Effective Mapping Composition for Biomedical Ontologies

Michael Hartung[1,2], Anika Gross[1,2], Toralf Kirsten[2], and Erhard Rahm[1,2]

[1] Department of Computer Science, University of Leipzig
[2] Interdisciplinary Center for Bioinformatics, University of Leipzig
{hartung,gross,rahm}@informatik.uni-leipzig.de
tkirsten@izbi.uni-leipzig.de

Abstract. There is an increasing need to interconnect biomedical ontologies. We investigate a simple but promising approach to generate mappings between ontologies by reusing and composing existing mappings across intermediate ontologies. Such an approach is especially promising for highly interconnected ontologies such as in the life science domain. There may be many ontologies that can be used for composition so that the problem arises to find the most suitable ones providing the best results. We therefore propose measures and strategies to select the most promising intermediate ontologies for composition. Experimental results for matching anatomy ontologies demonstrate the effectiveness of our approaches.

Keywords: ontology matching, mapping composition

1 Introduction

In recent years ontologies have become increasingly important in the life sciences [5,18]. For instance, Bio2RDF [3], the OBO Foundry [24] or BioPortal [20,28] distribute a growing number of biomedical ontologies from different domains such as anatomy and molecular biology. The ontologies are primarily used to annotate objects such as proteins, genes or literature to achieve a better information exchange. Often there are different ontologies from one domain containing overlapping or related information. As an example information about mammalian anatomy is available in NCI Thesaurus [19], Adult Mouse Anatomy [1] or the Unified Medical Language System [27]. In such cases *ontology mappings* can be used to express correspondences between different but related ontologies, e.g., which concepts of two different ontologies are equivalent.

Mappings between related ontologies are useful in many ways, in particular for data integration and enhanced analysis [20,14]. They are needed to merge ontologies, e.g., to create an integrated cross-species anatomy ontology such as the Uber ontology [26] or may also be useful to transfer knowledge from different experiments between species [6]. There are already numerous mappings between ontologies available, e.g., BioPortal provides mappings between approx. 300 ontologies. However, there is still a strong need for increasing the number of mappings as most ontologies are interlinked to only one or a few other ontologies.

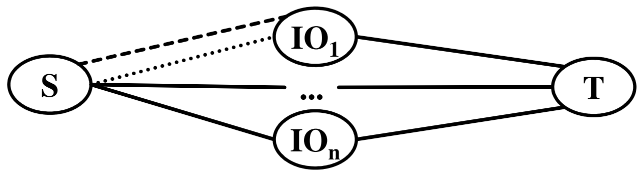


Fig. 1. Mapping composition with multiple intermediate alternatives

Furthermore, new ontologies need to be connected to existing ones. The size of biomedical ontologies makes a manual generation of new mappings unfeasible, hence (semi-) automatic match algorithms are required.

We focus on the reuse and composition of existing mappings between ontologies to indirectly determine new ontology mappings and correspondences. Such an approach is especially promising for the life science domain where many mappings can be reused (e.g., from BioPortal). A main advantage of such a composition approach is its simplicity and high efficiency even for large ontologies. As shown in Fig. 1, one can use multiple alternatives (routes) to establish a new mapping between a source (S) and target (T) ontology using composition. First, there can be multiple intermediate ontologies IO ($IO_1 \ldots IO_n$) leading to questions like: "Is it better to use IO_1 instead of IO_2 or both?". Second, for one single intermediate ontology there can be several alternatives if there are multiple mappings between two ontologies (dotted/dashed lines between S and IO_1), e.g., determined by different match approaches. Considering a large number of possible composition alternatives we need an automatic approach to select the most suitable intermediates that likely result in the best composed mappings.

In this paper we study such selection methods and make the following contributions:

- We propose an efficiently computable measure to determine the effectiveness of composition routes via intermediate ontologies. For the case of composing two mappings, the effectiveness measure helps to find the most promising intermediate ontology.
- We describe two strategies using the proposed effectiveness measure to rank and select the top-k intermediates for mapping composition. Combining the derived mappings for the top-k routes helps to improve the overall mapping quality.
- We evaluate the proposed approach on the OAEI [21] anatomy match task by using existing mappings determined by different match approaches. The obtained mapping quality results demonstrate the effectiveness of the proposed selection strategies.

In Sec. 2 we introduce our ontology and mapping model. Sec. 3 presents the composition-based match approach. We describe our effectiveness measure and outline two strategies for selecting the most promising routes. We evaluate the approach in Sec. 4. After a discussion of related work (Sec. 5), we summarize and outline possible future work.

2 Preliminaries – Ontologies and Mappings

An ontology $O = (C, R, A)$ consists of a set of concepts C which are interrelated by directed relationships R. Each concept has an unique identifier (e.g., accession number, URI) that is used to reference the concept, e.g., the concept 'Vertebra' in NCI Thesaurus is unambiguously referenced by `C12933`. A concept typically has further attributes $a \in A$ to describe the concept, e.g., `C12933` has the name 'Vertebra' and a synonym 'Vertebrae'. A relationship $r \in R$ forms a directed connection between two concepts and has a specific type, e.g., is_a or part_of. In our case `C12933` is a special 'Bone' (`C12366`): [`C12933`, is_a, `C12366`].

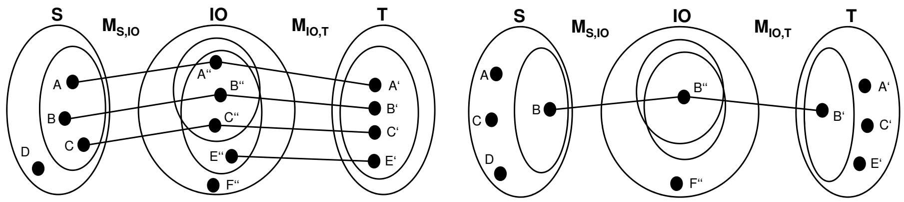
A *mapping* between two ontologies S and T, $M_{S,T} = \{(c_1, c_2, sim) | c_1 \in S, c_2 \in T, sim \in [0,1]\}$, consists of a set of correspondences between these ontologies, e.g., as determined by some ontology match method (see Related Work). Each correspondence interconnects two related concepts c_1 and c_2. Their relatedness is represented by a similarity value sim between 0 and 1 determined by the used match approach. The greater the sim value the more similar are the corresponding objects. Note that we focus on equality correspondences and leave the consideration of other correspondence types for future work. For already validated mappings we assume a similarity of 1 for each correspondence.

3 Rating and Selection of Composition Routes

In this section we present our approach to rate composition routes and to select the most promising ones. After introducing the concept of mapping composition, we propose an effectiveness measure to rate the value of routes in Sec. 3.2. Using this measure we describe the strategies *topKByEffectiveness* and *topKByComplement* for ranking and selecting the routes (Sec. 3.3). We finally describe in Sec. 3.4 the combined use of multiple selected routes to create a new mapping.

3.1 Composition for Generating new Mappings

The general idea behind mapping composition is to derive new mappings between two ontologies by reusing already existing mappings. Thus, new mappings are generated indirectly via one or more intermediate ontologies instead of a direct match between the two input ontologies. The typical situation for one intermediate is depicted in Fig. 2a. The input consists of two ontologies S/T and two mappings $M_{S,IO}/M_{IO,T}$ w.r.t. an intermediate ontology IO. The `domain` and `range` of the mappings can be used to find out which concepts are covered by the given mappings. For instance, all concepts of S covered by the mapping to IO are in its domain: `domain`$(M_{S,IO})$. Similarly, IO concepts covered by this mapping are in its range: `range`$(M_{S,IO})$. Mapping composition is then applied in the following way. A `compose` operator takes as input two mappings (from S/T to IO) and produces new correspondences between concepts of S and T if correspondences share the same concept in IO. The result is a new mapping $M_{S,T}$:



(a) Positive example (b) Negative example

Fig. 2. Examples for applying the effectiveness measure

$$M_{S,T} = \texttt{compose}(M_{S,IO}, M_{IO,T}) =$$
$$\{(c_1, c_2, aggSim(sim_1, sim_2)) | c_1 \in S, c_2 \in T, b \in IO :$$
$$\exists (c_1, b, sim_1) \in M_{S,IO} \wedge \exists (b, c_2, sim_2) \in M_{IO,T}\}$$

The similarity values of input correspondences are aggregated (aggSim) into new similarity values, e.g., by computing their maximum or average.

3.2 Effectiveness of Routes

The result of a mapping composition heavily depends on which intermediate ontologies are used and how the mappings to these intermediates look like. First, compose can at best create correspondences between concepts of S/T that are covered by the input mappings to an IO. The more concepts are covered by an input mapping the more likely it is that they can be interlinked to concepts in the other ontology. Thus, an intermediate for which mappings only cover a small portion of S/T are less effective compared to those covering larger portions. Second, there should be a high overlap of mapped objects in IO, i.e., many IO concepts should be in both $\texttt{range}(M_{S,IO})$ and $\texttt{domain}(M_{IO,T})$. This is because new correspondences can only be created if there are intermediate concepts for the composition. By contrast, a small overlap will only permit the creation of few correspondences, i.e., small and likely incomplete mappings. Based on these observations we define a measure to rate the effectiveness of a route between sources S and T via an intermediate IO:

$$eff(S, IO, T) = \frac{2 \cdot |range(M_{S,IO}) \cap domain(M_{IO,T})|}{|S| + |T|}$$

The measure is largely based on the size of the overlap of concepts in the intermediate ontology, i.e., the larger the overlap the better the effectiveness. Second, we relate this overlap to the sizes of the ontologies to be matched S and T. Only mappings with many correspondences can produce a high overlap and a good coverage of concepts in S and T. Fig. 2 shows two examples for applying the measure. The left example results in a good effectiveness of ($\frac{2 \cdot 3}{4+4}$=0.75) because the overlap in the intermediate ontology covers a large part of S and T. By contrast, in the right example there is only one overlapping concept in the

Algorithm 1: topKByComplement

Input: set of intermediates all_{IO}, input ontologies S and T, number of intermediates to consider k
Output: top intermediates $topK$

$topIO \leftarrow$ `getMostEffectiveIntermediate(`all_{IO}`)`;
$topK$`.add(`$topIO$`)`, all_{IO}`.remove(`$topIO$`)`;
$cov_{all} \leftarrow$ `domain(`$M_{S,topIO}$`)` $\cup$ `range(`$M_{topIO,T}$`)`;
while $|topK| < k$ **do**
  $compl_{max} \leftarrow \emptyset$; $topIO \leftarrow null$;
  foreach $IO \in all_{IO}$ **do**
    $compl_{IO} \leftarrow$ (`domain(`$M_{S,IO}$`)` $\cup$ `range(`$M_{IO,T}$`)`) $\setminus cov_{all}$;
    if $|compl_{IO}| > |compl_{max}|$ **then**
      $compl_{max} \leftarrow compl_{IO}$;
      $topIO \leftarrow IO$;
  $cov_{all} \leftarrow cov_{all} \cup compl_{max}$;
  $topK$`.add(`$topIO$`)`; all_{IO}`.remove(`$topIO$`)`;
return $topK$;

intermediate ontology resulting in a poor effectiveness of $\frac{2 \cdot 1}{4+4}$=0.25. The compose operator would produce the following mappings (without similarity values): (a) $M_{S,T} = \{(A, A'), (B, B'), (C, C')\}$ and (b) $M_{S,T} = \{(B, B')\}$. This shows that the better rated intermediate ontology is able to produce more correspondences and thus a more complete mapping.

3.3 Ranking and Selection of Routes

Mapping composition using only one route may lead to insufficient (incomplete) match results. Composing mappings for several routes via different intermediates and combining their results is likely to improve the mapping to be determined. This is because other intermediate sources may provide additional correspondences between the input ontologies. The question thus arises which of the available routes should be selected for mapping composition. In the following, we describe two selection strategies that we will also evaluate later.

The first strategy *topKByEffectiveness* simply uses a ranking based on the effectiveness measure described in Section 3.2. Hence, we perform composition only on the k most effective routes and combine their results.

The second strategy *topKByComplement* also selects the most effective route but selects the remaining routes based on the number of complementary correspondences they can provide. The strategy determines how much additional gains can be achieved by considering further routes. For instance, if one has to match two anatomy ontologies, an ontology about the skeletal system would be complementary to one about the nervous system or blood circuit. Hence, it makes sense to consider intermediate ontologies that contain additional knowledge that others do not provide.

Algorithm 2: topKComposeMatch

Input: set of possible intermediates all_{IO}, input ontologies S and T, selection strategy *selectionStrategy*, merge strategy *mergeStrategy*, number of intermediates to consider k

Output: mapping between S and T $M_{S,T}$

1 $all_{IO} \leftarrow$ `computeEffectiveness`(all_{IO},S,T);
2 $topK \leftarrow$ `getTopRoutes`(all_{IO},*selectionStrategy*,k);
3 *mapList* $\leftarrow$ *empty*;
4 **foreach** $IO \in topK$ **do**
5     $M_{S,IO} \leftarrow$`getMapping`(S,IO);
6     $M_{IO,T} \leftarrow$`getMapping`(IO,T);
7   *mapList*.`add`(`compose`($M_{S,IO}$,$M_{IO,T}$));
8 **return** `merge`(*mapList, mergeStrategy*);

Alg. 1 shows the implementation of this strategy. It first selects the most effective intermediate based on our effectiveness measure (lines 1–3). It then iteratively (while loop) adds the intermediate possessing the maximum complement ($compl_{max}$) compared to the already covered objects (cov_{all}) in S and T (lines 5–12). Particularly, we compare the covered concepts of the current intermediate with the covered concept set (cov_{all}) from already selected intermediates. In each round we select the intermediate which brings us the maximum complement.

3.4 Overall Composition Algorithm

We use the algorithm *topKComposeMatch* (see Alg. 2) to perform the composition for the k selected intermediates and to combine the composition results to obtain the overall mapping between two input ontologies.

We first apply our effectiveness measure on each route (line 1). Based on the given selection strategy (*topKByEffectiveness*, *topKByComplement*) we filter the top k promising intermediates (line 2). We then iteratively compose the mappings between S and T along each selected intermediate (lines 4–7). The generated mappings are temporarily stored in a *mapList* and are finally merged according to a specified merge strategy, such as union or intersection.

4 Evaluation

We evaluate our approach by composing mappings between anatomy ontologies. In particular, we focus on generating mappings between the Adult Mouse Anatomy (MA) and the anatomy part of NCI Thesaurus (NCIT) which is a challenging task in the yearly OAEI [21] match contest. This has the advantage that we can use the publicly available OAEI gold standard (perfect mapping) to assess the quality of computed mappings (using precision, recall and F-measure) and to compare the achieved results with the published results of other approaches. Furthermore, we can reuse a lot of already existing mappings, in particular mappings provided by BioPortal [28] and mappings that we previously generated using our GOMMA ontology management infrastructure [16].

We first describe our experimental setup in more detail (Sec. 4.1). We then correlate the effectiveness measure with the achieved match results by composing the mappings according to different intermediate ontologies (Sec. 4.2). Finally, we adopt our selection strategies and present results of performing composition-based matching via the most promising intermediate ontologies (Sec. 4.4).

4.1 Experimental Setup

The experiment focuses on generating mappings between the ontologies MA (2,737 concepts) and NCIT anatomy part (3,298 concepts) as available in June 2011. We use 28 input mappings interrelating MA/NCIT via 11 different intermediate ontologies. The input mappings are separated in two different sets. The first mapping set (referred to as Mapping set 1) is taken from the community platform BioPortal [28] and comprises 20 mappings from MA or NCIT to 10 ontologies including BRENDA Tissue Ontology (BTO), Cell Line Ontology (CL), Foundational Model of Anatomy (FMA), Galen (Galen), Logical Observation Identifiers Names and Codes (LOINC), Medical Subject Headings (MeSH), RadLex, Uber Anatomy Ontology (Uber), Teleost Anatomy (TAO), and ZebraFish Anatomy (ZF). These mappings have been created with the LOOM match approach [11]. LOOM takes all names and synonyms of the ontology concepts as input and returns concept pairs as matching when one of their name or synonym differ in at most one character. We use the mappings as provided by the BioPortal web page[1].

The second set of mappings (called Mapping set 2) consists of eight mappings interrelating MA and NCIT with four intermediate ontologies including Unified Medical Language System (UMLS), Uber, FMA, and Radlex. These mappings have been automatically created by a GOMMA match process. It uses a high trigram string similarity between concept name and synonyms to generate correspondences between concepts. Moreover, post-processing steps are applied to select only the best correspondence(s) per concept (MaxDelta selection (see [7])) and removal of crossing correspondences [15].

4.2 Route Effectiveness

We focus on routes involving a single intermediate ontology since there are many such routes. Typically, routes with chains of two or more intermediate ontologies may result in a reduced effectiveness. Table 1 shows selected statistics for the considered routes over different intermediates indicated in the columns. The routes are grouped by mapping set and ordered by the computed effectiveness (last row) starting with the route having the highest effectiveness. The first two rows characterize the input mappings for each route by showing the number of correspondences they comprise. These numbers are very different in both

[1] BioPortal: http://bioportal.bioontology.org, http://rest.bioontology.org

Routes via intermediate	Mapping set 1 Uber	FMA	CL	Galen	Radlex	MeSH	BTO	LOINC	ZF	TAO	Mapping set 2 UMLS	Uber	FMA	Radlex
$\|M_{MA,IS}\|$	1882	1176	1131	699	679	569	513	403	250	206	2975	2300	1601	1082
$\|M_{IS,NCIT}\|$	1330	1804	1851	1083	902	941	684	650	375	334	4214	1703	2337	1347
$\|\texttt{range}(M_{MA,IS}) \cap \texttt{domain}(M_{IS,NCIT})\|$	1048	825	793	564	479	438	372	364	169	140	2029	1320	1051	709
eff	0.35	0.27	0.26	0.19	0.16	0.15	0.12	0.12	0.06	0.05	0.67	0.44	0.35	0.23

Table 1. Mappings between MA and NCIT included in the evaluation according to the two used mapping sets

mapping sets ranging from approx. 1,900 (4,300) of the largest to about 200 (1,000) correspondences of the smallest mapping in Mapping set 1 (Mapping set 2). For the ontologies used in both mapping sets (Uber, FMA, and Radlex), the mappings in Mapping set 2 are larger than in Mapping set 1.

The third row displays the sizes of the mapping overlap in the intermediate ontology that is decisive for the effectiveness. In Mapping set 1, the route via Uber has the largest overlap (1,048 objects) and the highest effectiveness value of 0.35. In Mapping set 2, the number of referenced concepts in the intermediates is larger resulting in higher effectiveness values, but the relative order Uber, FMA, and Radlex remains. However, the route via UMLS has the highest effectiveness measure (0.67) and, is thus the most promising route for Mapping set 2.

4.3 Correlation of Routes Effectiveness and Composition Quality

Fig. 3 correlates the effectiveness (dashed line, z-axis on the right) for each route with the match quality of the composed mapping in terms of precision, recall and F-measure (bars, y-axis). The routes are decreasingly ordered by their effectiveness from left to right and separated for both mapping sets. Overall, there is an excellent correlation between the effectiveness values and achieved match quality for both mapping sets. This means that the composed correspondences are indeed valuable and contribute to the match result so that higher effectiveness values translate into higher F-measure values. For instance, for Mapping set 1 the route via Uber has the best effectiveness and the highest F-measure of 0.76 whereas the route via TAO with the lowest effectiveness (0.05) results in the worst F-measure of only 0.16. The same holds for Mapping set 2: the route via UMLS (Radlex) with the highest (lowest) effectiveness generates a mapping with the best (worst) F-measure of 0.87 (0.6). Therefore, using the effectiveness metric is a valid and reliable means to select the intermediate ontology providing the best match quality.

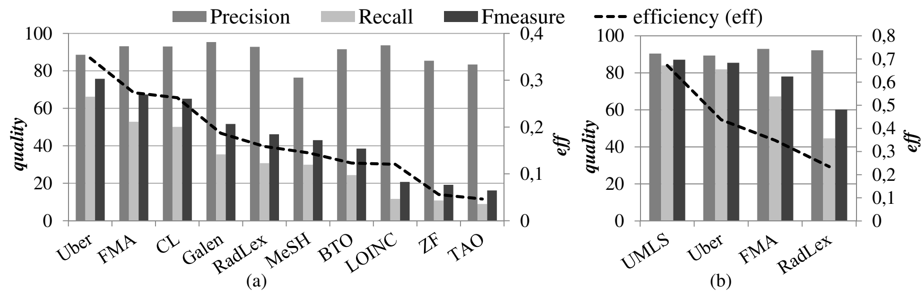


Fig. 3. Match quality for mapping compositions of routes with a single intermediate (sorted by effectiveness) for Mapping set 1 (a) and Mapping set 2 (b)

4.4 Top K Selection and Composition

In the next experiment, we evaluate whether the match quality (F-measure) can be increased when using the proposed selection strategies *topKByEffectiveness* and *topKByComplement* for selecting k routes and combining their composition results. We set k to 3 and use union as merge operation in both selection strategies. According to the effectiveness values for each route (see Table 1 and Algorithm 1) we select routes via Uber, FMA, and CL (UMLS, Uber, and FMA) in Mapping set 1 (Mapping set 2) for the *topKByEffectiveness* strategy and routes via Uber, FMA, and Galen (UMLS, Uber, and FMA) in Mapping set 1 (Mapping set 2) for the *topKByComplement* strategy. For comparison, we consider several additional selection strategies. They include the single route with the highest F-measure in the mapping set (BestSingle) and the strategies resulting in the worst (Min3), average (Avg3), and best (Max3) F-measure result for combining any three routes. Moreover, we computed the combination of all routes per mapping set (All).

Fig. 4 shows the F-measure for all selection strategies and both mapping sets. The results show that in both cases the *topKByComplement* strategy focusing on complementary mappings produces the max. possible match quality, i.e., it is able to identify the best and most effective composition routes. Interestingly, doing a compose-based match on only three out of the 10/4 possible routes results in better match quality than using all available routes since it apparently avoids wrong correspondences introduced by weaker routes. For instance, in Mapping set 1 F-measure is increased by 3% (74.2% → 77.4%) compared to the 'All' strategy. For Mapping set 2, the F-measure is improved by 0.2% compared to 'All'. The resulting F-measure of 91.5% is comparable to the best result in the OAEI 2011 contest (91.7% F-measure). While the OAEI contest poses certain restrictions, the participating prototypes did also exploit background knowledge for the Anatomy test case. Our *topKByEffectiveness* strategy shows marginally worse results compared to *topKByComplement* (76.2% vs. 77.4% for Mapping set 1), apparently since CL complements Uber and FMA less well than using Galen as intermediate ontology.

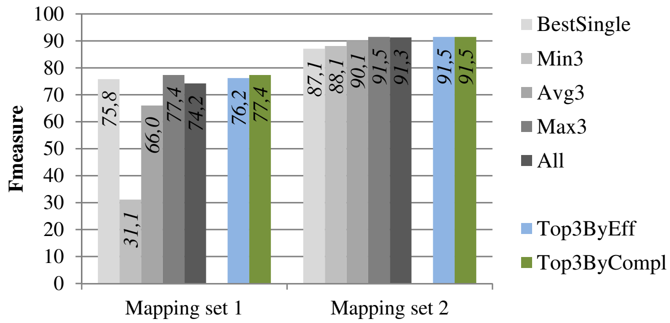


Fig. 4. Match results of combinations of multiple routes

5 Related Work

Ontology matching is the process of determining a set of semantic correspondences (ontology mapping) between concepts of two ontologies. A manual matching by domain experts is very time-consuming and for large ontologies almost infeasible. Thus, many (semi-)automatic matching algorithms have been developed for ontology matching (see [9,22,23] for surveys). Common match approaches follow a direct matching by employing lexical and structural methods; some approaches also consider the similarity of ontology instances. State-of-the art match systems such as COMA++ [2], Falcon [13] or SAMBO [17] combine multiple matchers within a match strategy to achieve better match quality. Results of matching biomedical ontologies showed that linguistic matching methods based on the similarity of concept names and synonyms produce very good results [30,11].

The composition of mappings has mainly been studied for schemas [8,10] and in model management [4]. Only a few approaches consider mapping composition for deriving new mappings in ontology matching. For instance, [29] utilizes FMA as an intermediate to indirectly generate a mapping between MA and NCIT. Similarly, the SAMBO system [17] utilizes background knowledge (e.g., UMLS) to find additional correspondences in the match process. [25] presents an empirical analysis of mapping composition available in BioPortal. In own related work [12], we already studied mapping composition. The primary focus of this work was on match quality (F-measure) by a manual intermediate selection but not on automatic strategies to select the best intermediates according to their expected contribution to the overall match quality.

In contrast to these approaches this paper differs in the following points. First, we apply mapping composition with multiple routes, while most match approaches only consider one route or purely apply a direct match. Second, we focus on finding the most valuable routes for mapping composition out of a pool of possible routes in two different mapping sets. A ranking of routes w.r.t. their effectiveness allows us to compose mappings for a reduced number of routes saving time and possibly improving match quality as shown in the evaluation.

6 Conclusion and Future Work

We proposed a new approach to rank and select promising routes for composing mappings between biomedical ontologies. The introduced effectiveness measure can be easily computed and allows a reliable identification of the most promising intermediate ontologies for composition-based ontology matching. We further proposed the selection of the k top routes and the combination of their composition results for improved match quality. Our evaluation for an OAEI match task on large anatomy ontologies showed the effectiveness of the proposed approach. In particular we found that the effectiveness metric for different routes correlates excellently with their achievable F-measure quality. Furthermore, we found that the *topKByComplement* ranking strategy is most effective that combines the route with the best effectiveness with routes providing most complementary correspondences. Our approach could effectively exploit existing mappings and achieved an excellent 91.5% F-measure for the challenging OAEI anatomy task. This shows that mapping composition is not only an efficient method to derive new mappings but can also increase the match quality, e.g., by finding additional correspondences compared to a direct match approach.

In future work we plan to apply and extend the approach for other domains, ontologies and data sources, e.g., matching Linked Data sources. In particular, we want to investigate inter-linking of instance objects and to consider further correspondence types. We further like to study longer mapping chains consisting of multiple intermediates.

Acknowledgment

This work is supported by the German Research Foundation (DFG), grant RA 497/18-1 ("Evolution of Ontologies and Mappings").

References

1. Adult Mouse Anatomy: `http://www.informatics.jax.org/searches/AMA_form`
2. Aumueller, D., Do, H.H., Massmann, S., Rahm, E.: Schema and ontology matching with COMA++. In: Proc. SIGMOD (2005)
3. Belleau, F., et al.: Bio2rdf: Towards a mashup to build bioinformatics knowledge systems. Journal of biomedical informatics 41(5), 706–716 (2008)
4. Bernstein, P., Melnik, S.: Model management 2.0: manipulating richer mappings. In: Proc. of SIGMOD (2007)
5. Bodenreider, O., Stevens, R.: Bio-ontologies: current trends and future directions. Briefings in bioinformatics 7(3), 256–274 (2006)
6. Bodenreider, O., et al.: Of mice and men: Aligning mouse and human anatomies. In: Proc. of AMIA Annual Symposium (2005)
7. Do, H., Rahm, E.: Matching large schemas: Approaches and evaluation. Information Systems 32(6), 857–885 (2007)
8. Dragut, E.C., Lawrence, R.: Composing mappings between schemas using a reference ontology. In: Proc. of CoopIS/DOA/ODBASE. pp. 783–800 (2004)

9. Euzenat, J., Shvaiko, P.: Ontology matching. Springer-Verlag New York (2007)
10. Fagin, R., Kolaitis, P., Popa, L., Tan, W.: Composing schema mappings: Second-order dependencies to the rescue. ACM Transactions on Database Systems (TODS) 30(4), 994–1055 (2005)
11. Ghazvinian, A., Noy, N., Musen, M.: Creating mappings for ontologies in biomedicine: Simple methods work. In: Proc. of AMIA Annual Symposium (2009)
12. Gross, A., Hartung, M., Kirsten, T., Rahm, E.: Mapping composition for matching large life science ontologies. In: 2nd Intl. Conference on Biomedical Ontology (ICBO) (2011)
13. Hu, W., Qu, Y.: Falcon-AO: A practical ontology matching system. Web Semantics: Science, Services and Agents on the World Wide Web 6(3), 237–239 (2008)
14. Jakoniene, V., Lambrix, P.: Ontology-based integration for bioinformatics. In: VLDB Workshop on Ontologies-based techniques for DataBases and Information Systems-ODBIS 2005. pp. 55–58 (2005)
15. Jean-Mary, Y., Shironoshita, E., Kabuka, M.: Ontology matching with semantic verification. Web Semantics: Science, Services and Agents on the World Wide Web 7(3), 235–251 (2009)
16. Kirsten, T., Gross, A., Hartung, M., Rahm, E.: GOMMA: A Component-based Infrastructure for managing and analyzing Life Science Ontologies and their Evolution. Journal of Biomedical Semantics 2, 6 (2011)
17. Lambrix, P., Tan, H.: Sambo–a system for aligning and merging biomedical ontologies. Web Semantics: Science, Services and Agents on the World Wide Web 4(3), 196–206 (2006)
18. Lambrix, P., Tan, H., Jakoniene, V., Strömbäck, L.: Biological ontologies. In: Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences, pp. 85–99. Springer Verlag (2007)
19. NCI Thesaurus: `http://ncit.nci.nih.gov/`
20. Noy, N., et al.: Bioportal: ontologies and integrated data resources at the click of a mouse. Nucleic acids res. 37(suppl 2), W170–W173 (2009)
21. Ontology Alignment Evaluation Initiative: `http://oaei.ontologymatching.org/`
22. Rahm, E.: Towards Large Scale Schema and Ontology Matching. In: Schema Matching and Mapping, chap. 1, pp. 3–27. Springer (2011)
23. Rahm, E., Bernstein, P.A.: A survey of approaches to automatic schema matching. The VLDB Journal 10(4), 334–350 (2001)
24. Smith, B., et al.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature biotechnology 25(11), 1251–1255 (2007)
25. Tordai, A., et al.: Lost in translation? empirical analysis of mapping compositions for large ontologies. In: Proc. Ontology Matching Workshop (2010)
26. UBERON: `http://obofoundry.org/wiki/index.php/UBERON:Main_Page`
27. Unified Medical Language System: `http://www.nlm.nih.gov/research/umls`
28. Whetzel, P., et al.: Bioportal: enhanced functionality via new web services from the national center for biomedical ontology to access and use ontologies in software applications. Nucleic Acids Research 39(suppl 2), W541 (2011)
29. Zhang, S., Bodenreider, O.: Alignment of multiple ontologies of anatomy: Deriving indirect mappings from direct mappings to a reference. In: AMIA Annual Symposium Proceedings (2005)
30. Zhang, S., Bodenreider, O.: Experience in aligning anatomical ontologies. International journal on Semantic Web and information systems 3(2), 1–26 (2007)