J Data Semant (2012) 1:219–236
DOI 10.1007/s13740-012-0011-z

ORIGINAL ARTICLE

Instance-Based Ontology Matching by Instance Enrichment

Balthasar Schopman · Shenghui Wang · Antoine Isaac · Stefan Schlobach

Received: 30 September 2011 / Revised: 15 June 2012 / Accepted: 22 June 2012 / Published online: 31 July 2012

Abstract The ontology matching (OM) problem is an important barrier to achieve true Semantic Interoperability. Instance-based ontology matching (IBOM) uses the extension of concepts, the instances directly associated with a concept, to determine whether a pair of concepts is related or not. While IBOM has many strengths it requires instances that are associated with concepts of both ontologies, (i.e) dually annotated instances. In practice, however, instances are often associated with concepts of a single ontology only, rendering IBOM rarely applicable. In this paper we discuss a method that enables IBOM to be used on two disjoint datasets, thus making it far more generically applicable. This is achieved by enriching instances of each dataset with the conceptual annotations of the most similar instances from the other dataset, creating artificially dually annotated instances. We call this technique *instance-based ontology matching by instance enrichment* (IBOMbIE). We have applied the IBOMbIE algorithm in a real-life use-case where large datasets are used to match the ontologies of European libraries. Existing gold standards and dually annotated instances are used to test the impact and significance of several design choices of the IBOMbIE algorithm. Finally, we compare the IBOMbIE algorithm to other ontology matching algorithms.

Keywords Ontology matching · Semantic Web · Semantic interoperability

1 Introduction

1.1 Motivation

Over the past decade the progress in Information and Communication Technology has made an immense quantity of information available. As the amount of information and the number of sources grow, the need for enhanced accessibility and interoperable data representation increases. The *Web of Data*, or the Semantic Web—a recently growing network that connects data resources (as opposed to the World Wide Web which links documents)—uses standards that enable uniform data representation to improve semantic interoperability. By means of formal languages such as *RDF*(*S*) and *OWL*, *ontologies* can be specified, sometimes for generic knowledge, most often, however, for specific application domains. Other specific models like SKOS can be used to represent less formal Knowledge Organization Systems (KOS), such as thesauri or subject heading lists.[1] In an open environment such as the Web, different parties tend to use their own concept definitions when publishing data, i.e., use their own ontologies. In order to achieve full interoperability on the Web of Data these different ontologies need to be matched.

1.2 Instance-Based (or Extensional) Ontology Matching

IBOM aligns ontologies using the extension of concepts, (i.e) their *instances*; the set of objects associated with (or: annotated by) that concept. The intuitive principle is that when a

B. Schopman (✉) · S. Wang · A. Isaac · S. Schlobach
Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
e-mail: bschopman@gmail.com

[1] This terminology corresponds to the broad view taken by the Ontology Mapping community (as can be witnessed, e.g., in the OAEI test-cases over the years and the OM literature in general).

pair of concepts is associated with the same set of objects, they are likely to be similar.[2]

With respect to lexical and structural algorithms an advantage of IBOM is that it is not negatively affected by ambiguous linguistic phenomena, such as synonyms and homonyms. This is an inherent advantage as matches are generated based on the actual usage of the concepts, as opposed to using their lexical metadata. A disadvantage is that to apply IBOM, a sufficient number of *dually annotated instances* is required, (i.e) instances that are associated with the two ontologies that we aim to align. In practice, dually annotated instances are rarely available since it requires extra effort to annotate instances using two different ontologies. This inherent problem of instance-based ontology matching has been recognized as the biggest bottleneck for its applicability in practice. The algorithm presented in this paper provides a solution for this problem for any KOS where objects are associated with concepts and where similarity between those objects can be established across datasets.

1.3 Method

In this paper we describe a method to match two ontologies using two disjunct datasets by enriching instances. To enrich an instance i, the concept associations of one or more similar instances of the other dataset are added to i. By doing so, we convert two disjunct datasets into an artificially dually annotated dataset, enabling the application of IBOM. This method tackles the practical problem of the rarity of dually annotated instances, as described above. We call this method instance-based ontology matching by instance enrichment (IBOMbIE).

To illustrate instance enrichment: our goal is to align the vocabularies[3] SWD and Rameau, which are used to annotate books by the German and French national libraries, respectively. Librarians use their own vocabularies to annotate their books, so the books in their corpora are all annotated with a single vocabulary. In the corpus of the German library the book i_{swd} is annotated with the SWD concept *Dachshunds*. Our instance matching algorithm finds a very similar book i_{rameau} in the corpus of the French library. The book i_{rameau} is annotated with the Rameau concept *Teckel*. Therefore, we add the latter annotation to the metadata of the book i_{swd}, which now becomes a dually annotated instance, because it instantiates concepts from both the SWD and Rameau vocabularies.

This paper extends our previous work [27,37] in two ways: first, we apply the proposed method in the large-scale, multilingual setting of the TELplus project,[4] featuring datasets (book catalogs) and vocabularies of the French and British national libraries.[5] Second, we investigate the influences of core parameters[6] of the IBOMbIE method, namely

- When enriching instance i_s we need to decide how many instances of the other dataset are used to enrich i_s. We can choose to enrich i_s with a constant number N of instances, (i.e) the $topN$ most similar instances. We may also enrich i_s with a variable number of instances depending on a *similarity threshold* (*ST*).
- The method used to measure similarity between instances is sensitive to the word distribution of the datasets. Therefore, we investigate the influence of using either word distributions of the source dataset, the target dataset or both datasets on the quality of the resulting alignments.
- Given the multi-lingual setting we evaluate the influence of a translation component on the mapping results.

1.4 Research Questions

The main research questions we will answer in this paper are as follows:

- does the IBOMbIE method work in a large-scale, possibly multi-lingual, scenario?
- How do the parameters influence the results of the IBOMbIE method?
- Is IBOMbIE effective as compared to other ontology matching techniques?

1.5 Experiment and Evaluation

To empirically test our method we apply IBOMbIE to a real-world OM scenario, where the controlled vocabularies of British and French national libraries[3] are matched using their book catalogs as sets of instances. Our test datasets contain hundreds of thousands of instances, which are used to match ontologies containing several thousands of concepts. To measure the quality of results we apply two evaluation methods: a gold standard comparison and a reindexing evaluation. For the first evaluation method we compare results to a manually created alignment, which is produced by the

[2] Throughout this paper we use the term *instance* very broadly, namely as whatever experts consider the extension of a concept in an application requiring some kind of associations of objects with concepts.

[3] Vocabularies are considered a kind of ontology. See Sect. 2 for our definition of the word *ontology*.

[4] http://www.theeuropeanlibrary.org/telplus/.

[5] In a more elaborate report [26], we match three ontologies: that of the French, German and British national libraries, namely Rameau, SWD and LCSH, respectively.

[6] We describe these parameters in more detail in Sect. 3 and report on their influence on the performance of IBOMbIE in Sect. 5.

MACS project.[7] The second evaluation method is novel: a bidirectional reindexing method based on the unidirectional method proposed in [14]. In this method a separate set of dually annotated instances is used to measure the correctness of mappings.

1.6 Findings

Taking the different word distribution into consideration and translating instances improves performance slightly. As the increase of computational complexity is minimal those optimizations seem worthwhile. The two parameters of the IE process, $topN$ and ST, have significant influences on the final mapping results. However, the most simple configuration outperforms the rest (namely $topN = 1$ and $ST = 0$).

Comparing the performance of IBOMbIE with other OM algorithms, we see that both in terms of run time and quality of the end result IBOMbIE is a competitive algorithm that can significantly increase the applicability of instance-based matching methods.

1.7 What to Expect from this Paper?

This paper presents an extensional ontology matching method that works in the absence of dually annotated corpora, and assess the viability of the method in a specific use-case, where we show that it can be a very useful extension of existing methods. Given problem-driven approach, driven by a real-world application in the library domain that started this line of research, we focus on technical aspects of the approach, rather than performing a broad, domain cross comparison.

This paper extends previous work in two ways: we apply the method introduced in [27] on a large-scale, multi-lingual (and thus very challenging) use-case, and second, exhaustively evaluate the possible parameters of the algorithm using two different ways of evaluating the matching results. In addition to the evaluation results in [27] we consider this sufficient proof for the power of IBOMbIE for Ontology Matching, especially when instances are available of which the similarity can be measured, as is the case in our application domain.

The rest of this paper is structured as follows: in Sect. 2 we discuss related work before we explain IBOMbIE in detail in Sect. 3. In Sect. 4 we introduce the scenario that we use to test the performance of different configurations of IBOMbIE. In Sect. 5 we describe our experiments and the results thereof. IBOMbIE is compared with other OM algorithms in Sect. 6. Finally we state our conclusions in Sect. 7.

[7] In the MACS project the vocabularies of European national libraries were aligned manually. http://macs.cenl.org.

2 Related Work

2.1 Instance Matching

Instance matching is a fundamental problem in many application domains, such as E-business, data migration and integration, information sharing and communication, web service composition, semantic query answering, etc. Diverse solutions to the matching problem have been proposed during the past few decades. In the database community particular efforts were put into schema matching, which corresponds to ontology matching in the Semantic Web context. An overview over these efforts is provided in [24]. However, there are significant differences between the two types of problems: databases schemas are usually much smaller than the thesauri we consider (with several thousands of concepts), and instances formalised in ontologies are normally richly described with formal semantics. This means that the extension of a concept of an ontology is far more characteristic for its overall, i.e. including intensional, semantics as compared with the extension of an attribute in a database.[8] This implies that instance-based methods for schema matching are in general not applicable in cases as considered in this paper and that most relevant work comes from the ontology matching community. The readers are referred to [8] for a broader overview of this field of research.

2.2 Ontology Matching

There are many different kinds of conceptual and data-structures that need integrating: database and XML schemas, ER models and conceptual graphs, etc. In [8] the authors argue that most work in matching such structures has been done in matching database and XML schemas, as well as ontology matching, most recently in the context of research on the Semantic Web.

Common to database schemas and ontologies is that they provide vocabularies for terms and constrain their meaning, but ontologies usually come with a richer formal semantics, which creates specific challenges and opportunities for the matching task. In this paper we will use the term "ontology" in the broad sense throughout this paper, (i.e) as a KOS relating concepts and instances. This includes controlled vocabularies, thesauri and "canonical" Semantic Web ontologies in RDF(S) or OWL.

There are four elementary automatic ontology matching methods: terminological, structure-based, semantic-based and instance-based methods [8]. Terminological methods use lexical data in ontologies to discover concept mappings.

[8] This fact is reflected in the fact that the similarity of extensions is often used to evaluate the quality of a concept mapping, see, e.g. [1]. Similarly our reindexing evaluation uses this principle.

Structure-based methods use the internal or external structure of concepts to deduce specific relations between concepts. Semantic-based methods use generic or domain-specific rules and/or background information to find correspondences between ontologies. Instance-based methods, finally, use concept extensions to align ontologies, where the extension of a concept consists of the set of instances that are associated with it.

2.3 Instance-Based Matching Methods

Instance-based matching has several advantages: first, it focuses on the *active part* of ontologies, (i.e) the instances, which reflects what those concepts really refer to in practice. Second, it is less subjected to lexical issues, such as the use of synonyms in labels, as the similarity is determined by their extensions/instances rather than by the labels or descriptions of the concepts. Third, this method is resistant to a small percentage of errors on the manual annotations, which is inevitable due to variations in the annotation strategy.

We follow [23] who identifies two main cases for instance-based ontology matching:

1. those that compare common extensions, (i.e) dually annotated instances, and
2. those for which no common extension exists.

2.4 Instance-Based Matching in the Presence of Dually Annotated Instances

When a dually annotated dataset is available, many statistical co-occurrence based measures can be directly applied to quantify the overlap of extensions of concepts, which produces candidate mappings [13,15,38]. In a survey in 2006 Choi et.al [4] reported that 4 out of 9 systems they studied used instance-based methods, namely LSD [5], GLUE [6], MAFRA [21] and FCA-Merge [29]. Many modern systems, such as RiMOM [19], apply combinations of mapping techniques and often include an instance-based component. This even holds for approaches in rather expressive representation languages [9].

The most common approach to extensional matching is using Jaccard-like similarity measures, such as in [18]. Udrea et al. [33] use such measures as a basis, which is later extended with logical inference. Other variants use the DICE similarity [30], or the Jensen-Shannon distance [38]. In [15] a number of alternative measures, including Jaccard coefficient and variation, point-wise mutual information and log-likelihood ratio are compared in a case of matching two Dutch thesauri based on the books they were annotated with. This work was extended in [36].

Common to all those approaches is that the concepts to be matched are associated with a sufficient number of instances. That is often not the case. There are two approaches to instance-based matching when no dually annotated instances are available:

1. Aggregate the information of the instances into virtual documents representing the concepts of two ontologies, and match the concepts based on those virtual documents.
2. Match instances from two data-sets, enrich each instance with annotations from the most similar instance(s) of the other ontology, thus creating a double annotation.

2.5 Instance-Based Matching Without Dually Annotated Instances: Aggregation-Based Approaches

When the instance sets of two ontologies are disjoint or have little overlap, one solution is to aggregate instance information as features of concepts and derive concept similarity from such aggregated instance-based representation (those aggregated representations are often called virtual documents). The Semantic Category Matching approach [12] compares feature vectors for each concept pair using keywords found in the instances and then determines similar feature vectors by a structural matcher. Another idea is to use *Formal Concept Analysis-Merge* [29] to extract instances from text documents. Based on the hypothesis that concepts that always appear in the same documents are supposed to be merged, *Formal Concept Analysis* techniques can be applied to compute concept lattices, which are subsequently used to merge two ontologies. The authors of the GLUE [6] system proposed a notion of concept similarity in terms of the joint probability distribution of instances of the concerned concepts. Using a Naive Bayes text classification, instances of one ontology are classified to concepts of the other ontology based on their textual information. Zaiss [39] presents two more instance-based matching methods, one of which is based on aggregation of both the properties and the instances of the concepts that are to be mapped. A similar idea was exploited earlier by Wang et al. [35] where a classifier was trained to classify pairs of source and target concepts into matches and non-matches. Todorov et al. [31] use Support Vector Machines for weighting features of similarities between classes of instances, in [32] they extend this method to the heterogeneous case. Finally, Li [20] uses Neural Networks to similar ends.

2.6 Instance-Based Matching Without Dually Annotated Instances: Instanced-Based Approaches

Common to all the approaches discussed above is that they aggregate over the instances of two concepts to find semantic similarity between them. Given that in many ontologies instances are richly formalized an alternative is to focus on

similarity of the instances themselves. As it has been shown that extensional overlap is a strong indication for similarity of concepts the idea is to identify same or similar instances from two ontologies and use them as dually annotated instances to derive concept mappings.

Of course, this approach requires instances to be matched. Instance matching, also called object matching, entity resolution or instance unification, is a core problem for the Semantic Web, and has recently attracted increased research attention [10]. We will restrict the discussion of the related work in instance unification to pointing the reader to a very useful overview [17] for work in the database and XML matching community, and the instance matching tracks at the recent OAEI evaluation initiatives [7].

The main contribution of this paper is to formally introduce in detail, and to provide a thorough analysis of instance-based matching by instance enrichment, an idea we first introduced in [15] and which is to calculate an artificially dually annotated corpus. We neither introduce new instance mapping nor extensional matching methods, but use well-established, and simple, techniques from both fields. It is the combination of both that, to the best of our knowledge, is an idea that has been hitherto unexplored.

3 Matching and Enriching Instances

This section gives an overview of the IBOMbIE algorithm, and discusses the issues that inspired the empirical research reported in Sect. 5.

As previously mentioned this paper addresses the specific problem of ontology matching: we focus on vocabularies with concepts and instances that are specified in a semantically rich formal ontology language, such as RDF(S) or OWL. As said in Sect. 2, we use the term ontology in a broad sense, to include less formal KOS. With this definition, knowledge definitions with fewer formal axioms, such as SKOS, FOAF and schema.org, are also considered ontologies. The generic definition of *ontology matching* is then the task of finding mappings between entities in two ontologies [8]. Here we will tackle the problem of mapping *concepts* of two ontologies, (i.e) entities that are clearly distinguished as classes of objects. Most ontologies make such a distinction between concepts and instances annotating those concepts, either as direct extensions (using, e.g. the `rdf:type` predicate), or more loosely (as in SKOS, in which the use of Dublin Core `dc:subject` is recommended).

In [16] we argued that the *meaning* of a mapping depends strongly on the context and the purpose of the application of a mapping. A good example is *extensional ontology mapping*, where the mapping between two concepts is determined by the similarity of usage of objects related to concepts. This paper extends existing method for extensional ontology matching where these extensions are disjoint, but comparable. Later in the evaluation and in our specific usecase *instances of a concept* will be sets of books annotated with that concept, as usual in the Information Science domain [28].[9] As a shortcut we will call the instances of an ontology *its dataset*. The IBOMbIE algorithm then matches concepts from two ontologies O_1 and O_2 which annotate instances of two datasets D_1 and D_2, respectively, (we also say that we match O_1 and O_2).

From a bird's eye view, the IBOMbIE algorithm consists of three independent steps:

1. match instances of D_1 (resp., D_2) with most similar instance(s) of D_2 (resp., D_1) and
2. enrich the instances of D_1 (resp., D_2) by adding the annotations of their most similar instance(s) of D_2 (resp., D_1).
   This second step is the simple, but crucial idea of IBOMbIE. The final step is to apply a classical *instance-based ontology matching* method:
3. match O_1 and O_2 using a co-occurrence based similarity measure, in our case JC_c (taken from [15][10])

$$\mathrm{JC}_c(c_1, c_2) = \frac{\sqrt{|i_1 \cap i_2| * (|i_1 \cap i_2| - 0.8)}}{|i_1 \cup i_2|}, \tag{1}$$

where i_x is the set of instances that are annotated with concept c_x.

Instance matching and enriching critically depends on the type and richness of information that is available for the instances in the ontology. Without loss of generality we assume in the following that each instance can be described as a set of features, which could be words in a document, concepts from the metadata or other related objects. This allows us to use the well-known Vector Space [25] model to determine similarity between instances.[11]

In our use-cases instances are documents, and the features are the words in those documents. In order to keep the standard terminology of the model as used in Information Retrieval we directly refer to the words in the documents as our features. More formally, in the following we consider

[9] There are common use-cases, e.g. reindexing of books in a library, where objects annotated by a SKOS concept in a thesauri can be considered its extension. This is not strictly the extension of a concept in a model-theoretic way but compliant with the practice in Ontology Matching.

[10] We use a simple adaptation of Jaccard similarity that was identified in [15] as the most simple, reliable and successful measure. A more exhaustive study of the impact of the choice of similarity for extensional mapping would be interesting, but is out of the scope of this paper.

[11] In other application areas other notions of similarity might be more appropriate, but as our metadata are mostly textual the VSM is the obvious choice.

our datasets D_1 and D_2 to consist of textual documents annotated with a concept in O_1 and O_2, respectively. Without loss of generality each document will be represented as a vector of words. In the following we will give more details on our methods for matching and enriching instances.

3.1 Instance Matching

In order to enrich an instance with the annotations of its most similar instance(s) in the other ontology, we need to determine which instance(s) actually is (are) most similar. *Instance matching* (IM) is the first step.

Instance matching is straightforward in the presence of inverse functional properties or shared keys, such as the *International Standard Book Number* (ISBN). Otherwise, approximate IM algorithms are required that use features to predict similarity between objects. The Vector Space model provides an abstract model, where documents are represented as vectors of features (in our case words) in a *vector space*. Let us briefly recall some basic notions: the similarity between two documents is negatively correlated with the angle between the vectors representing those documents. The similarity between two documents is quantified by the *cosine similarity*:

$$\text{cosine\_sim}(d_1, d_2) = \frac{\mathbf{d_1} \cdot \mathbf{d_2}}{|\mathbf{d_1}||\mathbf{d_2}|} = \frac{\sum_{j=1}^{n} w_{j,d_1} w_{j,d_2}}{\sqrt{\sum_j w_{j,d_1}^2}\sqrt{\sum_j w_{j,d_2}^2}},$$

where $\mathbf{d_1}$, $\mathbf{d_2}$ are the vectors representing the two documents being compared, n is their dimension and w_{j,d_k} is the coordinate of $\mathbf{d_j}$ along dimension j.

A commonly used weight to represent textual data in VSM is TF-IDF, which expresses the significance of a word w in a document d that is part of dataset D. The TF-IDF weight is the product of the *term frequency* (TF) of w in d and the *inverse document frequency* (IDF) of w in the set D:

$$\text{tf-idf}_{w,d,D} = tf_{w,d} * idf_{w,D}$$

The TF of w in d is defined as dividing the word frequency ($n_{w,d}$) by the document size ($|d|$). This division by $|d|$ is meant to prevent the measure from having a bias towards large documents, since large documents contain many words and therefore have higher word frequencies on average:

$$tf_{w,d} = \frac{n_{w,d}}{|d|}$$

The IDF of w is defined as the logarithm of the size of the dataset ($|D|$) divided by the number of documents in which the word w occurs:

$$idf_{w,D} = \log \frac{|D|}{|d \in D : w \in d|}.$$

If a word w occurs in many documents, the IDF will be low. If a word w occurs in few documents, the IDF will be high. Thus the IDF quantifies the significance of the occurrence of a word in a corpus.

Traditional IR scenarios consider a single *word distribution*, namely the word distribution of the dataset. In the IBOM algorithm there are always two datasets, each with their own word distribution. This gives us three options to consider:

1. IDF^{single}: only consider the word distribution of the source dataset. This is how the IDF of a word is generally determined in traditional IR algorithms.
2. IDF^{local}: use the local word distribution of w to calculate the IDF value of w, (i.e) when w is part of a document in dataset D_1 we consider the word distribution of D_1 to calculate $IDF(w)$. In the IBOMbIE algorithm there are always two word distributions: the word distributions of D_1 and D_2.
3. IDF^{global}: consider a single word distribution on a global scale, (i.e) when calculating the IDF of any word in D_1 or D_2, consider the word distribution of the union of the datasets: $D_1 \cup D_2$ (an approach similar to the one followed by GLUE [6]).

IDF^{single} is the simplest option, which may fail to correctly quantify the importance of a word w when w is rare in the dataset, but common in the dataset that is being enriched. We expect that the IDF^{local} option will give the best results, because it quantifies the importance of a word within its own dataset. IDF^{global} may also provide a reliable IDF quantification, when the importance of a word differs significantly in two datasets. This observation leads to the first research question.

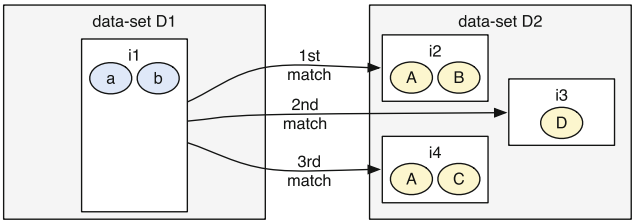
RQ1: What is the impact of using different word frequency distributions over sets of documents on the performance of IBOMbIE?

Using empirical results we will answer this question in Sect. 5.2.1.

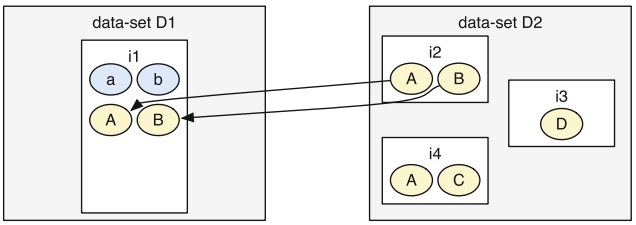
A well-known problem when dealing with words as features is that they need to be reduced to their common forms, or stemmed, for the similarity between the vectors to be reliable.[12] When comparing documents from multi-lingual datasets, the features need to be translated to a common language. Our approach to instance translation is very simple: all words are translated by the *Google translate* web service,[13] and all these translations constitute a new, now translated, document. In Sect. 5.2.2 we will answer the following research question:

[12] Through stemming *watching* and *watched* both become *watch* after stemming. We have used the Snowball stemmer http://snowball.tartarus.org/.

[13] http://translate.google.com/.



(a) instance matching



(b) instance enrichment ($topN$=1)

Fig. 1 Instance enrichment process. The is are instances (documents); a, b, A, B, C, D are concepts used to annotate these documents

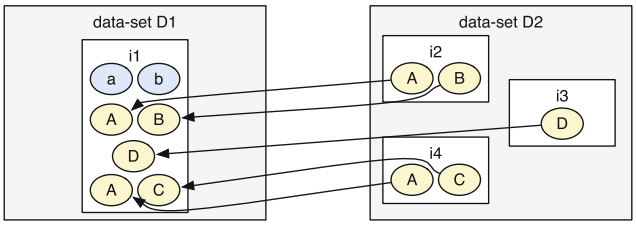


Fig. 2 Instance enrichment parameter: $topN = 3$

RQ2: Does even a naive instance translation method have a positive impact on the IBOMbIE process?

After having discussed how to identify similar instances let us study the options for enriching instances once the most similar instances have been determined.

3.2 Instance Enrichment

Consider the following scenario: we have two datasets D_1 and D_2, where the instances of D_1 and D_2 are associated with concepts of ontologies O_1 and O_2, respectively. As depicted in Fig. 1a, when comparing i_1 to the instances of D_2, the IM process ranked i_2, i_3 and i_4 as the first, second and third most similar instances, respectively. To enrich instance $i_1 \in D_1$ with $i_2 \in D_2$ we associate i_1 with the concepts that i_2 is associated with, (i.e) we add the annotations of i_2 to i_1 as shown in Fig. 1b. The result is that instance i_1 has become a dually annotated instance, because it is annotated with concepts of both O_1 and O_2.

There are two crucial parameters of the IE process: the $topN$ and the *similarity threshold* (ST) parameters. Tuning these two parameters may have a significant influence on the quality of the end result.

The $topN$ parameter defines from how many instances we add the associated concepts to the instance that will be enriched. To illustrate the dynamics of the $topN$ parameter, we re-use the scenario as depicted in Fig. 1a. If N has been set to 1, i_1 will be enriched with the concepts of the single, most similar instance of D_2, as shown in Fig. 1b. With N set to three instance i_1 will be enriched with the three most similar instances, as depicted in Fig. 2.

A larger value of the $topN$ parameter means that instances will be enriched with more concepts. Therefore, a larger N causes more concept associations to be created, resulting in a higher number of mappings generated by applying JC_c and thus a final result with a potentially higher coverage. With a smaller N we can say the enrichment algorithm is more selective, meaning instances will be enriched with relatively more similar instances, which implies better quality mappings in the final result.

RQ3: How does the $topN$ parameter influence the performance of IBOMbIE ?

This question will be answered in Sect. 5.2.3.

The ST parameter dictates a minimum similarity ST between i_1 and i_2 before i_1 is enriched with the concepts of i_2. This implies that, unlike the $topN$ parameter where an instance is always enriched with the N most similar instances, it is possible that i_1 is not enriched at all.

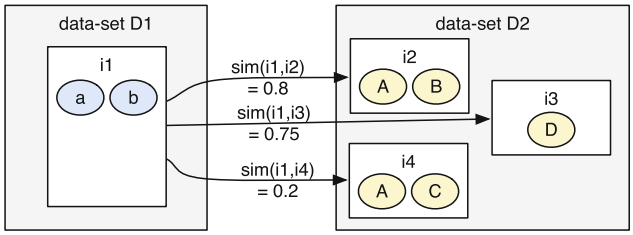
To illustrate the dynamics of the ST parameter we depict a scenario in Fig. 3. Figure 3a shows the results of the IM process: the similarity values between i_1, the instance that will be enriched, and the instances of the other dataset: i_2, i_3 and i_4. In Fig. 3b the threshold ST is smaller than both the similarity between i_1 and i_2 and that between i_1 and i_3, so i_1 is enriched with the concepts of i_2 and i_3.

As in the case with the $topN$ parameter, we have to balance the selectiveness and the number of concept associations. When using a low ST, the IBOMbIE algorithm will enrich an instance with the conceptual annotations of relatively many instances. As ST increases, the selectiveness of the IBOMbIE algorithm increases, resulting in fewer but potentially higher quality annotations, which may lead in turn to fewer but better mappings.

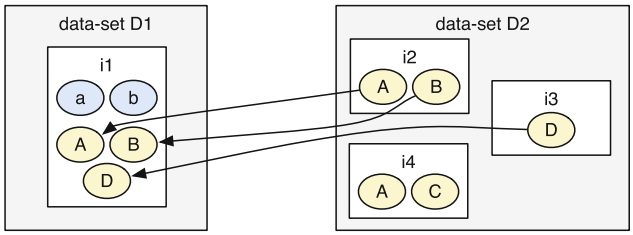
RQ4: What is the influence of a Similarity Threshold on performance of IBOMbIE ?

This question will be answered in Sect. 5.2.4.

Instead of using either one of the $topN$ or ST parameters, we may also use both to tune the selectiveness of the IBOMbIE algorithm. Naturally that raises the question, which we answer in Sect. 5.2.5:



(a) instance matching



(b) 0.75 >ST> 0.2, e.g., ST=0.5

Fig. 3 Instance enrichment parameter scenario: ST

RQ5: Does a combination of the two instance enrichment parameters improve performance as compared with using a single parameter?

Even more so than when configuring a single parameter, we have to find a balance between the selectiveness of the algorithm and the number of concept associations, when configuring both the $topN$ and the ST parameters. This trade-off is analog to the precision versus recall problem [2]: when we desire high precision we need to be selective, which will be at expense of the recall. Vice versa, when we want a high recall we have to be less selective, which will most likely decrease the precision.

4 Evaluation Scenarios and Methods

4.1 Datasets

We use a real-life OM scenario to empirically test the IBOMbIE method: *TEL*, named after the *TELplus* project.[14] In this scenario we match the controlled vocabularies of the English and French national libraries, using their book catalogs as collections of instances.

The controlled vocabularies in question, *LCSH* and *Rameau*, contain, respectively, 339,612 and 154,974 concepts. All concepts in the controlled vocabularies have a preferred label and a variable number of alternative labels. Partial hierarchical and associative concept relations are also present. Both vocabularies are accessible as Linked Data over the Web.[15]

The datasets of the English and French libraries[16] contain, respectively, 2,505,801 and 1,457,143 annotated books. Though the book texts are not available for our experiment, we can exploit the metadata in the *records* that are created for them: title, author, publisher, sometimes abstract, etc. An example is shown in Listing 1. A challenging aspect of this scenario is that the collections of book records originate from different countries and are thus in different languages. Since we use text-based instance similarity measures, this aspect is a significant handicap for the IBOMbIE algorithm.

Listing 1 Example of an English book instance

```
<record>
    <identifier>000084547</identifier>
    <dc:title>The Indian earthquake</dc:title>
    <dc:creator>
        Andrews, C. F. (Charles Freer),
        1871–1940.</dc:creator>
    <dc:publisher>London : G. Allen & Unwin,
        1935.</dc:publisher>
    <dcterms:issued>1935.</dcterms:issued>
    <dcterms:extent>130 p. ; 19 cm.
        </dcterms:extent>
    <dc:language>eng</dc:language>
    <dc:abstract>Describes the scene of the
        earthquake in North Bihar in 1934 and
        efforts made for relief.</dc:abstract>
    <dc:type>text</dc:type>
    <mods:location>British Library HMNTS
        07108.a.9.</mods:location>
    <telplus:topicalSubject xml:lang="en"
        identifier="sh2005000327">
        Earthquakes—India
        </telplus:topicalSubject>
</record>
```

4.2 Evaluation

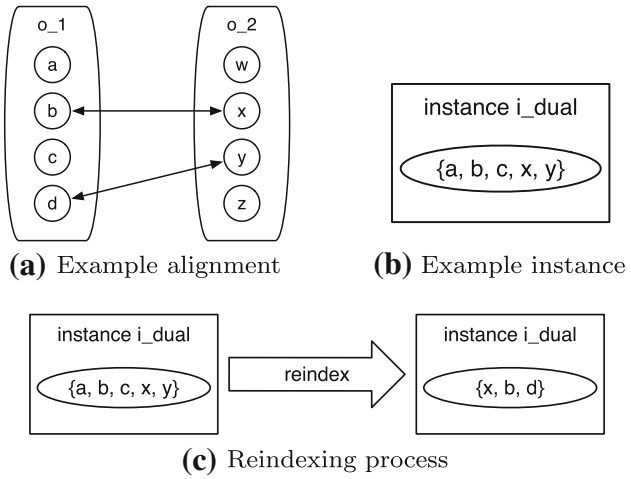
We use two evaluation methods to evaluate alignments: *gold standard* and *reindexing* evaluation methods.

In the first evaluation method, we use the alignment between the LCSH and Rameau vocabularies that are manually created during the MACS project.[5] Since the alignments are manually created, the mappings are of good quality. The 57,650 mappings in the MACS alignment (the version we obtained) identify correspondences between 55,623 LCSH concepts and 55,963 Rameau concepts, covering 16 % of LCSH and 36 % of the Rameau vocabulary. We do not know whether the alignment focuses on specific subsets of the vocabularies.

[14] The TELplus project (http://www.theeuropeanlibrary.org/telplus) stems from *The European Library* initiative (http://www.theeuropeanlibrary.org/) which offers access to 48 national libraries of European countries.

[15] See http://id.loc.gov and http://stitch.cs.vu.nl/rameau.

[16] See http://catalogue.bl.uk and http://catalogue.bnf.fr.



(a) Example alignment **(b)** Example instance

(c) Reindexing process

Fig. 4 Reindexing example

Although the MACS alignment does not provide the complete list of all correct mappings, it is an invaluable means for the automatic evaluation of alignments that are produced by the IBOMbIE algorithm. We consider a mapping *judgeable* when one of the concepts occurs in the MACS alignment, and a mapping is *non-judgeable* when neither of the concepts is used in the MACS alignment. To quantify the quality of an alignment, we apply, for its judgeable mappings, the well-known precision (P) and recall (R) formulas:

$$P = \frac{|\text{Correct} \cap \text{Found}|}{|\text{Found}|}, \quad R = \frac{|\text{Correct} \cap \text{Found}|}{|\text{Correct}|}$$

where Correct is the set of mappings from the MACS gold standard and Found is the set of (judgeable) mappings from the evaluated alignment.

The second automatic evaluation method is an adaptation of the *reindexing scenario*, in which an alignment and the original conceptual annotations are used to yield new annotations using concepts from a different vocabulary [14]. When a corpus of already dually annotated documents is available, these documents can be used to automatically evaluate the quality of an alignment in that application scenario.

To illustrate the reindexing method, consider the alignment in Fig. 4a, where two ontologies are shown (o_1 and o_2) and a double arrow indicates a mapping between two concepts. Figure 4b shows a dually annotated instance. To reindex an instance, the original annotations are replaced by the concepts that alignment A maps them to, as depicted in Fig. 4c. In this example, A maps concept b to x, so b is replaced by x. In the same fashion, x is replaced by b and y by d. Annotations to concepts that are not mapped are replaced with the empty set. Therefore, after re-indexing the instance has three annotations, since the concepts a and c are not mapped.

To calculate the precision and recall of an alignment considering a full set of dually annotated instances we use the following equations:[17]

$$P = \frac{\sum^{\text{Reindexed}} \frac{|\text{Ref} \cap R(\text{Ref})|}{|R(\text{Ref})|}}{\text{Reindexed}}$$

$$R = \frac{\sum^{\text{Total}} \frac{|\text{Ref} \cap R(\text{Ref})|}{|\text{Ref}|}}{\text{Total}}$$

In these equations Ref is the reference set, (i.e) the original (and thus correct) conceptual annotations of a book, $R(\text{Ref})$ is the set of concepts obtained by reindexing the original annotations, Total is the total number of books that were used to evaluate and Reindexed is the number of books that could be reindexed, (i.e) for which $R(\text{Ref})$ was not an empty set.

In order to apply the reindexing evaluation method in the TEL scenario, dually annotated instances are required. Fortunately, many books in the TEL datasets have shared *ISBN* identifiers. ISBN is an international book identification standard. When a book in the French collection has the same ISBN as a book in the English collection, we know that those records correspond to the same actual book. There are 182,460 books in the English and French datasets that share an ISBN identifier, which concerns 7 % of the English and 12.5 % of the French book records. Although this is a relatively small number of instances, we assume that the number of ISBN matching books is sufficient for creating dually annotated instances to perform a reindexing evaluation. Note that these dually annotated instances are excluded when generating an alignment using IBOMbIE, as using those instances for both evaluation purposes and alignment generation would bias the evaluation results.

5 Experiments and Results

In [27] we have provided initial results indicating that the IBOMbIE algorithm is capable of producing an alignment based on the extension of concepts using two *disjoint* datasets. In this section we discuss the results of a more exhaustive empirical study inspired by the issues identified in Sect. 3.

5.1 Experimental Setup

The IBOMbIE algorithm has been implemented using the Java programming language. We use a custom VSM implementation, which allows us to include several optimizations

[17] A directional approach of the reindexing evaluation method is applied in [14]. In this paper we use a bidirectional approach, as we are interested in the general quality of an alignment, as opposed to converting instance annotations in a specific direction.

(for details, see [26]). The IBOMbIE algorithm uses a single thread; no multi-threading techniques were used to speed up the process.

All the experiments were conducted on a single machine, with 32GB internal memory. For performance reasons, the index of the source dataset is stored in main memory during the IM process. The TEL scenario features large datasets: the two book catalogs are each 4–9 GB. Therefore, a large quantity of main memory is ideal (for the TEL scenario, IBOMbIE uses approximately 7GB of the main memory). As the price of RAM memory steadily decreases, we do not consider this requirement as a limitation of the IBOMbIE algorithm.

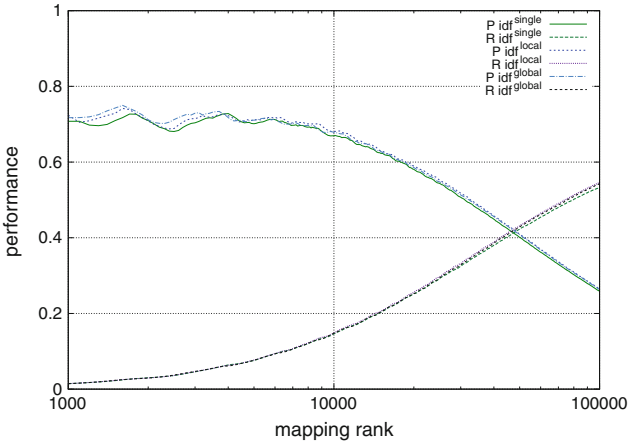


Fig. 5 IDF experiment evaluation results: gold standard comparison

5.2 Experiment Results

The evaluation data are often presented on a logarithmic scale, because that allows us to examine the quality of the early mappings, as well as the global performance of a whole alignment in a single figure.

The quality of alignments are plotted against the mapping ranks. Mappings are ranked with respect to their similarity values estimated using the corrected Jaccard measure (Eq. 1). Therefore, in these plots it is clearly visible as to how many mappings a certain evaluation result applies.

Consider Fig. 5, where recall and precision of three different alignments are plotted (to be explained later). The plot shows that if we take the 1,000 most confident mappings we get a recall of almost 0 % and a precision of around 70 %. Considering the first 100K most confident mappings we can read a recall of approximately 55 % with a precision of around 25 %. A good alignment is thus represented by a sharply rising recall curve, and a stable precision curve of maximal height.

As a default setting, the parameters of the IBOMbIE algorithm are set as follows: instance translation and word stemming are enabled, we use IDF^{local}, we set $topN$ to 1 and ST to 0. All options are set to default, unless stated otherwise in the presentation of the experiment results below. In all experiments we show results of both the gold standard and reindexing evaluation methods, except in Sects. 5.2.1 and 5.2.2 due to space limitations. In those experiments the findings in the reindexing evaluation are similar to those in the gold standard evaluation.

In this paper we use only the precision and recall measures to clarify our findings. In a more elaborate document we also include f-measure figures [26]. We chose to omit the f-measure figures for two reasons: (1) the large amount of figures can be overwhelming and (2) the precision and recall are most informative to explain our findings.

5.2.1 Word Distributions

In this experiment, we answer RQ1 regarding the influence of the choice of the word-distribution in the weighting of attributes, by testing the performance of the IBOMbIE algorithm using different definitions of the IDF, as explained in Sect. 3.1. Thus the IDF option is set to either IDF^{single}, IDF^{local} or IDF^{global} in each of the experiments.

Figure 5 displays the evaluation results of the alignments generated using the three different IDF configurations.[18] We see that the differences in quality of the alignments are marginal. The quality of the alignment produced with IDF^{single} is slightly worse than the other two alignments, which indicates that taking into account the word distributions of the different datasets increases the performance of the IM process in the context of the alignment task. From Fig. 6, we can see the alignments generated using different word distributions do have substantial overlaps. Here, the overlap is the proportion of the common mappings over all mappings generated by the two methods.

In conclusion, we see that taking both word distributions into account has a tangible impact on the performance of IBOMbIE, while having minimal impact on the run time. As the results show that IDF^{local} leads to the best performance, we use IDF^{local} in the following experiments.

5.2.2 Instance Translation

To answer RQ2, about the influence of translation on the mapping process, we have generated two alignments: one with and one without instance translation.

[18] We are aware that the precision-at-n representation is often used in evaluations. However, we have chosen to use precision and recall curves, because these enable a more in-depth analysis of the evaluation results.

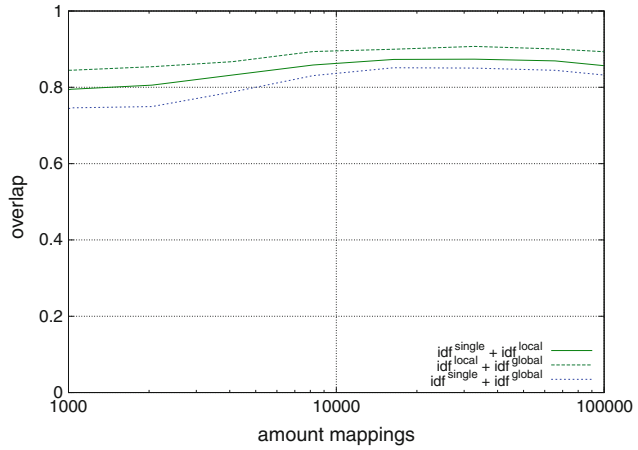


Fig. 6 IDF experiment: overlap of alignments

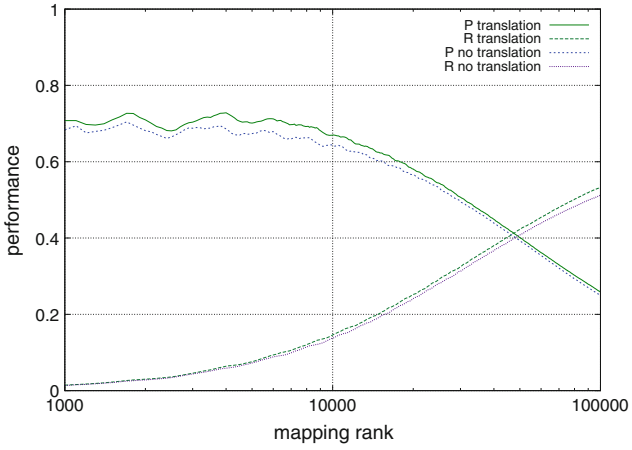


Fig. 7 Instance translation experiment evaluation results: gold standard comparison

When instance translation is disabled, we do not apply word stemming. The two languages of the datasets are different, and thus different stemming algorithms would need to be used. Using different stemming algorithms negatively influences the IM process: words that are lexically equal might be stemmed in different ways, rendering the words no longer lexically equal. This can prove especially harmful for language independent text, (i.e) proper nouns (places, persons), that we cannot shield from stemming. Also, words that are not otherwise related may be assigned a same stem, as different languages use different flection mechanisms.

In the evaluation results that are displayed in Fig. 7 we see that, as expected, without translation the algorithm performs relatively well, due to language independent text. With translation enabled, the performance is strictly better. As shown in Fig. 7, the precision of the top 10,000 mappings has been improved substantially. The improvement in precision decreases for the lower ranked mappings. But recall improves when these mappings are considered. Translation basically brings more elements for detecting instance mappings. We expect that it strengthens the robustness of the measures we use to rank concept mappings, especially for the candidates that are derived from a larger amount of linguistic evidence—the influence of individual translation errors will be lower for them. The lower ranked mappings will comparatively suffer more from translation errors. But these errors do not seem to write off the early recall gains brought by lifting more precise mappings higher in the ranks. Given the low complexity of the translation process, these results suggest that adopting translation is a reasonable approach to bring valuable performance improvement.

5.2.3 Parameter: $topN$

To answer RQ3, regarding the influence of the number of similar instances involved in the enrichment, we evaluate the performance of the IBOMbIE algorithm using six different settings of the $topN$ parameter: $topN \in \{1 \ldots 6\}$.

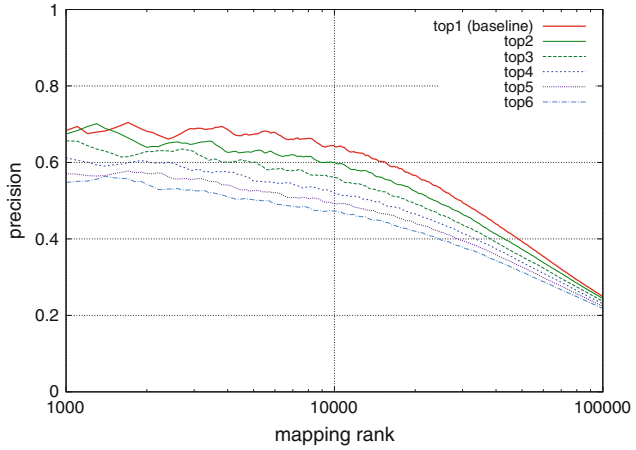
Figures 8 and 9, respectively, show the evaluation results of the $topN$ experiments regarding the gold standard comparison and reindexing evaluation methods. The evaluation results show that a low N results in better precision and recall in the early mappings. As N increases, the difference in performance in the late mappings decreases. The deteriorating performance that accompanies the increasing $topN$ parameter is most likely due to the applied concept similarity measure: the JC_c. The JC_c does not use multi-sets, but sets. Therefore, multiple concept associations that refer to the same concept are counted as a single concept association. An example of a double concept association can be seen in Fig. 2, where instance i_1 has two references to concept A.

We see in Fig. 9a that at approximately 90K mappings the performance w.r.t. precision with a higher $topN$ value eventually exceeds that of lower $topN$ values. Given that the corrected Jaccard measure used for IBOM assigns higher similarity to concepts with more joint instances, more similar instances boost concept similarity, which explains the higher recall. On the other hand, the aforementioned problem of dealing with multi-sets counts far less in case of the mappings with lesser confidence, as there are few overlapping instances anyway.

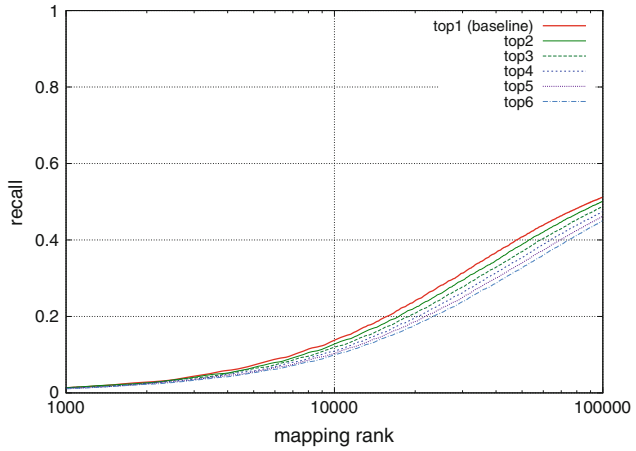
In the following experiments we will use the performance of IBOMbIE with $topN$ set to 1 as our *baseline*, as it is the simplest configuration and results in optimal performance.

5.2.4 Parameter: Similarity Threshold

In this section we answer RQ4, regarding the influence of a similarity threshold on the mapping performance, by studying the effect of using different values of ST. To test the ST parameter independently from the $topN$ parameter, we set the $topN$ parameter to infinity.

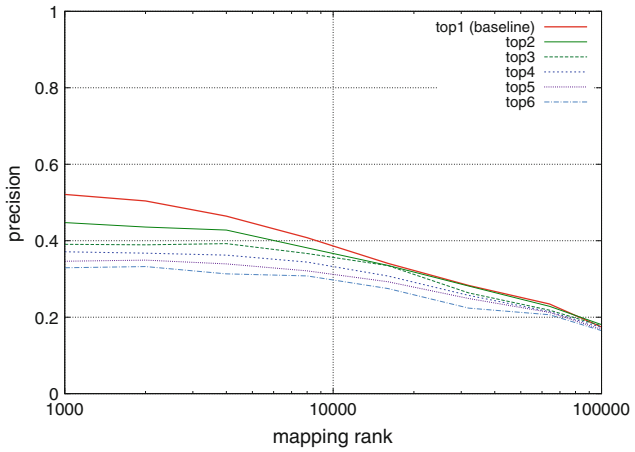


(a) precision

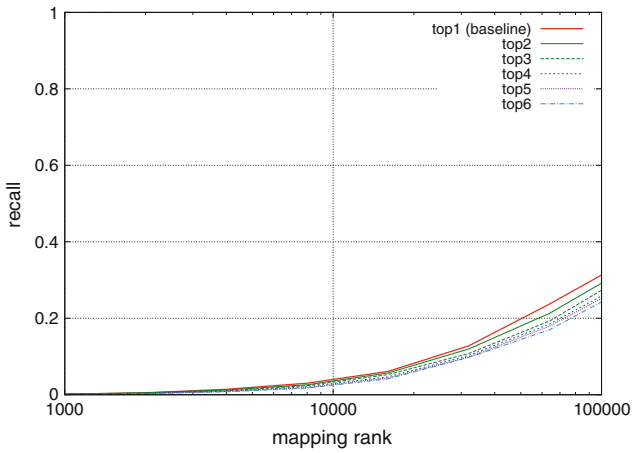


(b) recall

Fig. 8 *topN* experiment evaluation results: gold standard comparison



(a) precision



(b) recall

Fig. 9 *topN* experiment evaluation results: reindexing

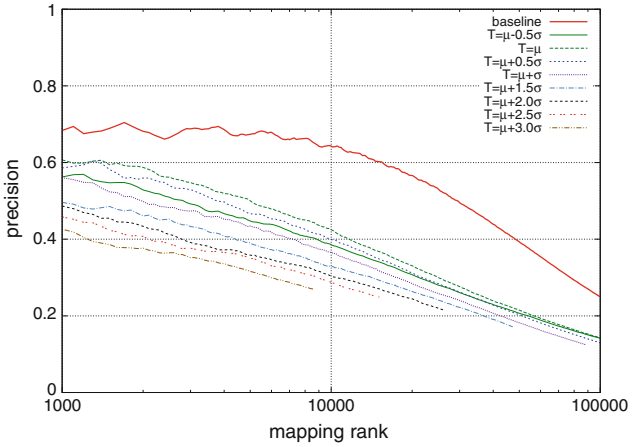
Experience with ST values in several OM scenarios has shown that ST is a context dependent parameter. For example, the average similarity in a multi-lingual environment is lower than when the text of all instances is in a single natural language (see [26] for concrete examples). To obtain default settings of the ST parameter we calculate the *mean* (μ) and *standard deviation* (σ) of the similarity between instances and their closest match in the other dataset. The settings of the ST that were tested are in the range $[\mu-\sigma, \mu+2\frac{1}{2}\sigma]$ with a step size of $\frac{1}{2}\sigma$. The lower bound of ST is set to $\mu-\sigma$ for technical reasons, since the amount of enrichments increases quickly as ST decreases, increasing both the run time and required disk space.

As expected, a higher ST results in less instance mappings (and therefore less concept mappings), but with higher quality, as depicted in Fig. 10. However, Fig. 11 shows that a higher ST leads to a higher recall. This counterintuitive phenomenon is related to the fact that the concept usage is not uniform. The loss of instance mappings can cause the disappearing of the mappings whose concepts are rarely used, while only putting the mappings with regularly used concepts to lower ranks. The way of calculating the recall in the reindexing evaluation (see Sect. 4.2) is heavily influenced by the usage frequency of concepts. If two regularly used concept are mapped correctly, this *mapping between actively used concepts* is counted much more often, and results in a boost of the recall. So even when the total amount of correct mappings decreases, these *mappings between actively used concepts* can result in a higher recall, as measured in the reindexing evaluation method.

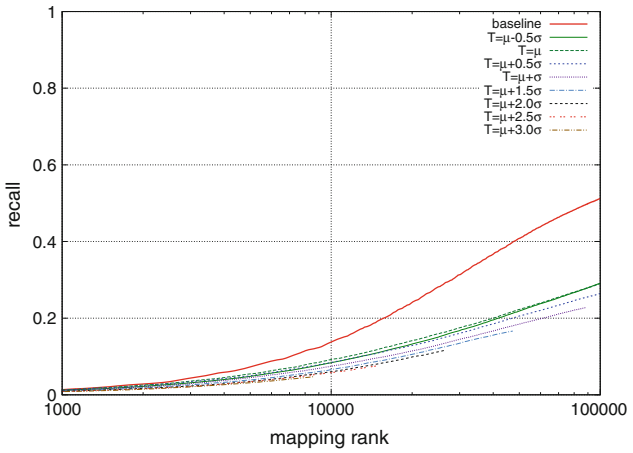
In conclusion, in Figs. 10 and 11 we can see that the IBOMbIE algorithm performs best with $ST = \mu$. Running up settings are $ST = \sigma \pm \frac{1}{2}\mu$. However, the baseline is the best performing configuration, which implies that the number of chosen instances has a higher impact on the mapping results than the similarity between instances.

5.2.5 Combining Parameters

Combining the $topN$ and ST parameters gives fine-grained control over the selectiveness of the IBOMbIE algorithm. In

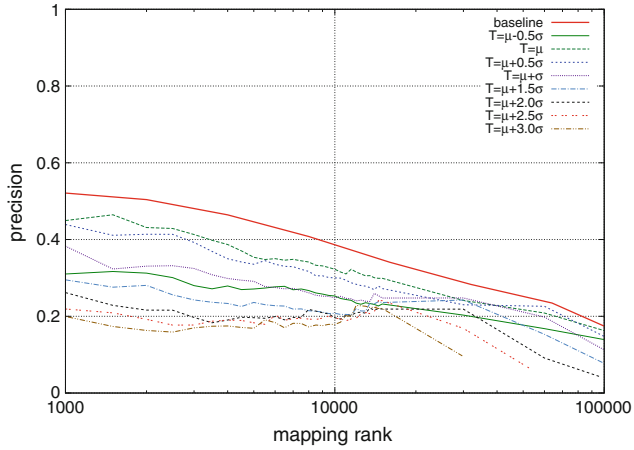


(a) precision

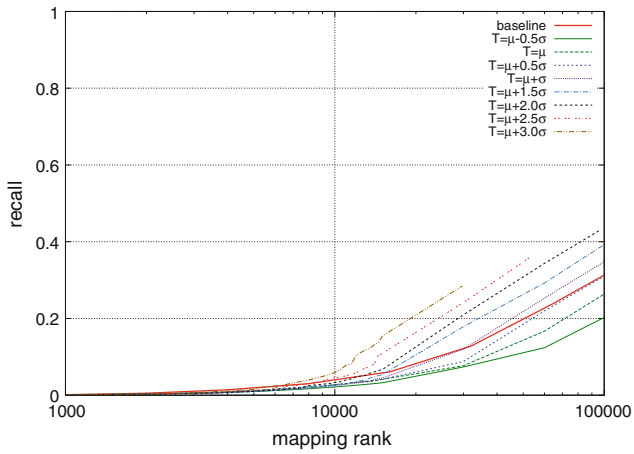


(b) recall

Fig. 10 *ST* experiment evaluation results: gold standard comparison



(a) precision



(b) recall

Fig. 11 *ST* experiment evaluation results: reindexing

this case, only candidates which are ranked within top N and have similarity higher than the threshold ST are selected. We are interested in (1) what combination of the parameters gives the best performance and (2) whether this gives better performance than when using a single parameter. This experiment is conducted to answer RQ5.

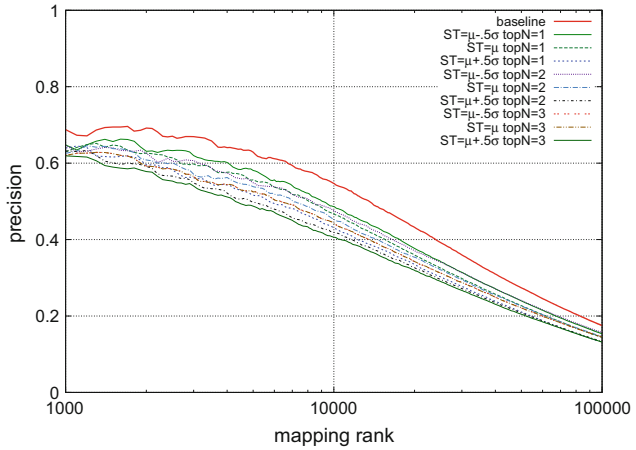
During this experiment we restrict the setting of ST to μ and $\mu \pm \frac{1}{2}\sigma$, because in Sect. 5.2.4 we have seen that these configurations result in the best performance. We will set the $topN$ parameter to 1, 2 and 3, as in Sect. 5.2.3 we have seen that low values of $topN$ result in the best results.

The results of the gold standard evaluation in Fig. 12 show that the baseline still performs best. However, the results of the reindexing evaluation in Fig. 13 show that the alignments generated using the parameters in conjunction match the quality of the baseline. Considering the precision figures we see that the configurations with $topN$ set to 1 and ST set to $\mu - \frac{1}{2}\sigma$ and μ show the best performance in the alignment portion between 1K and 10K mappings.

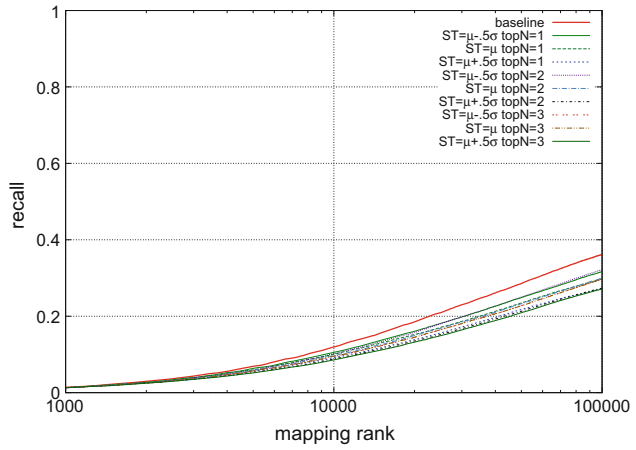
The difference in quality between the baseline and the alignments generated using the combination of the two parameters is significantly smaller than the difference between the baseline and the alignment generated in the previous experiments. In Fig. 13a we see that in early mappings, (i.e) in mappings with higher confidence, the precision of IBOMbIE using two parameters can be better than the baseline. It is safe to conclude that by combining the $topN$ and ST parameters, the performance of IBOMbIE can be better than when using either one of the $topN$ or ST parameters. However, it is very hard to tune the two parameters, as the optimal values may differ in different scenarios.

5.3 Experiment Conclusions

To answer the research questions in Sect. 1: we have seen that the IBOMbIE algorithm can be successfully applied in a large-scale, multilingual ontology matching scenario.

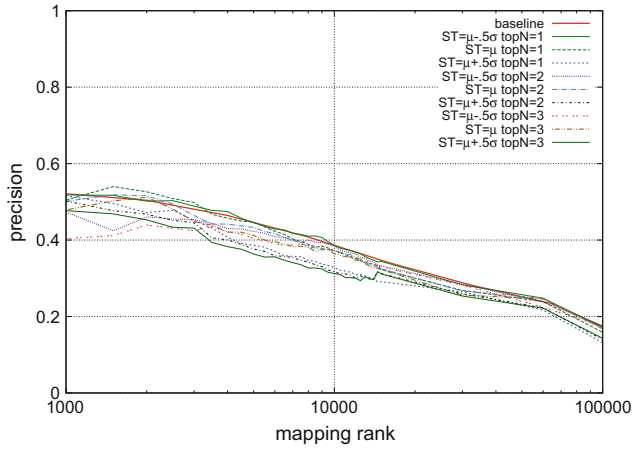


(a) precision

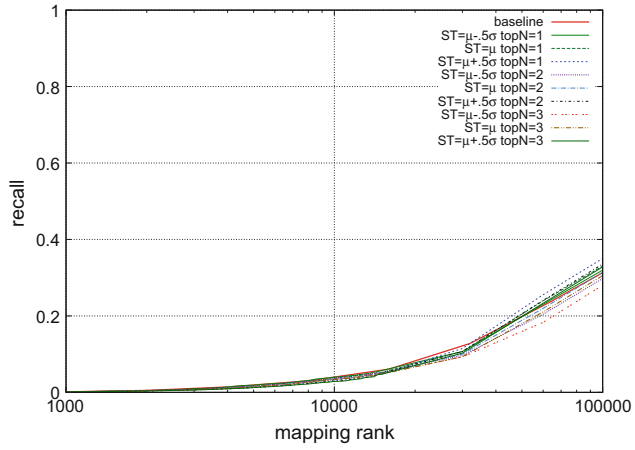


(b) recall

Fig. 12 Combining parameters experiment evaluation results: gold standard comparison



(a) precision



(b) recall

Fig. 13 Combining parameters experiment evaluation results: reindexing

Conclusions concerning the parameters of IBOMbIE are as follows:

- Taking into account the word distribution of both the source and target dataset have proved to marginally influence the quality of the alignments. Do note that the resulting alignments have tangible differences, as observed when considering the overlap between the alignments.
- We have seen that a simple translation algorithm results in a marginal, but tangible improvement of performance.
- The $topN$ and ST parameters appear to be important, as they visibly influence the results. Either by using one or both, the $topN$ and ST parameters provide great control over the IBOMbIE algorithm.

In the following section we compare the performance of IBOMbIE to other ontology matching algorithms to answer the final research question as formulated in Sect. 1.

6 Comparing with Other OM Algorithms

In this section we compare the IBOMbIE algorithm with several other OM algorithms. This section is based on evaluation efforts that have been carried out in the TELplus project [37] and in the context of the Ontology Alignment Evaluation Initiative,[19] which we introduce in Sect. 6.3.

6.1 Comparison with IBOM by Exact IM

As mentioned in Sect. 4 there is a substantial number of shared instances in the TELplus datasets. These instances can be used to generate an alignment by merging the annotations of shared instances and applying JC_c. Note that we cannot use the reindexing evaluation method to evaluate this alignment, because that would generate biased results (as

[19] http://oaei.ontologymatching.org/.

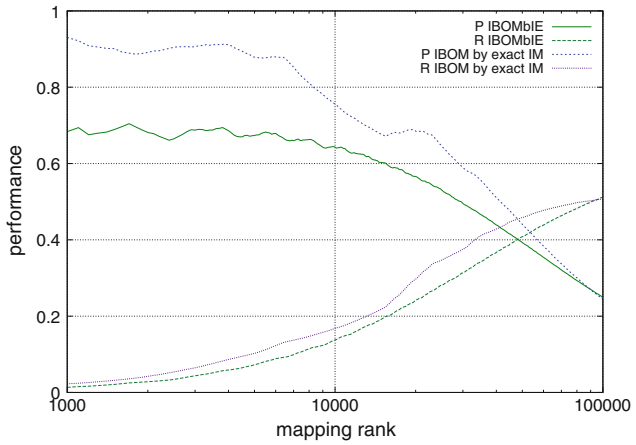


Fig. 14 IBOMbIE versus IBOM using exact IM: gold standard comparison

we would then use that same set of shared instances to both generate and evaluate the alignment).

The evaluation results of the alignments generated with IBOMbIE and exact matching are displayed in Fig. 14. We see that exact matching outperforms IBOMbIE in the early ranks. The superior performance of exact matching is not surprising since the concept associations of shared instances are more reliable. With an increasing number of mappings produced the quality becomes comparable. This indicates that the noise introduced through the instance enrichment process has a smaller impact when the similarity between concepts is small.

6.2 Comparison with a Lexical Matcher

As part of the TELplus project, we conducted experiments using a lexical OM algorithm based on the *CELEX* lexical database.[20] We applied this lexical matcher to (1) the vocabularies as they stand and (2) versions where concept labels were translated. The latter is done by querying the *Google Translate* web-service, translating the English concepts to French and vice versa. When a query is successful, the translated label is added to the concept.

The evaluation results of the alignments produced with the lexical matcher and the IBOMbIE baseline are shown in Figs. 15 and 16. The lexical matcher produces mappings with three different confidence levels, corresponding to different ways of involving lemmatization in the matching process. Mappings with the same confidence level are treated as having the same rank and hence the three horizontal lines in the evaluation results.

We first observe that the number of lexical mappings is greatly enhanced by translating the concept labels: without

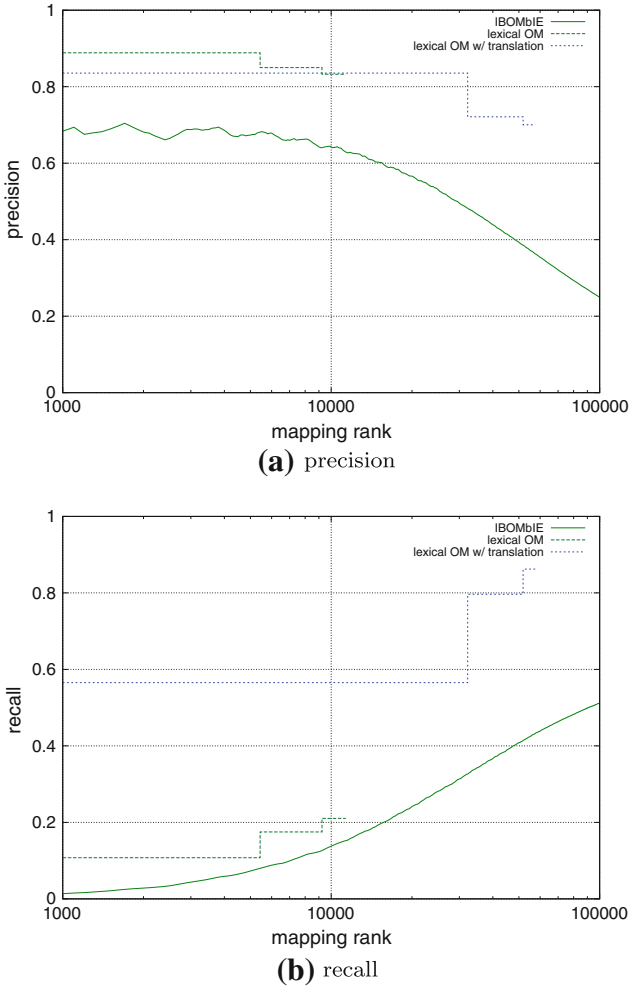


Fig. 15 IBOMbIE versus lexical OM algorithm: gold standard comparison

translation 11K mappings are produced, covering 13 % of the gold standard. Translating the concept labels increases the number of mappings generated by the lexical matcher to 58K, which covers 86 % of the gold standard.

We then observe a striking difference between the results obtained using the gold standard evaluation and the ones obtained from the reindex evaluation. This indicates that the alignment created by the lexical matcher is very similar to the gold standard. This is possibly due to the way experts discover and validate mappings, using lexical aids such as their own translation abilities, or dictionaries. Similarly, the precision of the lexical matcher without translation is strictly higher than that of the lexical matcher with translation in the gold standard evaluation (see Fig. 15a), but this holds vice versa in the reindex evaluation (see Fig. 16a). This discrepancy in the precision indicates that many mappings in the gold standard are lexically equal concept pairs, and concept pairs that are lexically similar after translation are not in the gold standard.

[20] http://celex.mpi.nl/.

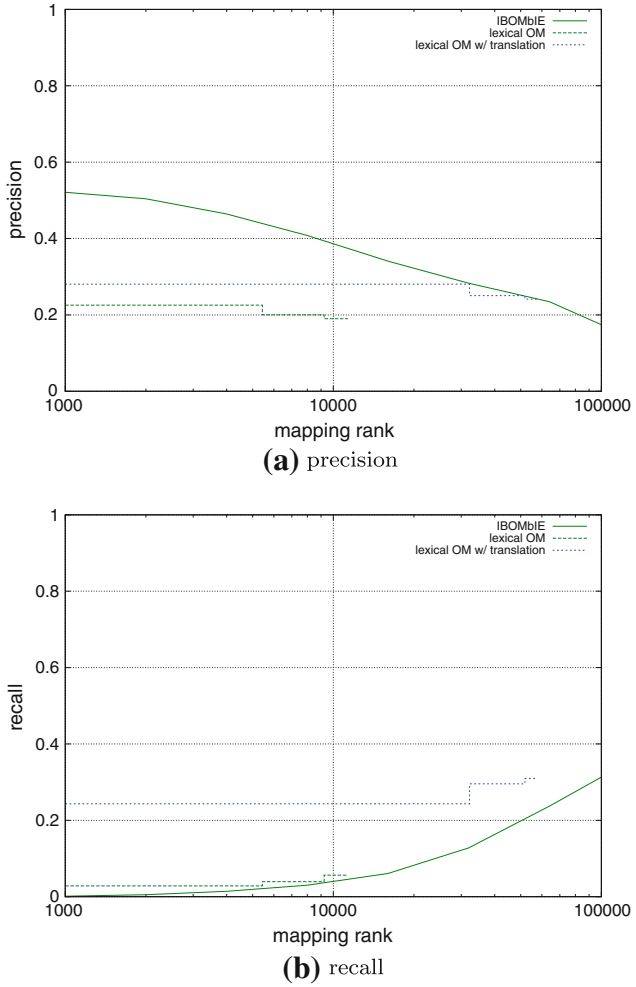


Fig. 16 IBOMbIE versus lexical OM algorithm: reindex

The results of the reindexing evaluation in Fig. 16 indicate that IBOMbIE outperforms both lexical matchers in terms of precision. As for the recall, the lexical matchers outperform IBOMbIE if we consider the ranks for which the lexical matchers produce mappings. However, IBOMbIE generates many more mappings, which—at the cost of precision—enables it to eventually achieve a higher recall than both lexical matchers.

Figure 17 shows the overlap of the alignments generated by IBOMbIE and the lexical matchers. We see that these overlaps does not exceed 17 %. This small overlap hints at that the two approaches are complementary to one another. In this application scenario, a hybrid approach is likely to outperform matchers that implement either an instance-based or a terminological method.

6.3 Comparison with OAEI Participants

The Ontology Alignment Evaluation Initiative is a yearly event where OM systems are evaluated in many tracks, such

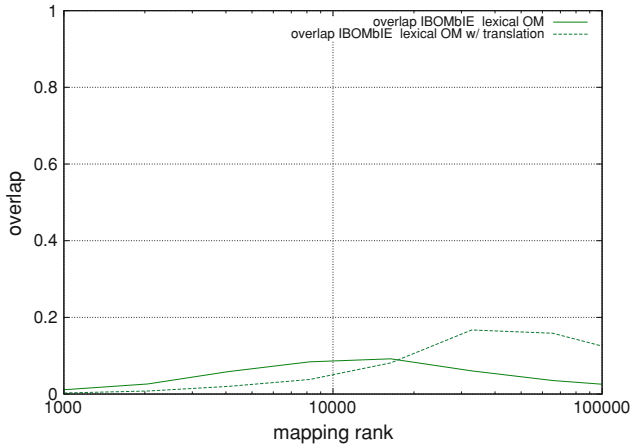


Fig. 17 Overlap IBOMbIE versus lexical OM algorithm with and without translation

Table 1 Run times for OAEI participants and IBOMbIE

Matcher	Tun time (h:min)
DSSim	12:00
Lily	Not included in OAEI report
TaxoMap	2:40
IBOMbIE	1:54

as the *Library* track. The 2008 edition of this track [3] focused on two large thesauri (resp., 5,000 and 5,000 concepts) from the *National Library of the Netherlands* (KB, which stands for *Koninklijke Bibliotheek*, (i.e) National Library). The KB track provides book instances—some of which dually annotated—enabling the application of IBOMbIE. In [26] we describe results of applying IBOMbIE to the KB scenario in detail. The OAEI 2008 Library track and the KB alignment scenario in [26] use the exact same vocabularies, rendering the alignments highly comparable (NB: the IBOMbIE experiments were conducted in 2009 and did not participate the official competition).

Three participants submitted results for the 2008 OAEI Library track: *DSSim*, *Lily* and *TaxoMap*. All three use terminological, structure-based and semantics-based techniques. Table 1 lists the run time of the OAEI participants and IBOMbIE.[21] We see that IBOMbIE is highly competitive in terms of run time, since it is faster than both DSSim and Taxomap.

Figure 18 compares the precision and the recall as obtained using the directional reindexing evaluation method [3]. The precision and recall of the OAEI contestants are constant, as the OAEI report provides single-valued evaluation results.

[21] The runtime of IBOMbIE is for a complete run—including the enrichment process. The run times in Table 1 were taken from the OAEI result reports of Taxomap [11] and DSSim [22]. The Lily OAEI result report [34] does not list the run-time.

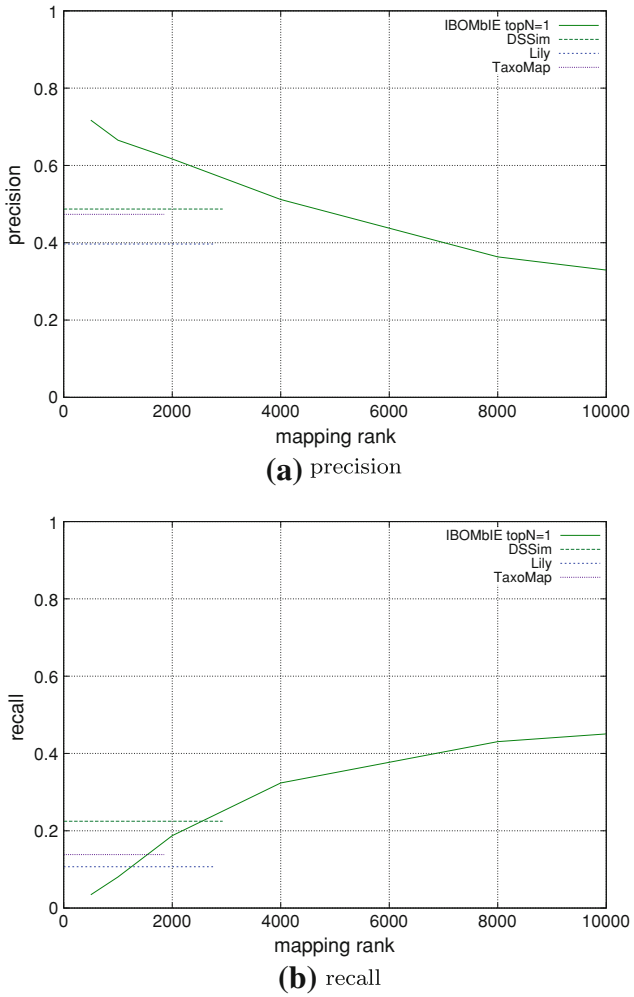


Fig. 18 Alignment quality of OAEI contestants and IBOMbIE baseline: Library track, reindexing evaluation

For any rank of mappings covered by the OAEI contestants, IBOMbIE has a higher precision than the OAEI participants. With respect to the recall, IBOMbIE performs better at the ranks corresponding to the number of mappings produced by each of the OAEI contestants, (i.e) 1,851, 2,797 and 2,930 mappings for TaxoMap, Lily and DSSim, respectively.

This comparison shows that in a library context, in which concepts have strong extensional semantics, the instance-based OM method work exceptionally well. The terminological, structure-based and semantics-based methods of the OAEI competitors perform relatively poor in this scenario, due to the non-English language and the flat taxonomy of the KB ontologies. In conclusion, the usefulness of instance-based OM in this particular application scenario shows that broadening the applicability of instance-based OM methods, as described in this paper, can be highly rewarding.

7 Conclusions

In this paper we describe and thoroughly investigate instance-based ontology matching by instance enrichment (IBOMbIE), a method which significantly expands the applicability of instance-based mappings to scenarios where no joint instances are available.

We identify several parameters, two of which influence the instance enrichment process and enable fine-grained control the selectiveness of the IBOMbIE algorithm. The effect of these parameters was evaluated using a real-life, large-scale and multi-lingual OM scenario in the Library domain. We have shown that simple word-by-word translation improves the results of the algorithm. Also, basing the IDF on the word distribution of both the indexed and the query dataset has a positive impact on performance. Furthermore, it turns out that refined instance enrichment methods do not significantly exceed the performance of a simple instance enrichment method.

The comparison with other OM algorithms shows that IBOMbIE is a promising OM method. The advantages of IBOM in general, such as the ability to deal with lexical ambiguity or the application in multi-lingual scenarios, make it lucrative to use IBOMbIE when many instances are available. The results of our experiments suggest that IBOMbIE is especially valuable as an approach that complements the mappings created by other techniques, such as terminological matching.

This paper presented an extensional ontology matching method that works in the absence of dually annotated corpora. The focus has been on technical aspects of the approach, and we had to restrict ourself to showing its usefulness in a specific case in the library domain rather than more generically. It will be interesting future work to apply and evaluate IBOMbIE in different applications and different domains.

References

1. Avesani P, Giunchiglia F, Yatskevich M (2005) A large scale taxonomy mapping evaluation. In: Gil Y, Motta E, Benjamins VR, Musen MA (eds) International semantic web conference. Lecture notes in computer science, vol 3729. Springer, Berlin, pp 67–81. http://dblp.uni-trier.de/rec/bibtex/conf/swws/TodorovG09
2. Buckland M, Gey F (1994) The relationship between recall and precision. J Am Soc Inf Sci 45(1):12–19. http://www.bibsonomy.org/bibtex/1f75b35ab969ab89391cf6cbd2176ca67/dblp
3. Caracciolo.C, Euzenat.J, Hollink L et al (2008) Results of the ontology alignment evaluation initiative 2008. In: Proceedings of the 3rd international workshop on ontology matching, collocated with the 7th international semantic web conference (ISWC). http://ceur-ws.org/Vol-431

4. Choi N, Song IY, Han H (2006) A survey on ontology mapping. SIGMOD Record 35(3):34–41. http://dblp.uni-trier.de/db/journals/sigmod/sigmod35.html#ChoiSH06
5. Doan A, Domingos P, Halevy AY (2003) Learning to match the schemas of data sources: a multistrategy approach. Mach Learn 50(3):279–301. http://dblp.uni-trier.de/db/journals/ml/ml50.html#DoanDH03
6. Doan A, Madhavan J, Domingos P, Halevy A (2004) Ontology matching: a machine learning approach. In: Handbook on ontologies in information systems. Springer, Berlin, pp 397–416
7. Euzenat J, Meilicke C, Stuckenschmidt H, Shvaiko P, dos Santos CT (2011) Ontology alignment evaluation initiative: six years of experience. J Data Semant 15:158–192
8. Euzenat J, Shvaiko P (2007) Ontology matching. Springer-Verlag, Heidelberg (DE), p 341. ISBN 3-540-49611-4
9. Euzenat J, Valtchev P (2004) Similarity-based ontology alignment in owl-lite. In: de Mántaras RL, Saitta L (eds) ECAI. IOS Press, Amsterdam, pp 333–337
10. Ferrara A, Nikolov A, Scharffe F (2011) Data linking for the semantic web. Int J Semant Web Inf Syst 7(3):46–76
11. Hamdi F, Zargayouna H, Safar B, Reynaud C (2008) Taxomap in the oaei alignment contest. In: Shvaiko P, Euzenat J, Giunchiglia F, Stuckenschmidt H (eds) OM, CEUR workshop proceedings, vol 431. CEUR-WS.org. http://dblp.uni-trier.de/db/conf/semweb/om2008.html#HamdiZSR08
12. Hoshiai T, Yamane Y, Nakamura D, Tsuda H (2004) A semantic category matching approach to ontology alignment. In: Proceedings of the third international workshop on evaluation of ontology-based tools (EON)
13. Ichise R, Takeda H, Honiden S (2003) Integrating multiple internet directories by instance-based learning. In: Proceedings of the eighteenth international joint conference on artificial intelligence
14. Isaac A, Matthezing H, van der Meij L, Schlobach S, Wang S, Zinn C (2008) Putting ontology alignment in context: usage scenarios, deployment and evaluation in a library case. In: Hauswirth M, Koubarakis M, Bechhofer S (eds) Proceedings of the 5th European semantic web conference, LNCS. Springer, Berlin. http://data.semanticweb.org/conference/eswc/2008/papers/188
15. Isaac A, vann der Meij L, Schlobach S, Wang S (2007) An empirical study of instance-based ontology matching. In: ISWC/ASWC, pp 253–266
16. Isaac A, Wang S, Zinn C, Matthezing H, van der Meij L, Schlobach S (2009) Evaluating thesaurus alignments for semantic interoperability in the library domain. IEEE Intell Syst 24(2):76–86
17. Kopcke H, Rahm E (2010) Frameworks for entity matching: a comparison. Data Knowl Eng 69:197–210. doi:10.1016/j.datak.2009.10.003
18. Leme LAPP, Casanova MA, Breitman KK, Furtado AL (2009) Instance-based owl schema matching. In: Filipe J, Cordeiro J (eds) Enterprise information systems, Proceedings of 11th international conference, ICEIS 2009, Milan, May 6–10. Lecture notes in business information processing, vol 24. Springer, Berlin, pp 14–26. doi:10.1007/978-3-642-01347-8_2
19. Li J, Tang J, Li Y, Luo Q (2009) Rimom: a dynamic multistrategy ontology alignment framework. IEEE Trans Knowl Eng 21:1218–1232. doi:10.1109/TKDE.2008.202
20. Li WS, Clifton C, Liu SY (2000) Database integration using neural networks: implementation and experiences. Knowl Inf Syst 2:73–96
21. Maedche A, Motik B, Silva N, Volz R (2002) Mafra—a mapping framework for distributed ontologies. In: Gmez-Prez A, Benjamins VR (eds) EKAW. Lecture notes in computer science, vol 2473. Springer, Berlin, pp 235–250. http://dblp.uni-trier.de/db/conf/ekaw/ekaw2002.html#MaedcheMSV02
22. Nagy M, Vargas-Vera M, Stolarski P, Motta E (2008) DSSim results for OAEI 2008. http://ceur-ws.org/Vol-431/oaei08_paper5.pdf
23. Rahm E (2011) Towards large-scale schema and ontology matching. ReCALL 5:1–26. http://www.springerlink.com/index/M5055K8721752228.pdf
24. Rahm E, Bernstein PA (2001) A survey of approaches to automatic schema matching. VLDB J 10(4):334–350
25. Salton G, Wong A, Yang CS (1975) A vector space model for automatic indexing. Commun ACM 18(11):613–620. doi:10.1145/361219.361220
26. Schopman B (2009) Instance-based ontology matching by instance enrichment. Master's thesis, Vrije Universiteit, The Netherlands
27. Schopman B, Wang S, Schlobach S (2008) Deriving concept mappings through instance mappings. In: ASWC
28. Spero SE (2011) What, if anything, is a subdivision? In: International Society for Knowledge Organisation
29. Stumme G, Maedche A (2001) Fca-merge: bottom-up merging of ontologies. In: Proceedings of the 17th international conference on artificial intelligence (IJCAI '01), Seattle, pp 225–230
30. Thor A, Kirsten T, Rahm E (2007) Instance-based matching of hierarchical ontologies. In: Kemper A, Schning H, Rose T, Jarke M, Seidl T, Quix C, Brochhaus C (eds) BTW, LNI, GI, vol 103, pp 436–448. http://dblp.uni-trier.de/db/conf/btw/btw2007.html#ThorKR07
31. Todorov K, Geibel P (2009) Variable selection as an instance-based ontology mapping strategy. In: Arabnia HR, Marsh A (eds) SWWS. CSREA Press, USA, pp 3–9
32. Todorov K, Geibel P, Kuhnberger KU (2010) Mining concept similarities for heterogeneous ontologies. In: Perner P (ed) ICDM. Lecture notes in computer science, vol 6171. Springer, Berlin, pp 86–100
33. Udrea O, Getoor L, Miller RJ (2007) Leveraging data and structure in ontology integration. In: Proceedings of the ACM SIGMOD international conference on management of data, Beijing, June 12–14, pp 449–460
34. Wang P, Xu B (2008) Lily: ontology alignment results for OAEI 2008. In: Shvaiko P, Euzenat J, Giunchiglia F, Stuckenschmidt H (eds) OM, Proceedings of the 3rd international workshop on ontology matching (OM-2008), collocated with the 7th international semantic web conference (ISWC-2008), Karlsruhe, Germany, 26 October 2008, vol 431. CEUR-WS.org. http://ceur-ws.org/Vol-431/oaei08_paper7.pdf
35. Wang S, Englebienne G, Schlobach S (2008) Learning concept mappings from instance similarity. In: International semantic web conference, pp 339–355
36. Wang S, Isaac A, Schlobach S, van der Meij L, Schopman BAC (2012) Instance-based semantic interoperability in the cultural heritage. Semantic Web 3(1):45–64
37. Wang S, Isaac A Schopman B, Schlobach S, van der Meij L (2009) Matching multi-lingual subject vocabularies. In: Proceedings of the 13th European Conference on Digital Libraries (ECDL2009)
38. Wartena C, Brussee R (2008) Instanced-based mapping between thesauri and folksonomies. In: ISWC'08
39. Zaiss KS (2010) Instance-based ontology matching and the evaluation of matching systems. Ph.D. thesis, Heinrich Heine Universität Düsseldorf