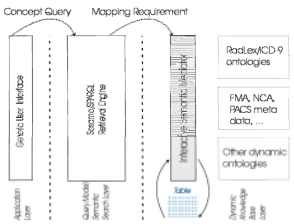
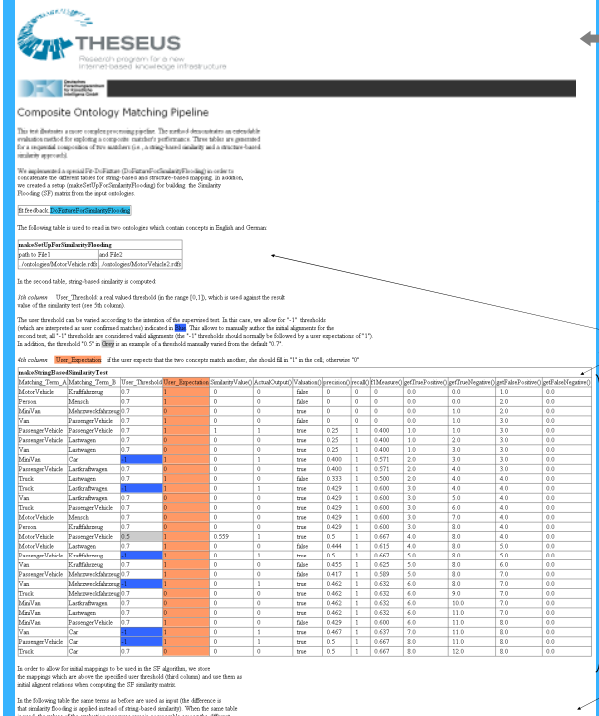


Figure 1. Three tier semantic search architecture accessing different ontologies to be integrated. The knowledge layer hosts the access ontologies and the interactive semantic mediator which is responsible for inducing an appropriate (partial) alignment between two heterogeneous information services,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

HTML Test Case



Evaluation Procedure

You can use input from, e.g., visual ontology matching tools to create test tables.

The expert user is involved in the specification of the test cases and provides suggestions on mappings.This makes the process of creating and validating mappings interactive and personalised to experts or expert groups.

Load and exchange ontologies to match and write supervised benchmark cases.

The user knows about the functional primitives which correspond to functions of a Java API.

Supervise intermediate results, interpret incremental precision, recall, and F-measure values, and sequentially combine matchers (which also may require input alignments)

The second matching phase benefits from pre-compiling the first alignment. Users can add comments; a summary is generated.

Conclusion

We described a tool for ontology mapping and alignment generation.

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.

```
<?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>
```

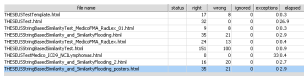
Enabling Technologies

Publicly available testing and matching software can be used.

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.

Write several individual tests for a complete perform. benchmark.



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, *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 Proceeding. CEUR-WS.org, 2008.

Supported by Federal Ministry of Economics and Technology

This research is supported by the German Federal Ministry of Economics and Technology under the grant number 01MQ07016 (THESEUS).
The responsibility for this publication lies with the author.