An Evolution-based Approach for Assessing Ontology Mappings - A Case Study in the Life Sciences

Andreas Thor[1], Michael Hartung[2], Anika Gross[2], Toralf Kirsten[2,3], Erhard Rahm[1,2]

[1]Dept. of Computer Science, University of Leipzig
[2]Interdisciplinary Centre for Bioinformatics, University of Leipzig
[3]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig
{thor, rahm}@informatik.uni-leipzig.de
{hartung, gross, tkirsten}@izbi.uni-leipzig.de

Abstract: Ontology matching has been widely studied. However, the resulting ontology mappings can be rather unstable when the participating ontologies or utilized secondary sources (e.g., instance sources, thesauri) evolve. We propose an evolution-based approach for assessing ontology mappings by annotating their correspondences by information about similarity values for past ontology versions. These annotations allow us to assess the stability of correspondences over time and they can thus be used to determine better and more robust ontology mappings. The approach is generic in that it can be applied independently from the utilized match technique. We define different stability measures and show results of a first evaluation for the life science domain.

1 Introduction

Ontology matching aims at identifying semantic correspondences between concepts of different ontologies. It has been recently studied in diverse scientific and commercial application domains and various match approaches and prototypes (see e.g., [RB01, KS03, Do06, ES07] for surveys) have been developed. The ontology mappings determined by ontology matching are used in many ways, such as for enhanced data analysis or data integration. However, ontologies underlie continuous modifications because domain knowledge or community agreements represented by ontologies often change over time. Typical changes include additions of new elements, deletions of outdated elements, modifications of element attributes, such as names and descriptions, and structural changes, such as element moves within an is-a hierarchy. Therefore new ontology versions are released periodically and, thus, ontology mappings need to be re-created or updated accordingly. This is also necessary when new versions of secondary sources used for matching become available, such as instance sources, thesauri and their associations to ontologies.

As a result of evolving ontologies and secondary sources the generated ontology mappings can be rather unstable, i.e., the similarity between two given concepts may vary significantly over several versions. As an example from the life science domain, Figure 1 shows the history of concept similarity for two selected correspondences over 21 mapping versions. The considered mapping relates two sub-ontologies of the Gene Ontology
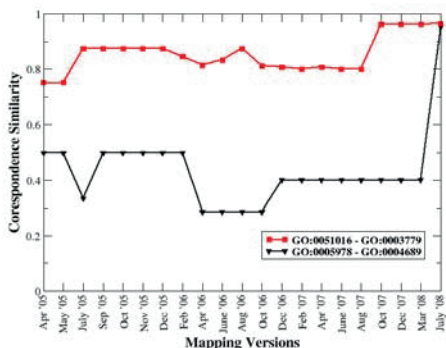
Figure 1: History of similarity values for Gene Ontology correspondences

[Go04], Molecular Functions and Biological Processes, and was determined by an instance-based match technique which derives the similarity of two concepts from their associations to the same instances (see [KTR07] for details). Both correspondences have a relatively high similarity value of 0.95 in the last mapping version but significantly differ in their history. The first correspondence (GO:0051016-GO:0003779) is more stable and at a higher level of similarity than the second correspondence (GO:0005978-GO:0004689). Since the similarity value of the latter correspondence jumps from 0.4 to 0.95 its correctness might be questionable and should be further verified. This observation of fluctuating concept similarities is mainly influenced by evolving ontologies, modifications on secondary sources as well as changes of associations between them and was quantitatively evaluated in [HKR08].

In order to obtain ontology mappings of high quality, previous work (e.g., [ADMR05]) has shown that a combination of different match approaches can be helpful. However, current ontology match approaches do not take into account that ontologies evolve. They only utilize information about the specific ontology versions which are matched with each other. Studying the correspondence history for past ontology versions is orthogonal to a match technique or matcher combination. One aim is to utilize the stability of correspondences to identify robust mappings, while unstable correspondences may be subject to a focused manual validation. To that end we make the following contributions:

· We propose a generic approach to annotate generated ontology mappings independently from the used match approach, i.e., the computation of the similarity values. Besides the similarity value, the mapping annotation shows the stability for each correspondence of the mapping by taking its historic changes into account. Hence, the correspondences can not only be ranked and evaluated using the similarity value but also by considering the computed stability values.

· We define two stability measures quantifying the evolution of similarity values for a given concept correspondence. While the average stability considers changes during evolution steps, the weighted maximum stability inspects stability w.r.t. the current similarity value.

· We show results of a first evaluation using the proposed approach in an example domain, the life sciences. In particular we use the stability measures to classify the correspondences of a mapping into several groups, such as accepted, candidates, and questionable.

The rest of the paper is organized as follows. In Section 2, we introduce the approach and define the stability measures. Section 3 shows the evaluation results for life science applications. We discuss related work in Section 4 before we conclude in Section 5.

2 Evolution-aware ontology mappings

In this section we first introduce a versioning scheme that is used for the definition of version-based correspondence similarity. We then motivate and define two stability measures that characterize the correspondence similarity over several versions.

2.1 Versioning scheme and correspondence similarity

Since different ontologies as well as secondary data are typically developed independently, new versions of different ontologies A and B are usually not introduced at the same time. Figure 2 schematically illustrates the evolution process for two ontologies A and B and a secondary data source D over time. Each diamond represents a new ontology version and each cylinder indicates changes in the (instance) data of D.
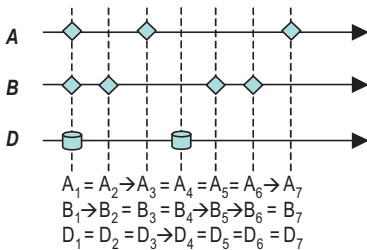

$A_1 = A_2 \rightarrow A_3 = A_4 = A_5 = A_6 \rightarrow A_7$
$B_1 \rightarrow B_2 = B_3 = B_4 \rightarrow B_5 \rightarrow B_6 = B_7$
$D_1 = D_2 = D_3 \rightarrow D_4 = D_5 = D_6 = D_7$

Figure 2: Schematic illustration of the versioning scheme

We start with two initial versions A_1 and B_1. After a while a new version B_2 comes up whereas A_1 is still the current valid version for A. We bridge this gap by introducing a *virtual version* for A (A_2) that is, of course, equivalent to A_1. By doing so we may only consider matching ontologies of the same virtual version, e.g., A_2 and B_2. Note from Figure 2 that new versions for A and B need not be introduced alternately but an ontology version (A_3) may outlast several versions of the other ontology (B_5 and B_6). Since changes in the secondary data source D may also affect the results of ontology matching, we additionally introduce new virtual versions when such changes occur (see A_4 and B_4). In summary, during an evolution step $i \rightarrow i+1$ from a virtual version i to $i+1$ at least one of the participating ontologies (A or B) or a secondary data source (D) used for matching has changed.

In the following, we consider concepts a and b from the ontologies A_i and B_i (of a virtual version i), respectively, and a match technique m. We do not make any assumption about the matcher to keep our approach generally applicable. For the similarity computation a matcher may not only use the concepts a and b itself, e.g., the concept names, but may also utilize other parts of the involved ontologies A_i and B_i, e.g., their structure or associated instances of a source D_i.

We denote the similarity between a and b as $sim(a,b,m \mid A_i,B_i)$ with $a \in A_i$ and $b \in B_i$, i.e., the concept similarity determined by a matcher m between a and b is defined in consideration of the ontologies in their versions A_i and B_i. Since ontology evolution may be triggered by concept additions and deletions, it is not given that a concept a is always contained in all versions A_i ($i=1...n$). Thus, we further define $sim_i(a,b,m)$ as the *correspondence similarity* between two concepts a and b w.r.t. a version i.

$$sim_i(a,b,m) := \begin{cases} sim(a,b,m \mid A_i,B_i), & \text{if } a \in A_i \wedge b \in B_i \\ 0 & , \text{otherwise} \end{cases} \in [0,1]$$

If (at least) one of the two concepts does not appear in the considered versions the correspondence similarity $sim_i(a,b,m)$ is set to the minimal value of 0. This definition allows for an easy similarity comparison between two concepts over different virtual versions.

2.2 Definition of stability measures

We aggregate the different similarity values of a given correspondence by calculating stability measures which consider the similarity value of a current version n as well as the $k>0$ previous versions. However, the maximum number $kmax$ of available previous versions is limited by both the correspondence (a,b) and the applied matcher m. Obviously we may only consider versions from the time when both concepts a and b have been appeared together in the involved ontology versions for the first time. Moreover, we further restrict $kmax$ to the first version with $sim_i(a,b,m)>0$, i.e., we determine the first version where matcher m calculates a positive similarity value for the correspondence (a,b). Thereby the "initial jump" from 0 to a positive similarity value is not considered for any stability calculation because we do not want to penalize this as instability. Hence, $kmax$ which will be used in later stability definitions is defined as follows:

$$kmax_n(a,b,m) = \max_{k=1..n-1} \left(\left\{ k \mid sim_{n-k}(a,b,m) > 0 \right\} \right)$$

Note that this definition is only well-defined if there is at least one correspondence with $sim_i(a,b,m)>0$ within the previous k versions. However, this is not a relevant restriction because correspondences with $sim_i(a,b,m)=0$ for all $i<n$ do not carry evolutionary information and, thus, can not be annotated with any stability values.

In the following, we define two stability measures *average* and *weighted maximum*. The first stability captures the average fluctuation of the similarity values over the last k evolution steps. For a correspondence (a,b) the *average stability* is defined as follows:

$$stabAvg_{n,k}(a,b,m) = \begin{cases} 1 - \dfrac{1}{k} \cdot \sum_{i=n-k}^{n-1} \left| sim_{i+1}(a,b,m) - sim_i(a,b,m) \right| & \text{, if } k \leq kmax_n(a,b,m) \\ stabAvg_{n,kmax_n(a,b,m)}(a,b,m) & \text{, otherwise} \end{cases} \in [0,1]$$

The measure captures both small as well as big differences of the correspondence similarity (as determined by m) as indicators for an unstable correspondence (a,b). Hence, we interpret (a,b) as stable if only few and small similarity changes occur during the evolution process. To that end, the absolute differences of correspondence similarities of consecutive ontology versions are summarized for all evolution steps from version $n-k$ to n. This sum is normalized by the total number of evolution steps (k) giving a value in a 0 to 1 range because each of the k evolution steps contributes a similarity change between 0 and 1. Finally, we compute 1 minus this normalized sum to have an average stability value of 1 (0) for perfect stability (complete instability).

For a correspondence (a,b), a current version n, and a matcher m we further define the *weighted maximum stability* for the last k evolution steps as follows:

$$stabWM_{n,k}(a,b,m) = \begin{cases} 1 - \max_{i=1...k}\left[\dfrac{|sim_n(a,b,m) - sim_{n-i}(a,b,m)|}{i}\right] & \text{, if } k \leq kmax_n(a,b,m) \\ stabWM_{n,kmax_n(a,b,m)}(a,b,m) & \text{, otherwise} \end{cases} \in [0,1]$$

The idea of weighted maximum stability for a correspondence (a,b) is to identify how close similarity values are to the current similarity $sim_n(a,b,m)$, i.e., we focus on version n to assess the stability within the last k evolution steps. For a version $n-i$ we therefore consider the distance of $sim_{n-i}(a,b)$ to the current similarity $sim_n(a,b)$. This distance is normalized (weighted) by the number of evolution steps (i) and, thus, differences in later versions have a higher impact than in previous versions. This allows us to inspect trends, i.e., to evaluate if the evolution of a correspondence similarity is constant, slightly increasing (decreasing) or gamboling within the last versions. The maximum value over all considered versions i defines the maximum deviation and indicates the stability. Again, we compute 1 minus this max-value so that complete stability (instability) is equal to 1 (0). If and only if all previous similarities sim_i of the considered correspondence (a,b) are equal to the current similarity sim_n, we calculate a perfect stability with $stabWM=1$. Contrary, the minimal weighted maximum stability of 0, i.e., complete instability, is obtained if and only if $|sim_n - sim_{n-1}| = 1$, i.e., the correspondence similarity has changed from 0 to 1 (or vice versa) during the last evolution step $n\text{-}1 \rightarrow n$.

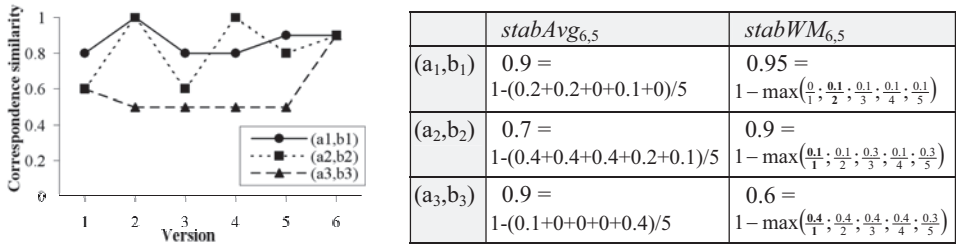|	$stabAvg_{6,5}$	$stabWM_{6,5}$
(a_1,b_1)	$0.9 =$ $1-(0.2+0.2+0+0.1+0)/5$	$0.95 =$ $1-\max\left(\frac{0}{1};\frac{0.1}{2};\frac{0.1}{3};\frac{0.1}{4};\frac{0.1}{5}\right)$
(a_2,b_2)	$0.7 =$ $1-(0.4+0.4+0.4+0.2+0.1)/5$	$0.9 =$ $1-\max\left(\frac{0.1}{1};\frac{0.1}{2};\frac{0.3}{3};\frac{0.1}{4};\frac{0.3}{5}\right)$
(a_3,b_3)	$0.9 =$ $1-(0.1+0+0+0+0.4)/5$	$0.6 =$ $1-\max\left(\frac{0.4}{1};\frac{0.4}{2};\frac{0.4}{3};\frac{0.4}{4};\frac{0.3}{5}\right)$

Figure 3: Computation of stability values for three example correspondences

Example: Figure 3 (left) shows an example for the evolution of similarity values for three correspondences. All correspondences have a similarity value of 0.9 in the current version ($n=6$) but show a different behavior in the previous $k=5$ versions. Figure 3 (right) also illustrates the computation of both stability measures, i.e., $stabAvg_{6,5}$ and $stabWM_{6,5}$. The first correspondence is very stable over all considered versions and, thus, achieves high stability values. By contrast, the second correspondence shows high fluctuations which diminish in the last versions. Thereby the average stability is rather low but the correspondence obtains a high weighted maximum stability. The similarity values of the third correspondence are in a very similar range for the first five versions but we observe a steep increase from version 5 to 6. This behavior is reflected by a high average stability together with a low weighted maximum stability because the change in the last evolution step primarily accounts for the computation of the weighted maximum stability.

The given example illustrates that the proposed stability measures quantify different aspects of the stability over a period of versions. We will further demonstrate the usefulness of our stability measures in the following evaluation.

3 Evaluation

In this section we show the results of a first evaluation of our approach in order to demonstrate its benefit and practicability for assessing (available) ontology mappings. For this initial work we utilize the life sciences as an example domain. In this domain a huge amount of (meta-)data is generated and the very active research community often revises data due to new (experimental) findings. Among others this process regularly triggers new versions of ontologies and instance data sources that use ontology concepts for semantic annotation of their molecular-biological objects, e.g., genes or proteins.

For this evaluation we consider the two sub-ontologies Biological Processes (BP) and Molecular Functions (MF) from the popular Gene Ontology (GO) [Go04]. These two ontologies consist of 15,131 and 8,827 concepts, respectively (as of Apr. 2008). Furthermore, the evaluation takes into account associated instance data of Ensembl [HAB+06], a data source containing proteins annotated with ontology concepts of the GO. Particularly, Ensembl of July 2008 comprises 46,704 proteins which exhibit 80,705 (100,195) annotations to BP (MF). Referring to our versioning scheme we consider 26 different versions (from Feb. 2004 to July 2008) due to changes in the ontologies and modifications in the associated instance data.

3.1 Quantitative statistics

We utilized an instance-based match approach that we have already applied in the life sciences [KTR07] and in e-commerce [TKR07]. The similarity between two concepts is thereby derived from the overlap of the instances that are associated to each of the two concepts. More precisely, we used the similarity measure sim_{min-3} whereupon each correspondence must have an overlap of at least three instances. Overall, our instance-based match approach computed 3,280 match correspondences between MF and BP for the latest version (26). All correspondences exhibit a minimum similarity value ($sim_{26}(a,b,min-3)$) of 0.8.

In a first analysis we inspect the presence of match correspondences in different versions, i.e., we determine how many correspondences of version 26 are also present in previous versions (1-25). Figure 4 illustrates the cumulative frequency of correspondences w.r.t. the minimal version number representing the first occurrence of a correspondence. We observe an almost linear slope from version 1 to 22 (see Figure 4: linear fit) followed by a considerably increasing number of correspondences that are only present in the last five versions. Approx. 76% of all correspondences (2,497) have at least five versions, i.e., they exist since version 22 or earlier. Since the significance of stability information is limited for correspondences appearing in only a few versions, we consider these 2,497 correspondences in the following investigation.

We utilize our stability measures in the following way. On the one hand, we set $k=25$ (i.e., the complete history) for average stability ($stabAvg_{26,25}$) to reflect long-term stability of a correspondence. Note that this does not imply that all correspondences must appear in all previous 25 versions. The stability measure definitions make sure that
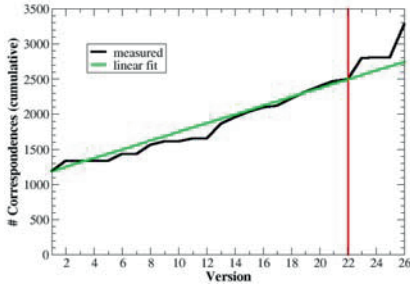
Figure 4: Cumulative frequency of corre-
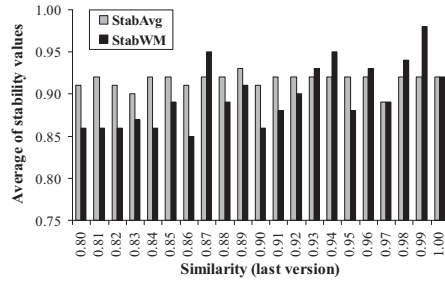spondences (w.r.t. their occurrence) in a
version

Figure 5: Average stability grouped by correspon-
dences with a certain similarity in the last version
(rounded to 2 decimals)

$stabAvg_{26,25}$ is well-defined for all 2,497 correspondences due to the use of $kmax$. On the
other hand, k=4 is used for weighted maximum stability ($stabWM_{26,4}$) to assess a corre-
spondence short-term stability, e.g., trends in the recent history.

A further evaluation scenario deals with the statistical independence of the similarity
measure compared to our stability measures. We aim at establishing new measures
which provide additional benefit to annotate ontology match results. For this purpose, we
distinguish groups of similarities (each of size 0.01) between 0.8 and 1 for the latest
version and calculate the average $stabAvg_{26,25}$ and $stabWM_{26,4}$ stability in each group (see
Figure 5). The values for $stabAvg_{26,25}$ ($stabWM_{26,4}$) range from 0.89 to 0.93 (0.85 to 0.98)
having a mean value of 0.92 (0.9) without definite order or trend. Thus, we observed that
the stability measures are statistically independent from the similarity computed by the
match technique. This indicates that the proposed stability measures may be very helpful
for classification of match correspondences.

3.2 Classification of match correspondences

The stability measures may be used to assess computed match correspondences and to
classify them more reliably as match or non-match compared to the sole use of the simi-
larity measure. Note that the following evaluation exemplifies a possible procedure for
supporting a manual match decision based on separating correspondences into several
groups of quality. We leave the use of automatic matching rules for future work.

We use an upper (t_{high}) and a lower (t_{low}) threshold for each considered measure
($stabAvg_{26,25}$, $stabWM_{26,4}$, sim_{26}). Correspondences with a value higher than t_{high} are con-
sidered as the best correspondences w.r.t. the stability criterion, whereas others between
t_{high} and t_{low} or lower than t_{low} are considered as intermediate or insufficient, respectively.

In order to set the thresholds for the stability measures we firstly analyze both,
$stabAvg_{26,25}$ and $stabWM_{26,4}$. Figure 6 reveals the cumulative frequency of correspon-
dences w.r.t. the two stability measures in the range of 0.6 to 1. A comparatively small
number of correspondences possesses stability values of about 0.85 or lower in both
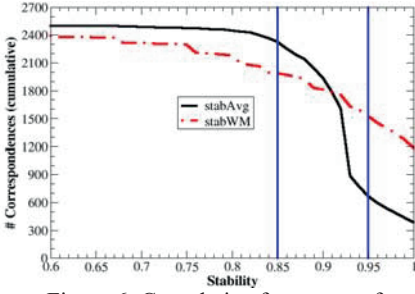measures. Particularly, about 6% (20%) of all correspondences show values lower than

Figure 6: Cumulative frequency of
stabAvg₂₆,₂₅ and *stabWM₂₆,₄*.

sim$_{26}$ > 0.9 / sim$_{26}$ ≤ 0.9	stabWM > 0.95	0.95 ≥ stabWM ≥ 0.85	0.85 > stabWM	Σ
stabAvg > 0.95	424 / 44	37 / 55	11 / 25	596
0.95 ≥ stabAvg ≥ 0.85	863 / 96	203 / 212	235 / 125	1734
0.85 > stabAvg	17 / 5	13 / 16	85 / 31	167
Σ	1449	536	512	2497

Table 1: Number of correspondences classified
by *stabAvg₂₆,₂₅*, *stabWM₂₆,₄* and *sim₂₆*.

0.85 in *stabAvg₂₆,₂₅* (*stabWM₂₆,₄*). This is supported by the fact that all cumulative frequencies are nearly stable in the range of 0.6 to 0.85. However, the behavior for *stabAvg₂₆,₂₅* and *stabWM₂₆,₄* differs for stability values higher than 0.85. While the curve for *stabAvg₂₆,₂₅* highly decreases between 0.85 and 0.95 (from 2,330 to 669), the cumulative frequency of *stabWM₂₆,₄* slightly decreases (from 1,985 to 1,528). Perfect stability for *stabAvg₂₆,₂₅* (*stabWM₂₆,₄*) is achieved by 377 (1,178) correspondences. As a result, we set a threshold of 0.85 for both $t_{stabAvg,low}$ and $t_{stabWM,low}$ and thresholds for $t_{stabAvg,high}$ and $t_{stabWM,high}$ are set to 0.95 (indicated as lines in Figure 6). Furthermore, based on experiences from former match tasks utilizing sim_{min-3} we set the higher threshold for sim_{26} ($t_{sim,high}$) to 0.9. Note that we implicitly adapted $t_{sim,low}$ to 0.8 before (see Section 3.1).

These threshold settings are specific to our match scenario and can differ for other match tasks since other match data, match techniques and application purposes need to be considered. For instance, the frequency of ontology modifications depends on the domain and therefore may influence stability thresholds. Moreover, the number of previous versions (k) that are respected for stability measuring may depend on the extent of ontology changes, e.g., only versions after a major ontology release might be considered.

To exemplify our approach we discuss one result table using the described threshold configuration. Table 1 represents a classification of the correspondences w.r.t. the two stability measures and the last similarity value. In general, the reliability of the sets of correspondences decreases from top left to down right. For a better understanding, we distinguish the results into three groups: I (white), II (light-gray) and III (dark-gray). The correspondences in group I represent 54.8% (1,368) of all correspondences. They possess the best quality as they exhibit high values for at least two criteria and no insufficient criterion. To our understanding they are classified as *accepted*. Group II covers 15.3% (382) and includes correspondences that mainly achieve intermediate and high values in each measure. As these correspondences still obtain quite high stability and similarity values we ranked them as *candidates*. Finally, group III contains a total of 29.9% (747). These correspondences predominantly possess insufficient values for at least one of the stability criteria and, thus, are considered as *questionable*.

In summary, these first evaluation results show that the proposed stability measures can be used to more precisely classify correspondences of an ontology mapping as match or non-match. We thereby utilized upper and lower thresholds (see Table 1) in this study.

4 Related Work

Overviews of approaches for ontology matching in general are given in [RB01, ES07]. Match techniques can be roughly categorized into the following groups: metadata-based, instance-based, or hybrid. More precisely, match techniques consider available information of ontologies (e.g., concepts names or descriptions, structure) or utilize associated data (e.g., instances that are annotated with ontology concepts) to compute match correspondences between different ontologies. However, none of these techniques incorporates historical information about match correspondences. Hence, our work complements previous match techniques to enhance their match results by considering evolutionary information of correspondences to produce more stable mappings.

The evolution of ontologies was primarily studied in the context of the Semantic Web. Versioning of ontologies was investigated by Klein [KF01,Kle04]. Furthermore, [NK04] defined change operations to describe the evolution between ontology versions. Supplementary, [SMM+02, SM02] formulized the process of ontology evolution by proposing strategies to unambiguously handle critical ontology changes. Whereas these approaches focus on consistent evolution and versioning of ontologies, we look at the evolution of correspondence similarities to produce more stable match results, i.e., we neither do change ontologies themselves nor migrate dependent instances if ontologies evolve.

The combination of both matching and evolution is investigated within the model management framework of Bernstein et al [BM07]. The schema evolution problem described in [Ber03] utilizes operators such as match, compose and diff to adapt views that would break if a schema evolves to a new version. The approaches proposed in [VMP03, YP05] can be seen as a part of this framework. Whereas [VMP03] presented a framework for incremental adaptation of schema mappings as participating schemas evolve, [YP05] studied the mapping adaptation by employing mapping composition and pruning techniques. These approaches differ from ours in that they reuse match results to efficiently compute new mappings (e.g., for views) if schemas change.

The evolution of association rules has been studied in data mining. Particularly, [AG95] proposed to monitor rules in different time periods by considering changes in their support and confidence. An advanced approach presented in [LHM01] aims to find fundamental association rule changes. Similar to mapping correspondences, association rules also relate (sets of) items to describe semantic relationships. However, data mining approaches attempt to detect rules that significantly differ in their evolution from others whereas we focus on finding the most stable correspondences in an ontology mapping.

5 Conclusions and future work

We presented an evolution-based approach for assessing ontology mappings. The approach utilizes two stability measures taking the history of correspondences w.r.t. their computed similarity values into account. While the average stability considers changes between all consecutive ontology versions in the analyzed evolution history, the weighted maximum stability inspects stability w.r.t. the computed similarity value of the

current mapping. In addition to the similarity values, the proposed stability values can be used to rank and evaluate the correspondences of an ontology mapping. Moreover, the stability values are useful to classify correspondences into several groups such as *accepted*, *candidate* and *questionable*. The approach is complementary to existing match approaches that do not consider historical information of correspondences. Furthermore, it is generic in that it can be applied independently from the utilized match technique. The defined stability measures offer the possibility for an individual (application-specific) adjustment. The results of a first evaluation using ontologies of the life science domain show the usability of our approach to classify mapping correspondences.

In future work, we will comparatively analyze and hence prove the applicability of our approach in different domains, e.g., product catalogs in e-commerce or matching web directories. Moreover, we will establish alternative stability measures by taking further knowledge into account, e.g., the structure of ontologies that participate in an ontology mapping. Finally, the management of versioned ontology mappings and the efficient computation of stability measures will be investigated in more detail.

References

[AP95]   Agrawal; Psaila: Active Data Mining. KDD, 1995
[ADMR05] Aumüller; Do; Massmann; Rahm: Schema and ontology matching with COMA++. SIGMOD, 2005
[Ber03]   Bernstein: Applying Model Management to Classical Meta Data Problems. CIDR, 2003
[BM07]   Bernstein; Melnik: Model Management 2.0-Manipulating Richer Mappings. SIGMOD, 2007
[Do06]   Do: Schema Matching and Mapping-based Data Integration. PhD thesis, 2006
[ES07]   Euzenat; Shvaiko: Ontology matching. Springer Verlag, 2007
[Go04]   The Gene Ontology Consortium: The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research 32, 2004
[HAB+06] Hubbard; Aken; Beal et al.: Ensembl 2007. Nucleic Acids Research 35, 2006
[HKR08]   Hartung; Kirsten; Rahm: Analyzing the Evolution of Life Science Ontologies and Mappings. DILS, 2008
[Kle04]   Klein: Change Management for Distributed Ontologies. PhD thesis, 2004
[KF01]   Klein; Fensel: Ontology versioning on the Semantic Web. Int. Semantic Web Working Symposium (SWWS), 2001
[KS03]   Kalfoglou; Schorlemmer: Ontology mapping: The state of the art. The Knowledge Engineering Review Journal 18(1), 2003
[KTR07]   Kirsten; Thor; Rahm: Instance-based matching of large life science ontologies. DILS, 2007
[LHM01]   Liu; Hsu; Ma: Discovering the Set of Fundamental Rule Changes. KDD, 2001
[NK04]   Noy; Klein: Ontology evolution: Not the same as schema evolution. Knowledge and Information Systems 6(4), 2004
[RB01]   Rahm; Bernstein: A survey of approaches to automatic schema matching. VLDB Journal 10(4), 2001
[SM02]   Stojanovic; Motik: Ontology evolution within ontology editors. OntoWeb-SIG3 Workshop, 2002
[SMM+02] Stojanovic; Maedche; Motik et al.: User-driven ontology evolution management. Conf. On Knowledge Engineering and Knowledge Management, 2002
[TKR07]   Thor; Kirsten; Rahm: Instance-based matching of hierarchical ontologies. BTW, 2007
[VMP03]   Velegrakis; Miller; Popa: Mapping Adaptation under Evolving Schemas. VLDB, 2003
[YP05]   Yu; Popa: Semantic Adaptation of Schema Mappings when Schemas Evolve. VLDB, 2005