A Harmony based Adaptive Ontology Mapping Approach

Ming Mao[1], **Yefei Peng**[2], **and Michael Spring**[3]
[1]SAP Research, Palo Alto, CA, USA
[2]Yahoo!, Sunnyvale, CA, USA
[3]School of Information Sciences, University of Pittsburgh, Pittsburgh, PA, USA

Abstract - *Ontology mapping seeks to find semantic correspondences between similar elements of different ontologies. Ontology mapping is critical to achieve semantic interoperability in the WWW. Nowadays most ontology mapping approaches integrate multiple individual matchers to explore both linguistic and structure similarity of different ontologies. Thus how to effectively aggregating different similarities is pervasive in ontology mapping. In current aggregation methods, people either have to manually set parameters in aggregation function or need "ground truth" in advance for machine learning based parameter optimization. Both of them have limitation. In this paper, we propose a measure harmony, which is the normalized number of mapping pair that suggests an unambiguous one-to-one mapping, and a harmony based adaptive ontology mapping approach, which can automatically adjust parameters of three kinds of similarities (i.e., edit distance based similarity, profile similarity and structure similarity) in aggregation functions according to different mapping tasks without given any ground truth. Experimental results show the harmony is indeed a good estimator of the performance (i.e., f-measure) of different similarities, and the harmony based adaptive aggregation method outperforms all other existing aggregation methods on OAEI benchmark tests.*

Keywords: ontology mapping, ontology matching, harmony, adaptive similarity aggregation, prior+

1 Introduction

The World Wide Web (WWW) is widely used as a universal medium for information exchange. However, semantic interoperability in the WWW is still limited due to the heterogeneity of information. Ontology, a formal, explicit specification of a shared conceptualization [8], has been suggested as a way to solve the problem. With the popularity of ontologies, ontology mapping, aiming to find semantic correspondences between similar elements of different ontologies, has attracted many research attentions from various domains. Different techniques have been examined in ontology mapping, e.g., analyzing linguistic information of elements in ontologies [18], treating ontologies as structural graphs [15], using heuristic rules [9] or applying machine learning techniques [2]. Comprehensive surveys of ontology mapping approaches can be found in [4][11][17].

Nowadays most ontology mapping approaches integrate multiple individual matchers to explore both linguistic and structure similarity of different ontologies. However they all face a problem of parameter setting when integrating different similarities. Currently they either use experience numbers or tentatively set parameters in aggregation functions, which is obviously unable to adjust to different mapping tasks, or alternatively applying machine learning techniques to search for an optimized parameter setting, which, however, needs "ground truth" that is usually unavailable in advance in real world cases.

To overcome the problem, we propose a measure harmony, which is the normalized number of mapping pair that suggests an unambiguous one-to-one mapping, and a harmony based adaptive ontology mapping approach, which can automatically adjust parameters of three kinds of similarities (i.e., edit distance based similarity, profile similarity and structure similarity) in aggregation functions according to different mapping tasks without given any ground truth.

To evaluate our approach, we adopt the benchmark tests from OAEI ontology matching campaign 20071. We follow the evaluation criteria of OAEI, calculating the precision, recall and f-measure of over each benchmark test. Experimental results show that the harmony has high correlation with the f-measure of different similarities and the harmony based adaptive aggregation method outperforms all other existing aggregation methods on OAEI benchmark tests.

2 Problem statement

Ontology is a formal, explicit specification of a shared conceptualization in terms of *classes*, *properties* and *relations* [6]. Figure 1 shows two sample ontologies in bibliography area, in which the ellipses indicate classes (e.g., "Reference", "Composite", "Book" and

[1] http://oaei.ontologymatching.org/2007/benchmark

"Proceedings" etc.), the dashed rectangles indicate properties (e.g., "publisher", "editor", "organization" etc.), the lines with arrowhead indicate "subClassof" relation between two classes, and the solid rectangle indicates an instance that is associated with the class of "Monograph" (i.e., "object-oriented data modeling" published by MIT Press at 2000). Each class and property has some information to describe and restrict it, for example, the information next to the bracket of "Book" in the ontologies (e.g., its ID, label, comments and some restrictions such as title, publisher etc.).

Ontology mapping aims to find semantic correspondences between similar elements in two ontologies. The input of an ontology mapping task is two homogeneous ontologies, O_1 and O_2, expressed in formal ontology languages. The output is a list of mapping pairs expressed in the statement of $m(e_{1i}, e_{2j}, r, s)$, where m specifies a specific element e_{1i} in O_1 maps to a certain element e_{2j} in O_2 with a relationship of r, and the mapping holds a confidence measure of s (also known as *similarity*), which is typically normalized in a range of [0..1]. In this paper, r refers to "=" relationship only and *elements* e_{ij} refer to *classes* and *properties* in ontologies. Sample mappings in Figure 1 include: *m(press, publisher, =, .8), m(Reference, Composite, =, .11), m(Monograph, Monography, =, .9), m(Collection, Collection, =, 1)* , *m(Proceedings, Proc, =, .36)* and etc.

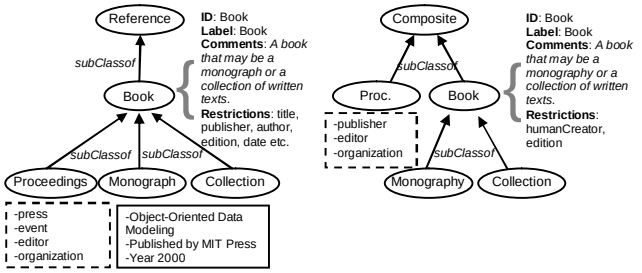

Figure 1. Two bibliographic ontologies

3 Our approach

Figure 2 illustrates the architecture of our approach. First, we parse ontologies using Jena[2], and preprocess them by removing stop words, stemming, and tokenizing. Next, we measure three kinds of similarities, i.e., edit distance based similarity, profile similarity and structure similarity, for each ontology. After that we calculate the harmony for each similarity by counting the number of mapping pairs that suggest unambiguously 1-to-1 mappings, and then we adaptively aggregate three similarities upon their harmonies. Finally we extract mapping results using naive descendant extraction algorithm [14]. In this paper we briefly introduce how we generate three similarities and focus on the harmony and the harmony based adaptive aggregation. More details about the similarity generation can be in our previous work [12][13].

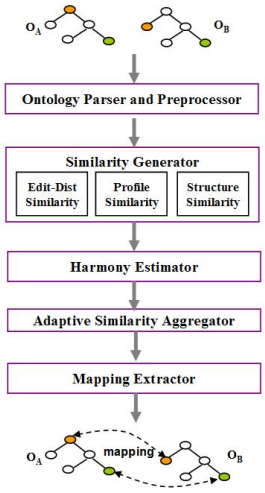

Figure 2. The architecture of PRIOR+

3.1 Similarity generation

The similarity generator generates three kinds of similarities, i.e., edit distance based similarity, profile similarity and structure similarity.

3.1.1 Edit distance based similarity

Edit distance is an intuitive measure of the similarity between elements. In our approach, the edit distance based similarity is calculated using normalized *Levenshtein* distance between the names (i.e. ID) of elements e_{1i} and e_{2j}.

3.1.2 Profile similarity

Though the edit distance based similarity is intuitive, it encounters trouble where synonyms exist or the name of an element is identified using meaningless symbols (e.g., digits). To overcome the problems, we propose the profile similarity, which utilize various descriptive data (e.g. name, label, comments, etc.) to build a profile for each element in ontologies, and thus enrich its information. In particular, the profile of a *class* = the class's ID + label + comments + other restriction + its properties' profiles + its instances' profiles. The profile of a *property* = the property's ID + label + its domain + its range. The profile of an *instance* = the instance's ID + label + other descriptive information. For example, the profile of *class* "Book" and "Proceedings", and the profile of *property* "editor" in the left ontology in Figure 1 are:

[2] http://jena.sourceforge.net/

Profile(Book)=(book, book, book, proceeding, monograph, collection, write, text)
Profile(Proceedings)=(proceeding, press, event, editor, organization)
Profile(editor)=(editor, proceeding, person[3]*)*

Then the *tf·idf* weight (i.e., term frequency and inverse document frequency) [22] is assigned for every term in profiles. When calculating *tf·idf* weights, *each profile* is treated as a *document* and *all profiles* in two ontologies are treated as the *collection of documents*. Afterwards, a set of vectors, each of which represents a profile of an element using a serial of *tf·idf* weights, are output. For example, the *Profile(Book)* can be represented as (.24, .24, .24, .4, .5, .3, .1, .1). Finally we calculate the cosine similarity between two elements in a vector space model (Raghavan and Wong 1986) based on their *tf·idf* weight vectors.

3.1.3 Structure Similarity

Structural similarity is considered for *classes* only. The structural similarity of the classes is calculated based on the normalized difference between the number of the classes' direct properties, the number of the classes' instances, the number of the classes' children, and the depth of the classes from their root. For example, assume the max depth of ontology O_1 is 5, the max depth of ontology O_2 is 6. The depth of element e_{1i} and e_{2j} to their root is 3 and 4 respectively, which will be normalized as 3/5 and 4/6. Then their depth difference is |3/5-4/6|, which can be further normalized by *max*(3/5,4/6) as .1. Finally, the average of the four normalized difference will be calculated as structure similarity.

3.2 Harmony estimation

We define the *harmony* as the normalized number of mapping pairs that suggests an unambiguous one-to-one mapping. The motivation of the harmony is: The ideal 1-to-1 mapping results should be consistent when mapping elements from O_1 to O_2 and vice versa. That is, the similarity score of two truly mapped elements should be the highest (which means unambiguous) comparing to other candidate mapping pairs of them. Therefore if we can estimate the performance of different similarities, then we know which similarity is more reliable and trustful, and thus we can give higher weights to the similarities that are more reliable when aggregating.

Table 1 gives an example of calculating the harmony of an edit distance based similarity matrix, in which the left table lists original similarity scores between each pair of elements of two bibliographic ontologies; the right table illustrates how the harmony is calculated, where "×" denotes the cell that has the highest similarity score in each row, "O" denotes the cell that has the highest similarity score in each column, and "⊗" denotes the overlapped cell that has the highest similarity in both the row and the column, i.e., totally 4. We further normalize it by the max number of elements of two ontologies, i.e., 5. Finally the harmony is *4/5 = .8*. Similarly we estimate the harmony of profile similarity and structure similarity. Please note the order of the elements in the table does not influence the value of the harmony, and the dimension of the matrix could be asymmetric, i.e., $m \times n$. Though there exists the particular case that most mappings holding the highest similarity score are totally wrong but they still result in a good harmony, we ignore it because the three similarities proposed in §3.1 are pretty reasonable based on the analysis of real world tasks and thus it is rare to meet such kind of corner case in practice.

3.3 Adaptive similarity aggregation

Aggregating different similarities is pervasive in ontology mapping systems that contain multiple individual matchers, for example, COMA [1], Falcon-AO [18], RiMOM [19], QOM [3], etc. Many strategies, e.g., *Max*, *Weighted*, *Average* and *Sigmoid*, have been proposed to aggregate different similarities in the approaches. However these strategies either select one extreme end of various similarities to be the representative of the final similarity (e.g., *Max*) or consider the individual similarities equally important and thus can not distinguish differences between them (e.g. *Average*). The *Weighted* strategy overcomes the drawbacks of *Max* and *Average* strategy by assigning relative weights to individual matchers, and the *Sigmoid* strategy emphasizes high individual predicting values and deemphasizes low individual predicting values.

However the *Weighted* strategy needs to manually set aggregation weights using experience numbers and the *Sigmoid* strategy need to tentatively set center position and steepness factor in the *sigmoid* function. Alternatively machine learning based parameter optimization is a solution to the problem of manual parameter setting. However, learning based approach needs extra information, i.e., ground truth, which is usually unavailable in real world mapping tasks.

To overcome the problems, we propose a new weight assignment method to adaptively aggregate different similarities. That is, we use the *harmony* of different similarities as their *weight* when aggregating them. Please see the *HADAPT* method in Table 2 for the definition of our aggregation function. Such method has two advantages: 1. It is easy to adapt to different mapping tasks. 2. It does not need any ground truth in advance.

[3] where the *person* is the *range* of the *editor*, which is not indicated in the Figure 1

Table 1. A sample of harmony calculation

	Composite	Book	Proc	Monography	Collection
Reference	.11	0	**.22**	0.1	.1
Book	0.22	**1**	.2	.2	.2
Proceeding	.18	.09	**.36**	.09	.18
Monograph	.11	.22	.11	**.9**	.1
Collection	**.3**	.2	.1	.1	**1**

Harmony = 4/5 = 0.8

		×		
⊗				
	⊗			
			⊗	
O				⊗

4 Evaluation

4.1 Data sets

To evaluate our approach we use the benchmark tests from OAEI ontology matching campaign 2007[4]. The reason why we choose it is: 1.The annual OAEI campaign has become an authoritative contest in the area of ontology mapping, and thus attracts many participants including both well-known ontology mapping systems and new entrants. 2. The campaign provides uniform test cases for all participants so that the analysis and comparison between different approaches is practical. 3. The ground truth of benchmark tests is open. Thus we can use it to comprehensively evaluate different components of our approach.

The OAEI benchmark tests include 1 reference ontology, dedicated to the very narrow domain of bibliography, and 50 test ontologies, 4 of 50 are real cases and the left 46 are artificially made tasks, each of which discards various information from the reference ontology so as to evaluate how algorithms behave when information is lacking.

4.2 Evaluation criteria

We follow the evaluation criteria from the OAEI campaign, calculating the *precision*, *recall* and *f-measure* over each benchmark test [5]. For the matter of aggregation of the measures weighted harmonic means [5] will be computed.

4.3 Experimental methodology and results

Two experiments are designed. The 1st experiment aims to verify whether the *harmony* reflects the reliability of different similarities, i.e., whether it correlates to the f-measure of different similarities. The 2nd experiment aims to verify the performance of harmony based adaptive aggregation method, i.e., whether it is better than other aggregation methods as discussed in Table 2.

[4] http://oaei.ontologymatching.org/2007/

4.3.1 The correlation between the harmony and the f-measure of different similarities

The experiment methodology is: For each OAEI benchmark test, we calculate 5 similarities, i.e., class edit distance based similarity, class profile similarity, class structure similarity, property edit distance similarity and property profile similarity. Based on each similarity matrix, we extract mapping results using naive descendant extraction algorithm [14]. After that we evaluate the results against the reference alignment and get the f-measure of each similarity. Meanwhile, we estimate 5 harmonies upon its corresponding matrix. Finally we compare the f-measure with the harmony on each test.

The results in Figure 3 show the harmony does linearly correlate with the f-measure of different similarities on both classes and properties. Especially it is a good estimator of f-measure for class' edit distance based similarity, property's edit distance based similarity, and property's profile similarity, the R^2 of which are .9718, .9609 and .9696 respectively.

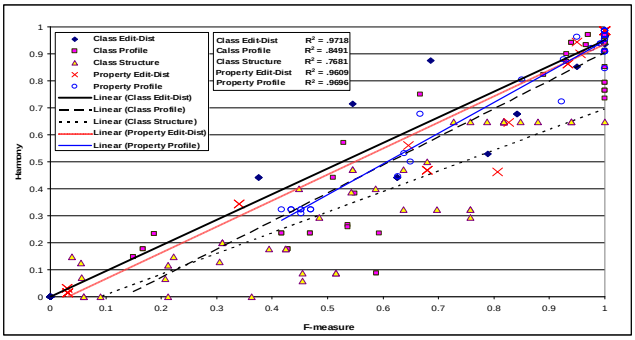

Figure 3 The correlation between the harmony and the f-measure of different similarity

4.3.2 The comparison of different aggregation methods

Similarity aggregation has been researched in many ontology mapping approaches as we discussed in previous section. Data aggregation, called data fusion, has been

widely investigated in information retrieval area [7] as well. To evaluate the harmony-based adaptive aggregation method, short as *HADAPT*, we conduct the experiment comparing its precision, recall and f-measure with 5 aggregation methods selected from both ontology mapping and information retrieval area. Table 2 lists the name and brief description of all aggregation methods used in the experiment, where s_i denotes the i^{th} similarity, fs denotes the final aggregated similarity, h_i denotes the harmony of i^{th} similarity, N denotes the number of individual similarity, Nz denotes the number of non-zero similarities.

The experiment methodology is: For each test, we first calculate three similarities (i.e. name similarity, profile similarity and structural similarity). Then we aggregate them using different aggregation methods as described in Table 2. After that, we extract mapping results using naïve descendant extraction algorithm [14][14]. We then evaluate the results against the reference alignment to get the precision, recall and f-measure on each test. Finally we calculate the precision, recall and f-measure over all tests.

Experiment results in Figure 4 show: 1. The f-measure of profile similarity is slightly better than all aggregated methods except *HADAPT*. The phenomenon tells us aggregation without right parameters can not boost the final result of multiple similarities. 2. The performance of *AVG*, *MAX*, *ANZ*, *MNZ*, and *SIGMOID* methods are competitive with each other. The f-measure of them is around .74-.79. 3. The harmony based adaptive similarity aggregation method (i.e., *HADAPT*) beats all other methods. It holds the highest precision, recall, and f-measure at .92, .83 and .87 respectively. Its improvement of f-measure is 7% and more.

Table 2. Different aggregation methods

Method	Description	Equation
HADAPT	harmony based adaptive aggregation	$fs = \text{sum}(h_i * s_i)/N$
MAX	maximum of individual similarities	$fs = \max(s_i)$
AVG	average of individual similarities	$fs = \text{sum}(s_i)/N$
ANZ	AVG ÷ number of nonzero similarities	$fs = (\text{sum}(s_i)/N)/Nz$
MNZ	AVG × number of nonzero similarities	$fs = (\text{sum}(s_i)/N)*Nz$
SIGMOID	average of individual similarities smoothed by sigmoid function (σ=.5)	$fs = \text{sum}(\text{sigmoid}(s_i))/N$

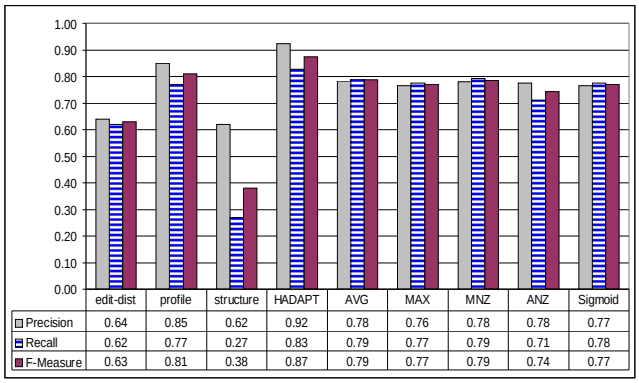

Figure 4. The comparison of different aggregation methods

5 Related work

Different approaches have been proposed to solve the ontology mapping problem. Comprehensive surveys of some famous ontology mapping systems, such as GLUE [2], QOM [3], Similarity Flooding [15], PROMPT [16], can be found in [4][11][17]. Here we only review 4 top-ranked systems that participated in OAEI campaign 2007, i.e., Falcon-AO [18], RiMOM [19], LILY [24] and ASMOV [10]. The reason of reviewing 4 OAEI campaign participants is: 1. The techniques used in 4 systems are diverse and based on the state-of-art approaches. In reviewing these systems, we are reviewing the latest developments in this area. 2. Like ours, all the systems explored multiple similarities, and thus face the problem of effectively aggregating different similarities in an effective way.

Falcon-AO [18] is a similarity-based generic ontology mapping system. It consists of three elementary matchers, i.e., V-Doc, I-Sub [23], and GMO, and one ontology partitioner, PBM. V-Doc constructs a virtual document for each URIref, and then measures their similarity in a vector space model. I-Sub compares the similarity of strings by considering their similarity along with their differences. GMO explores structural similarity based on a bipartite graph. PBM partitions large ontologies into small clusters, and then matches between and within clusters. The profile used in our approach is similar as the virtual document constructed in Falcon-AO. The difference is the virtual document only exploits neighboring information based on RDF model; whereas our profile does not have any limitation of information type, and thus can integrate any information including instance. From the aggregation view, though Falcon-AO measures both linguistic comparability and structural comparability of ontologies to estimate the reliability of matched entity pairs, it only uses them to form three heuristic rules to integrate results generated by GMO and LMO. In LMO Falcon-AO linearly combines two linguistic similarities with some experiential number.

Unfortunately neither experiential number nor heuristic rules can automatically adapt to different test cases. Furthermore, when estimating linguistic comparability Falcon-AO does not distinguish the difference between class and property; whereas our approach estimates harmony for class and property separately.

RiMOM [19] is a general ontology mapping system based on Bayesian decision theory. It utilizes normalization and NLP techniques and integrates multiple strategies for ontology mapping. Afterwards RiMOM uses risk minimization to search for optimal mappings from the results of multiple strategies. The difference between us is when integrating multiple strategies RiMOM adopts a *Sigmoid* function with tentatively set parameters, which has been demonstrated not as good as harmony-based adaptive similarity aggregation (see Figure 4). Furthermore, though RiMOM calculates two similarity factors to estimate the characteristics of ontologies, their estimation is suitable to some special situations only. For example, their linguistic similarity factor only concerns elements that have the same label. However, the harmony in our approach is more general. The harmony does not limit to a specific characteristics of ontologies.

LILY [24] is a generic ontology mapping system based on the extraction of semantic subgraph. It exploits both linguistic and structural information in semantic subgraphs to generate initial alignments. Then a subsequent similarity propagation strategy is applied to produce more alignments if necessary. Finally LILY uses classic image threshold selection algorithm to automatically select threshold, and extract final results based on the stable marriage strategy. One limitation of LILY is that it needs to manually set the size of subgraph according to different mapping tasks and the efficiency of semantic subgraph is very low in large-scale ontologies. Furthermore, as with most mapping approaches, LILY combines all separate similarities with experiential weights.

ASMOV [10] is an automated ontology mapping tool that iteratively calculates the similarity between concepts in ontologies by analyzing four features such as textual description and structure information. It then combines the measures of these four features using a weighted sum. The weights are adjusted based on some static rules. At the end of each iteration, a pruning process eliminates the invalid mappings by analyzing two semantic inconsistencies: crisscross mappings and many-to-one mappings. Due to the limited literature available we are unable to compare our approach with ASMOV in detail. What we can say is the aggregation method in ASMOV is heuristic rule based *Weighted* aggregation method.

6 Conclusions

In the paper we proposed a measure *harmony* to estimate the performance of different similarities without given ground truth, and a harmony-based aggregation method to adaptively aggregate multiple similarities for ontology mapping. Experiment results show the harmony is indeed a good measure to estimate the reliability of different similarities and the harmony-based adaptive aggregation method, *HADAPT*, outperforms all other existing aggregation methods on OAEI benchmark tests.

7 References

[1] Do, H.H., and Rahm, E. COMA - A System for Flexible Combination of Schema Matching Approaches. VLDB 2002, 610-621

[2] Doan, A., J. Madhaven, et al. "Learning to Match Ontologies on the Semantic Web." VLDB Journal 12(4): 303-319. 2002.

[3] Ehrig, M., and Staab, S. QOM - Quick Ontology Mapping. International Semantic Web Conference 2004, 683-697

[4] Euzenat, J., Bach, T., et al. State of the art on ontology alignment, Knowledge web NoE. 2004.

[5] Euzenat, J et al. Results of the Ontology Alignment Evaluation Initiative 2007. In Proceedings of ISWC 2007 Ontology Matching Workshop. Busan, Korea. 2007.

[6] Fensel, D. Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce, Springer-Verlag New York, Inc. 2003.

[7] E.A. Fox and J.A. Shaw. Combination of multiple searches. In Proeddings of the 2nd Text Retrieval Conference (TREC-2), pages 243-252, 1994.

[8] Gruber, T. (1993). "A Translation Approach to Portable Ontology Specifications." Knowledge Acquisition 5(2): 199-220

[9] Hovy, E. Combining and standardizing large-scale, practical ontologies for machine translation and other uses. In Proceedings of the 1st International Conference on Language Resources and Evaluation (LREC), Granada, Spain. 1998.

[10] Jean-Mary Y., Kabuka, M. ASMOV Results for OAEI 2007. In Proceedings of ISWC 2007 Ontology Matching Workshop. Busan, Korea.

[11] Kalfoglou, Y. and M. Schorlemmer "Ontology mapping: the state of the art." Knowledge Engineering Review 18(1):1-31. 2003.

[12] Mao, M. and Peng, Y. The PRIOR: Results for OAEI 2006. In Proceedings of ISWC 2006 Ontology Matching Workshop. Atlanta, GA. 2006

[13] Mao, M., Peng, Y. and Spring, M. A Profile Propagation and Information Retrieval Based Ontology Mapping Approach, In Proceedings of SKG 2007.

[14] C. Meilicke, H. Stuckenschmidt. Analyzing Mapping Extraction Approaches. In Proceedings of ISWC 2007 Ontology Matching Workshop. Busan, Korea.

[15] Melnik, S., H. Garcia-Molina, et al. Similarity flooding: a versatile graph matching algorithm and its application to schema matching. Proc. 18th International Conference on Data Engineering (ICDE). 2002.

[16] Noy, N. and Musen, M. The PROMPT Suite: Interactive Tools for Ontology Merging and Mapping. International Journal of Human-Computer Studies 59: 983-1024

[17] Noy, N. "Semantic Integration: A Survey of Ontology-Based Approaches." SIGMOD Record 33(4): 65-70.

[18] Qu, Y., Hu, W., and Cheng, G. Constructing virtual documents for ontology matching. In Proceedings of the 15th International Conference on World Wide Web. 2006.

[19] Tang, J., Li, J., Liang B., Huang, X., Li, Y., and Wang, K. Using Bayesian Decision for Ontology Mapping. Web Semantics: Science, Services and Agents on the World Wide Web, Vol(4) 4:243-262, December 2006.

[20] V. Raghavan and S. Wong. A Critical Analysis of Vector Space Model for Information Retrieval. Journal of the American Society for Information Science 37(5) (1986) 279-287.

[21] Ross, W. and Philip, H. "Using the cosine measure in a neural network for document retrieval," in Proceedings of the 14th SIGIR Conference. Chicago, IL. 1991.

[22] G. Salton, C. Buckley, Term Weighting Approaches in Automatic Text Retrieval, Cornell University, Ithaca, NY, 1987

[23] Stoilos, G., Stamou, G., and Kollias, S. A String Metric for Ontology Alignment. International Semantic Web Conference 2005, 623-637

[24] Wang, P., Xu, B. LILY: The Results for the Ontology Alignment Contest OAEI 2007. In Proceedings of ISWC 2007 Ontology Matching Workshop. Busan, Korea.