A Large Scale Dataset for the Evaluation of Matching Systems

Fausto Giunchiglia[1], Mikalai Yatskevich[1], and Paolo Avesani[2]

[1] Dept. of Information and Communication Technology,
University of Trento,
38050 Povo, Trento, Italy
{fausto,yatskevi}@dit.unitn.it
[2] ITC-IRST,
38050 Povo, Trento, Italy
avesani@itc.it

Abstract. In the last years the number of ontology matching techniques and systems has significantly increased, and this, in turn, has raised the issue of their evaluation and comparison. One of the key challenges is how to build large scale datasets. In fact the number of possible mappings between two ontologies grows quadratically in respect to the number of nodes in the graphs what, in turn, makes the manual construction of the reference mappings too demanding for large scale real world matching tasks. In this paper we present a new mapping dataset TaxME 2 extracted from Google, Yahoo and Looksmart web directories. TaxME 2 is computed in a semiautomatic way and it is an order of magnitude larger than the state of the art datasets. Moreover, to our knowledge, it is the only large scale dataset which can be used to compute both Precision and Recall. We have evaluated TaxME 2 exploiting results of twelve state of the art matching systems. The evaluation results have shown that the data set has the desired key properties, namely it is discriminative, error-free and hard to solve for state of the art matching systems.

1 Introduction

Match is a critical operator in many applications. It takes two graph-like structures, e.g., lightweight ontologies, such as Google [3] and Looksmart [4], or business catalogs, such as UNSPSC [5] and eCl@ss [6], and produces a mapping between the nodes that correspond semantically to each other. Many diverse solutions to the matching problem have been proposed so far, see for example [2, 19, 18, 16, 10, 6, 14]. This in turn has raised the issues of their evaluation and comparison. One of the key issues in this area is how to build large scale datasets, where a dataset for matching consists of a large set of *reference mappings* holding between two ontologies and in general, graph-like structures. In fact the number of possible mappings grows quadratically with the number

[3] http://www.google.com/Top/
[4] http://www.looksmart.com/
[5] http://www.unspsc.org/
[6] http://www.eclass.de/

of nodes, and this, in turn, makes the manual construction of the reference mappings too demanding for large scale real world matching tasks. One of the largest state of the art manually constructed datasets [7] is composed of several hundreds of reference mappings. The real world part of systematic tests designed in [9] contains tens of them. The reference mappings in these datasets are composed of positive mappings, namely the mappings that hold among the graph structures (e.g., *car* is equivalent to *auto*). All the other mappings are assumed to be negative (e.g., *car* is not related to *tree*). A first attempt to automate the process of the reference mappings acquisition is designed in [1]. This dataset, called TaxME, contains thousands of mappings. However, due to its inherent limitations (see [1, 9] for detailed discussion) TaxME allows only Recall estimation. However, Recall can be easily maximized at the expense of a poor Precision, for instance by returning all possible correspondences, i.e., the cross product of the input graphs.

In this paper we present a new mapping dataset TaxME 2 which has been extracted from Google, Yahoo and Looksmart web directories and has been constructed as an extension of TaxME. Differently from the previous datasets [7, 9] the reference mapping of TaxME 2 do not contain all the positive mappings holding between web directories. TaxME 2 contains two large subsets of positive and negative mappings for an overall number of 4500 mappings, which are computed in a semiautomatic way. TaxME 2 is order of magnitude larger than the state of the art datasets. At the same time, differently from its predecessor TaxME, TaxME 2 allows for the estimation of all commonly used matching quality measures, in particular Precision and Recall. We have evaluated TaxME 2 exploiting results of twelve state of the art matching systems. The evaluation results highlight the key properties of the dataset, namely that it is discriminative, error-free and hard to solve.

The rest of the paper is organized as follows. Section 2 presents a short introduction to the notions of matching and matching evaluation. Section 3 extends the results presented in [1] and discusses the features and properties of TaxME. Section 4 illustrates how TaxME 2 has been constructed by suitably extending TaxME and expanding some of its inherent properties. Section 5 presents the results of our experiments and shows that TaxME 2 posses the desired mapping dataset properties. Section 6 concludes the paper.

2 Basic notions

In order to motivate the matching problem and illustrate one of the possible situations which can arise in the data integration task let us use the (parts of the Google and Yahoo) directories depicted in Figure 1. Suppose that the task is to integrate these two directories. The first step in the integration process is to identify the matching candidates. For example, $Shopping_{O1}$ can be assumed equivalent to $Shopping_{O2}$, while $Board\_Games_{O1}$ is less general than $Games_{O2}$. Hereafter the subscripts designate the directory (either O1 or O2) of the node considered.

We define *matching* as the process of discovering mappings between two graph-like structures through the application of a matching algorithm, where a mapping can be ultimately sought as a pair of nodes that semantically correspond to each other.

Fig. 1. Parts of Google and Yahoo directories

The commonly accepted measures for a quantitative matching evaluation are based on the well known in information retrieval measures of relevance, namely *Precision* and *Recall*. Consider Figure 2; the calculation of these measures is based on the comparison between the mappings produced by a matching system (S in Figure 2) and a complete set of reference mappings H considered to be correct (the area inside the dotted circle in Figure 2). H is usually produced by humans. Here and further we refer to the set of all possible mappings (i.e., cross product of two input graphs) as M. Finally, the correct mappings found by the system are the *true positives*:

$$TP = S \cap H \tag{1}$$

The incorrect mappings found by the system are the *false positives*:

$$FP = S - S \cap H \tag{2}$$

The correct mappings missed by the system are the *false negatives*:

$$FN = H - S \cap H \tag{3}$$

The incorrect mappings not returned by the system are the *true negatives*:

$$TN = M - S \cap H \tag{4}$$

Further we call the mappings in H *positive mappings*, and the mappings in

$$N = M - H = TN + FP \tag{5}$$

negative mappings.

Precision is a correctness measure which varies from [0, 1]. It is calculated as

$$Precision = \frac{|TP|}{|TP + FP|} = \frac{H \cap S}{S} \tag{6}$$

Recall is a completeness measure which varies from [0, 1]. It is calculated as

$$Recall = \frac{|TP|}{|TP + FN|} = \frac{H \cap S}{H} \tag{7}$$

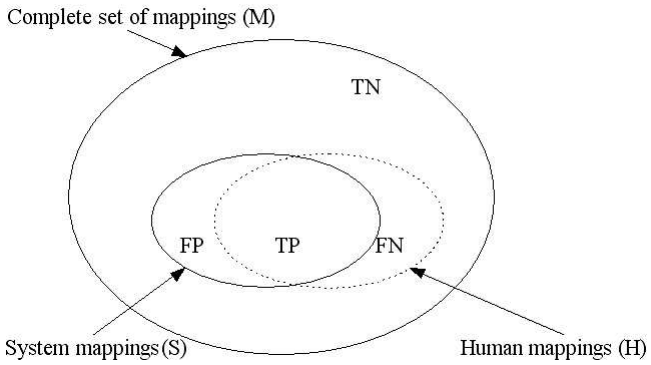

Fig. 2. Basic sets of mappings

However, neither Precision nor Recall alone can accurately evaluate the match quality. In particular, Recall can easily be maximized at the expense of a poor Precision by returning all possible correspondences, i.e. the cross product of two input graphs. At the same time, a high Precision can be achieved at the expense of a poor Recall by returning only few (correct) correspondences. Therefore, it is necessary to consider both measures or a combined measure.

F-measure is a global measure of the matching quality. It varies from [0, 1] and calculated as a harmonic mean of Precision and Recall:

$$F - Measure = \frac{2 * Recall * Precision}{Recall + Precision} \tag{8}$$

Notice that the complete reference mapping H must be known in advance in order to calculate both Precision and Recall. This opens a problem of its acqusition. The problem is that the construction of H is a manual process which, in the case of matching is quadratic in respect to the size of the graphs to be matched. This process turns to be unfeasible for large datasets. For instance, in the dataset we have exploited in this work, namely the Google, Yahoo and Looksmart web directories, each structure has the order of 10^5 nodes. This means that construction of H would require the manual evaluation of 10^{10} mappings.

3 A dataset for evaluating Recall

We compute an approximation of H proposed in [1]. As from [1] we apply the proposed methodology to the Google, Yahoo and Looksmart web directories. The key idea is to rely on a reference interpretation for nodes, constructed by analyzing which documents have been classified in which nodes. The assumption is that the semantics of nodes can be derived from their pragmatics, namely by analyzing the documents that are classified under the given nodes. In particular, the underlying intuition is that two nodes have equivalent meaning if the sets of documents classified under those nodes have a meaningful overlap. The basic idea is therefore to compute the relationship hypotheses based on the co-occurence of documents.

Consider the example presented in Figure 3. Let N_1 be a node in the first taxonomy

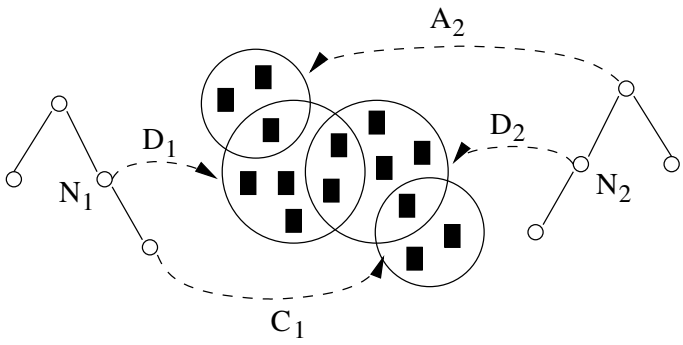

Fig. 3. *TaxME*. Illustration of a document-driven similarity assessment.

and N_2 be a node in the second taxonomy. D_1 and D_2 stand for the sets of documents classified under the nodes N_1 and N_2 respectively. A_2 denotes the documents classified in the ancestor node of N_2; C_1 denotes the documents classified in the children nodes of N_1.

A simple *equivalence* measure is defined as

$$Eq(N_1, N_2) = \frac{|D_1 \cap D_2|}{|D_1 \cup D_2| - |D_1 \cap D_2|} \tag{9}$$

Notice that the range of $Eq(N_1, N_2)$ is [0,∞]. The intuition is that the more D_1 and D_2 overlap the bigger is $Eq(N_1, N_2)$ with $Eq(N_1, N_2)$ becoming infinite with $D_1 \equiv D_2$. Following what described in [1] $Eq(N_1, N_2)$ is normalized to [0,1]. The special case of $D_1 \equiv D_2$ is approximated to 1.

Given the two nodes N_1 and N_2 and the related document sets D_1 and D_2, we introduce two additional sets: (i) the set of documents classified in the ancestor node of N_2, namely A_2, and (ii) the set of documents classified in the children nodes of N_1, namely C_1.

The *generalization* relationship holds when the first node has to be considered more general of the second node. Intuitively, it happens when the documents classified under the first node occur in the ancestor of the second node, or the documents classified under the second node occur in the subtree of the first node. Following this intuition we can formalize the generalization hypothesis as

$$Mg(N_1, N_2) = \frac{|(A_2 \cap D_1) \cup (C_1 \cap D_2)|}{|D_1 \cup D_2|} \tag{10}$$

The *specialization* relationship hypothesis $Lg(N_1, N_2)$ can be easily formulated exploiting the symmetry of the problem.

The TaxME dataset is computed starting from Google, Yahoo! and Looksmart. These web directories hold many interesting properties: they are widely known, they cover overlapping topics, they are heterogeneous, they are large, and they address the same space of contents. All of this makes the working hypothesis of documents co-occurrence sustainable. The nodes are considered as categories denoted by lexical labels, the tree structures are considered as hierarchical relations, and the URLs classified

Table 1. Number of nodes and documents processed in the *TaxME* construction process

Web Directories	Google	Looksmart	Yahoo!
number of nodes	335.902	884.406	321.585
number of urls	2.425.215	8.498.157	872.410

under a given node are taken to denote documents. The following table summarizes the total amount of processed data.

Let us briefly summarize the five steps process used in the *TaxME* reference mapping construction.

Step 1 All three web directories are crawled, their hierarchical structure and their web content;

Step 2 The URLs that do not exist in at least one web directory are discarded;

Step 3 The nodes with a number of URLs under a given threshold (10 in the experiment) are pruned;

Step 4 A manual selection is performed with the goal of restricting the assessment of the similarity metric to the subtrees concerning the same topic; 50 pairs of sub trees are selected;

Step 5 For each of the subtree pairs selected, an exhaustive assessment of correspondences holding between nodes is performed. This is done by exploiting the equivalence metric defined in Eq. 9 and the corresponding generalization and specialization metrics. The TaxME similarity metric is computed as the biggest of the three metrics, namely

$$Sim_{TaxME} = max(Eq(N_1, N_2), Lg(N_1, N_2), Mg(N_1, N_2)) \qquad (11)$$

The distribution of mappings constructed using Sim_{TaxME} is depicted in Figure 4, for varying values of the metric.

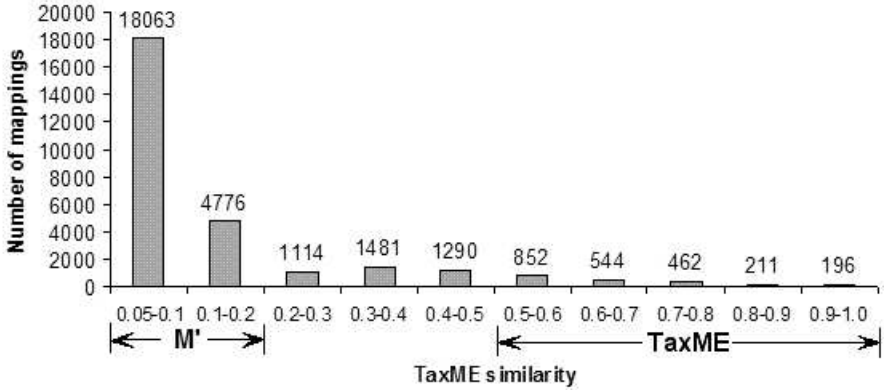

Fig. 4. Distribution of mappings according to TaxME similarity metric

Notice that Sim_{TaxME} is very robust. The number of mappings is in fact very stable and grows substantially, of two orders of magnitude, only with a value of the

metric less than 0.1. As a pragmatic decision, the mappings with Sim_{TaxME} above 0.5 are taken to constitute the reference mapping TaxME. As a result, TaxME is composed from 2265 mappings. Half of them are equivalence relationships and half are generalization relationships. As depicted in Figure 5, TaxME is an incomplete reference mapping since it contains only part of the mappings in H. The key difference between Figures 5 and 2 is the fact that a complete reference mapping (the area inside the dotted circle in Figure 5) is simulated by exploiting an incomplete one (the area inside the dashed circle in Figure 5).

However, if we assume that TaxME is a good representative of H we can use Eq. 7 for an estimation of Recall. In order to ensure that this assumption holds a set of

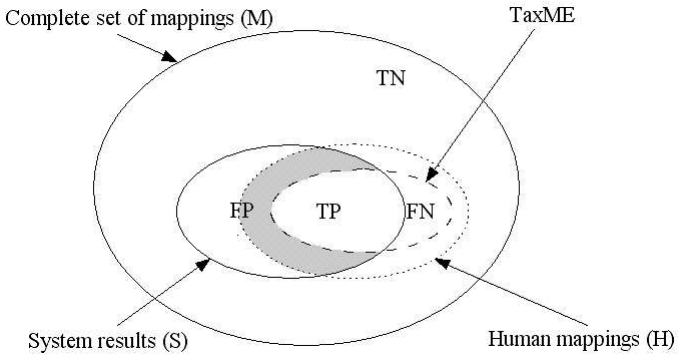

Fig. 5. Mapping comparison using TaxME. TP, FN and FP stand for true positives, false negatives and false positives

requirements have to be satisfied:

1. *Correctness*, namely the fact that $TaxME \subset H$ (modulo annotation errors).
2. *Complexity*, namely the fact that state of the art matching systems experience difficulties when run on TaxME.
3. *Discrimination Capability*, namely the fact that different sets of mappings taken from *TaxME* are hard for the different systems.
4. *Incrementality*, namely the fact that *TaxME* allows for the incremental discovery of the weaknesses of the tested systems[7].

As discussed in [1] *TaxME* satisfies these requirements.

In order to build TaxME 2, however, we need to verify another property of Sim_{TaxME}, namely its robustness. By robustness we mean the fact that the number of incorrect mappings is high only for very low values of Sim_{TaxME} and decreases very sharply as soon as these values increase. We need robustness as it highlights the correspondence between the values of the similarity measure and the human observed similarity. To test the robustness of Sim_{TaxME}, we have randomly selected 100 mappings in 9 intervals of range 0.1 and one interval of range 0.05 as depicted in Figure 6 and manually evaluated their correctness. This resulted in a relatively small amount of manual work as we have analyzed around one thousand of mappings. The results are presented in Figure 6.

[7] We do not consider this property here as insignificant to our goals.

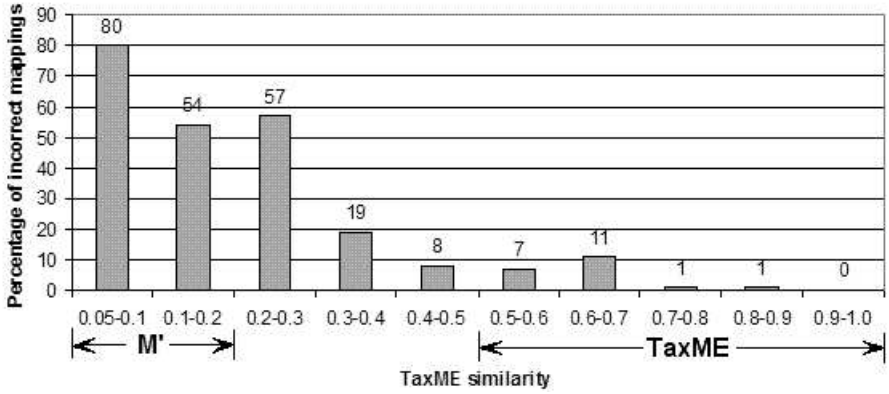

Fig. 6. Distribution of incorrect mappings. Each column is calculated evaluating 100 randomly selected mappings

The results of this manual evaluation show that Sim_{TaxME} is very robust as:

- it is very stable with a small percentage of incorrect mappings for a very large range [0.3,1];
- the number of incorrect mappings becomes substantial for very small values of Sim_{TaxME}, namely with threshold less than 0.1.

4 A dataset for evaluating Precision

As from Eq 6 in order to evaluate Precision, as defined in Eq. 6, we need to know FP, which in turn, as from Figure 2, requires that we know H. However, as from Section 2, computing H in the case of a large scale matching task requires an implausible human effort. Notice also that we can not use an incomplete reference mapping composed from positive mappings i.e, $TaxME$, either. In this case, as shown in Figure 5, FP can not be computed. This is the case because $FP_{unknown} = S \cap (H - TaxME)$, marked as a grey area in Figure 5, is not known.

Our proposal in this paper is to construct a reference mapping for the evaluation of both Recall and Precision, let us call it $TaxME\ 2$, defined as

$$TaxME\ 2 = TaxME \cup N_{T2} \tag{12}$$

where N_{T2} is an incomplete reference mapping containing *only* negative mappings (i.e., $N_{T2} \subset M - H$ in Figure 5). Of course $TaxME\ 2$ must be a good representative of M and therefore satisfy the three requirements described in the previous section and satisfied by *TaxME*. Notice that the request of correctness significantly limits the size of N_{T2} since each mapping has to be evaluated by a human annotator (i.e., $|N_{T2}| \ll |M - H|$). At the same time, N_{T2} must be big enough in order to be the source of meaningful results. Therefore, we require N_{T2} to be at least of the same size as $TaxME$, namely $|N_{T2}| \geq |TaxME|$.

N_{T2} is computed from the complete mapping set M (as from Figure 2) in the following two macro steps:

– *Step 1: Candidate mappings selection.* The goal of this step is to select a set M' where $M' \subseteq M$ which contains a big number of "hard" negative mappings.
– *Step 2: Negative mappings selection.* The goal of this step is to filter all positive mappings from M'. In order to achieve this goal M' is first pruned to the size that allows manual evaluation of the mappings. Finally the negative mappings are manually selected from the remaining mapping set.

Let us describe Step 1 and Step 2 in more detail.

4.1 Candidate mappings selection

The candidate mapping set M' is selected from M, as depicted in Figure 7. The goal of

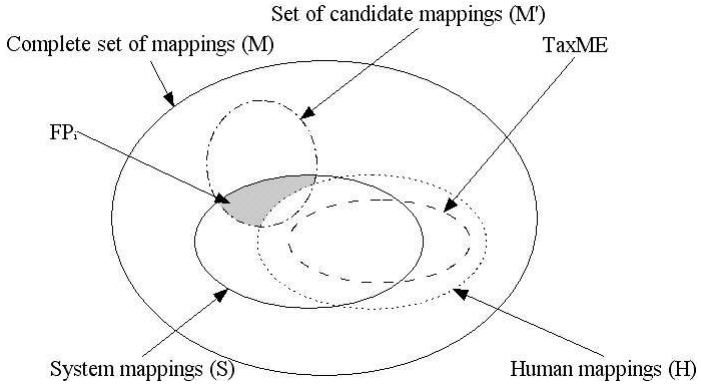

Fig. 7. Mapping sets in *TaxME 2*. Gray area stands for FP_i a set of FP produced by a matching system on M'

this step is to ensure that M' contains a big number of "hard" negative mappings. Intuitively a "hard" negative mapping is the mapping with high value of similarity measure which is incorrect according to manual annotation. Given the robustness of Sim_{TaxME} we have decided to exploit Sim_{TaxME} as the similarity measure for M' construction. Consider Figures 4 and 6. A big enough number of negative mappings can be obtained only for values of Sim_{TaxME} in the 0-0.2 range. As a pragmatic decision we have selected M' as the mappings having Sim_{TaxME} values in the 0.05-0.2 range. As from Figure 4, this allowed us to obtain 18063+4776=22836 candidate mappings.

4.2 Negative mappings selection

The negative mappings selection step is devoted to the computation of N_{T2}. The process is structured as follows:

– *Step 1: Matching systems selection.* The goal of this step is to select a set of matching systems whose results are exploited for constructing N_{T2}. The set of the selected systems should be heterogeneous. By this we mean that the selected systems should make mistakes on different sets of mappings. Thus, the selected systems

have to be the representatives of the different classes of the existing matching techniques. This also prevents N_{T2} from being biased towards a particular class of matching solutions.

– *Step 2: Computation of negative mappings.* The goal of this step is to compute $N_{T2}$by exploiting the results obtained by running the selected matching systems on M'. In particular N_{T2} is computed from FP as $N_{T2} = \bigcup_i FP_i$, where FP_i stands for the FP produced by running the i-th matching system on M' (i.e., incorrect mappings in the set $S \cap M'$). The result of this exercise is depicted in Figure 7, where the grey area stands for FP_i. This construction schema ensures that N_{T2} will be hard for all existing systems and discriminative given that the set of matching systems evaluated on M' is representative and heterogeneous. An implicit constraint is that the number of FPs produced by each of the systems should be comparable. This prevents the existence of a bias towards a particular class of matching solutions. Notice that the computation of FP (as from Eq. 2) requires the human annotation of the systems results.

Based on the classification of the matching systems originally presented in [11] and then largely extended and augmented in [20] [8] as part of Step 1 we have selected three matching systems namely COMA [14], Similarity Flooding (SF) [17] and S-Match (SM) [12]. The first, as from [1, 12], is one of the best syntactic matching systems. The matching process proposed in COMA has been further extended in [5] and parts of it have been reused in the number of matching systems including [15]. SF utilizes a matching algorithm based on the ideas of similarity propagation. SF computes an initial mapping exploiting a string based matcher. Then the mapping is refined using fix-point computation and filtered according to some predefined criteria. The ideas of similarity propagation have been further reused in [10] where the fix point algorithm is exploited for solving the system of linear equations. The SF mapping filtering techniques have been further reused in the system described in [13]. S-Match [9] [12] differs from the SF and COMA as it the implements semantic matching approach, as described in [11], namely it considers rather concepts than labels at nodes and produces a set of semantic relations rather than numerical similarity coefficients [0..1]). Other semantic matching systems, similar to S-Match, are [3, 4].

During Step 2 we have executed COMA, SF and S-Match on M'. We also have manually evaluated the mappings found by the systems and selected the FP from them. Notice that we have not distinguished among different semantic relations while evaluating the matching quality. Therefore, for example, the mapping $A \sqsubseteq B$ produced by S-Match and $A_1 \equiv B_1$ produced by COMA have been considered as TP if $A \equiv B$ and $A_1 \sqsubseteq B_1$ are TP according to the human judgement. Finally we have computed N_{T2} as the union of the FPs produced by the matching systems.

Table 2 provides a quantitative description of the content of N_{T2}, and of the effort needed to build it. As from the first row of Table 2 the total number of annotated mappings was 2553+2163+2151=6867. Notice that this is 6 orders of magnitude lower

[8] See also http://www.ontologymatching.org/

[9] In the evaluation discussed in this paper we have used the basic version of S-Match and not the enhanced version described in [1].

Table 2. Total number of mappings and number of FP computed by COMA, SF and S-Match on M'

	COMA	SF	SM
Found (S)	2553	2163	2151
Incorrect (FP)	870	776	781

than the number of mappings to be considered in the case of complete reference mapping. Notice also that the number of mappings per system is very balanced, as required. Figure 8 shows how the FPs produced by the systems are partitioned.

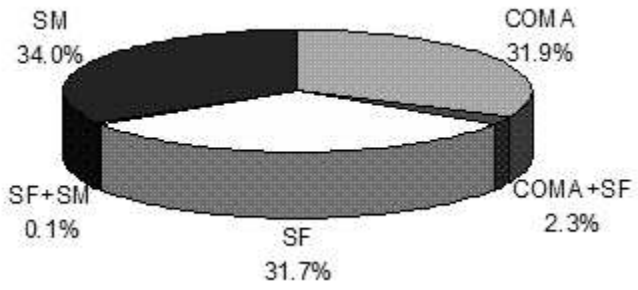

Fig. 8. Partitioning of the FPs computed by COMA, SF and S-Match on M'

As from Figure 8, there are no FPs found by SM, COMA and SF, or even by SM and COMA together. There are the small intersections between the FPs produced by SM and by SF (0.1%) or by COMA and by SF (2.3 %). These results justify our assumption that all 3 systems belong to different classes.

The final result is that N_{T2} consists of 2374 mappings. Notice that the size of N_{T2} is not equal to the sum of the FPs reported in the second row of Table 2 since, as from Figure 8, there is some intersection among these sets. The union of N_{T2} with TaxME has allowed us to compute a reference mapping TaxME 2, good for the evaluation of both Recall and Precision, of 2265+2374=4639 mappings.

5 Evaluating the dataset

In this section we present an evaluation of the Complexity and Discrimination Capability of TaxME 2. In particular we exploit the results of twelve matching systems (Apfel [6], CMS [15], ctxMatch2 [4], OLA [10], OMAP [21] and seven systems participated in OAEI-2006 [8] evaluation). For all the systems we use the default settings or, if applicable, the settings provided by the authors for the OAEI-2005,2006 [9, 8] evaluations. We also compare the results of the matching systems with the results of the systems exploited in the dataset construction process (COMA, SF and SM). The evaluation results, in terms of TP and FP, are presented in Table 3.

Table 3. Number of FP and TP on TaxME 2 dataset

	Apfel	CMS	ctxMatch	OLA	OMAP	COMA	SF	SM	Hmatch	Falcon	Automs	RiMOM	OCM	COMA++	Prior
FP	670	367	299	1356	1113	870	776	781	632	1513	730	1416	712	1343	1085
TP	269	319	298	724	694	876	218	669	303	1030	330	915	356	608	552

5.1 Complexity

Figure 9 presents the Precision of the systems when evaluated on $TaxMe$ 2. As from

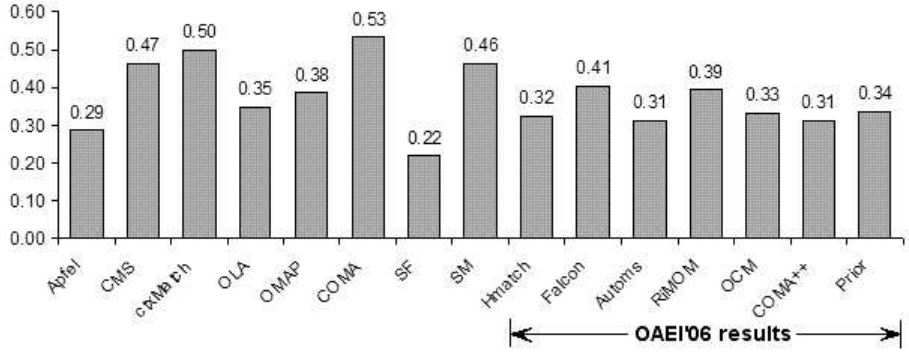

Fig. 9. Evaluation results. Precision on TaxME 2 dataset

Figure 9 the maximum Precision is about 0.5, a value which is significantly lower than the results obtained with the other datasets. For example, the average Precision demonstrated by Falcon, FOAM, CMS and OMAP on the real world part of the systematic tests (problems 301, 302, 303, 304) in the OAEI-2005 evaluation [9] was in the 0.91-0.93 range.

Figure 10 illustrates the Recall of the matching systems while Figure 11 presents the F-Measure as an aggregated matching quality measure. The best F-Measure is 0.45

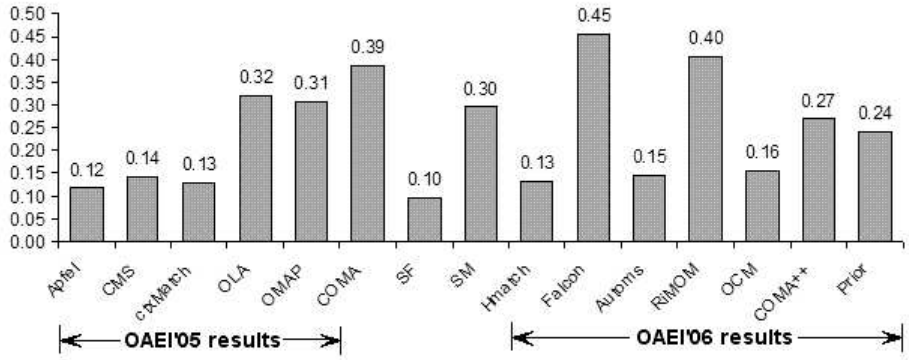

Fig. 10. Evaluation results. Recall on TaxME 2 dataset

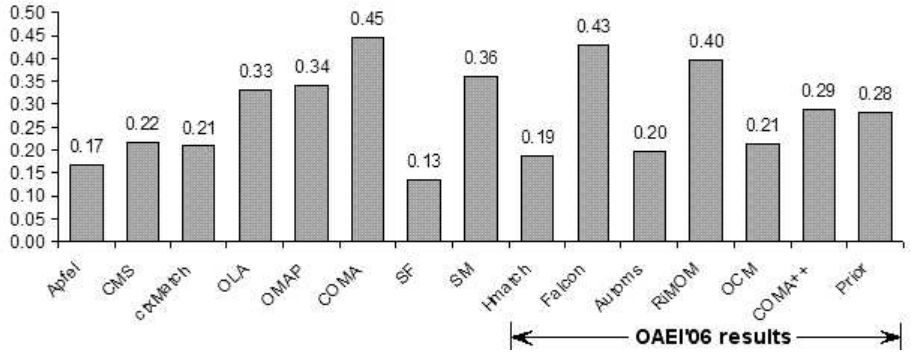

Fig. 11. Evaluation results. F-Measure on TaxME 2 dataset

what is significantly lower than the results demonstrated by the systems on the other datasets.

The results of our evaluation highlight the complexity of TaxME 2. The other interesting observation is that the systems exploited in the dataset construction process demonstrate a performance which is comparable with the other systems. In fact all evaluated systems have experienced the same problems as COMA, SF and SM. This fact justifies that TaxME 2 reflects the inherent properties of real world problems. At the same time, it is still very hard for the state of the art matching systems.

5.2 Discrimination Capability

Consider Figures 12 and 13. They present the partitioning of the FPs and the TPs

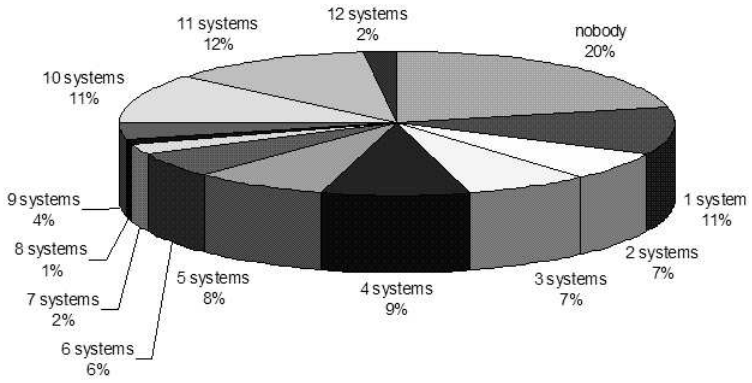

Fig. 12. Partitioning of FPs found by matching systems in TaxME 2 dataset according to the number of systems which found them

in TaxME 2 according to the results of the matching systems. As from Figure 12 all matching systems provided the correct results only for 20% of the FPs while 25% of the FPs are incorrectly found by ten or more matching systems. At the same time 29% of the TPs are not found by any of the matching systems and 65% of the TPs are

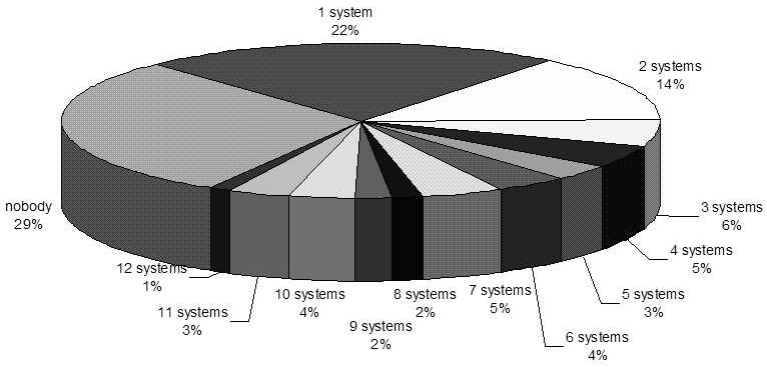

Fig. 13. Partitioning of TPs found by matching systems in TaxME 2 dataset according to the number of systems which found them

found by 2 or less of the matching systems. Figures 12 and 13 illustrate the fact that the different systems experience difficulties on different parts of the dataset (i.e., $TaxME\ 2$ is discriminative or it is hard for the different systems in the different ways).

6 Conclusion and Future Work

In this paper we have presented a large scale mapping dataset constructed starting from Google, Yahoo and Looksmart web directories. The dataset allows for the evaluation of Precision and Recall. Twelve state of the art matching solutions have been evaluated on the dataset. The evaluation results highlight the fact that the dataset posses the key important properties of Complexity and Discrimination capability. Notice that the dataset is correct by construction since the final decision for every mapping is taken by the human annotator.

As a future work we are going to investigate the mapping dataset construction process in the case of expressive ontologies. The other promising direction of research is devoted to the further automation of the mapping dataset construction process. The ultimate goal in this direction is to minimize the human effort while increasing the datasets size.

7 Acknowledgment

This work has been partially supported by the European Knowledge Web network of excellence (http://knowledgeweb.semanticweb.org/) and by STReP Open Knowledge (http://www.openk.org/).

References

1. P. Avesani, F. Giunchiglia, and M. Yatskevich. A large scale taxonomy mapping evaluation. In *Proceedings of International Semantic Web Conference (ISWC)*, 2005.

2. S. Bergamaschi, S. Castano, and M. Vincini. Semantic integration of semistructured and structured data sources. *SIGMOD Record*, (28(1)):54–59, 1999.
3. P. Bouquet, L. Serafini, and S. Zanobini. Semantic coordination: a new approach and an application. In *Proceedings of 2nd international semantic web conference (ISWC 2003)*, Sanibel Island, Florida, 20-23 October 2003.
4. P. Bouquet, L. Serafini, and S. Zanobini. Bootstrapping semantics on the web: Meaning elicitation from schemas. In *Proceedings of 15nd international World Wide Web conference (WWW 2006)*, Edinburgh, UK, 23-26 May 2006.
5. M. Ehrig and S. Staab. QOM - quick ontology mapping. In *Proceedings of the Third International Semantic Web Conference*, pages 683–697, Hiroshima, Japan, November 2004.
6. M. Ehrig, S. Staab, and Y. Sure. Bootstrapping ontology alignment methods with apfel. In *Proceedings of International Semantic Web Conference (ISWC)*, 2005.
7. M. Ehrig and Y. Sure. Ontology mapping - an integrated approach. In Christoph Bussler, John Davis, Dieter Fensel, and Rudi Studer, editors, *Proceedings of the First European Semantic Web Symposium*, volume 3053 of *Lecture Notes in Computer Science*, pages 76–91, Heraklion, Greece, MAY 2004. Springer Verlag.
8. J. Euzenat, Malgorzata Mochol, Ondrej Svab, Vojtech Svatek, Pavel Shvaiko, H. Stuckenschmidt, Willem Robert van Hage, and Mikalai Yatskevich. First results of the ontology alignment evaluation initiative 2006. In *Proceedings of Ontology Matching 2006 Workshop at ISWC'06*, 2006.
9. J. Euzenat, H. Stuckenschmidt, and M. Yatskevich. Introduction to the ontology alignment evaluation 2005. In *Proceedings of K-CAP 2005 Workshop on Integrating Ontologies*, 2005.
10. J. Euzenat and P. Valtchev. Similarity-based ontology alignment in OWL-lite. In *Proceedings of European Conference on Artificial Intelligence (ECAI)*, pages 333–337, 2004.
11. F. Giunchiglia and P. Shvaiko. Semantic matching. *The Knowledge Engineering Review Journal*, (18(3)):265–280, 2003.
12. F. Giunchiglia, P. Shvaiko, and M. Yatskevich. S-Match: an algorithm and an implementation of semantic matching. In *Proceedings of 1st european semantic web symposium (ESWS'04)*.
13. A. Hess. An iterative algorithm for ontology mapping capable of using training data. In *Proceedings of the Third European Semantic Web Conference*, Budva, Montenegro, 2006.
14. H.H.Do and E. Rahm. COMA - a system for flexible combination of schema matching approaches. In *Proceedings of Very Large Data Bases Conference (VLDB)*, pages 610–621, 2001.
15. Y. Kalfoglou and B. Hu. Crosi mapping system (cms). In *Proceedings of K-CAP 2005 Workshop on Integrating Ontologies*, 2005.
16. D. L. McGuinness, R. Fikes, J. Rice, and S. Wilder. An environment for merging and testing large ontologies. In *Proceedings of International Conference on the Principles of Knowledge Representation and Reasoning (KR)*, pages 483–493, 2000.
17. S. Melnik, H. Garcia-Molina, and E. Rahm. Similarity flooding: A versatile graph matching algorithm. In *Proceedings of International Conference on Data Engineering (ICDE)*, pages 117–128, 2002.
18. P. Mitra, N.F. Noy, and A.R. Jaiswal. Ontology mapping discovery with uncertainty. In *Proceedings of International Semantic Web Conference (ISWC)*, 2005.
19. N. Noy and M. A. Musen. Anchor-prompt: Using non-local context for semantic matching. In *Proceedings of workshop on Ontologies and Information Sharing at International Joint Conference on Artificial Intelligence (IJCAI)*, pages 63–70, 2001.
20. P. Shvaiko and J. Euzenat. A survey of schema-based matching approaches. *Journal on Data Semantics*, IV, 2005.
21. U. Straccia and R. Troncy. omap: Combining classifiers for aligning automatically owl ontologies. In *Proceedings of 6th International Conference on Web Information Systems Engineering (WISE'05)*, 2005.