

## RESEARCH ARTICLE

# Toward the Multilingual Semantic Web: Multilingual Ontology Matching and Assessment

SHIMAA IBRAHIM<sup>1,2</sup>, SAID FATHALLA<sup>3</sup>, JENS LEHMANN<sup>4,5</sup>, AND HAJIRA JABEEN<sup>6</sup><sup>1</sup>Institute of Computer Science, University of Bonn, 53115 Bonn, Germany<sup>2</sup>Institute of Graduate Studies and Research, University of Alexandria, Alexandria 5422004, Egypt<sup>3</sup>Faculty of Science, University of Alexandria, Alexandria 5413213, Egypt<sup>4</sup>Institute for Applied Informatics Association e.V. (InfAI), 01069 Dresden, Germany<sup>5</sup>Amazon, 01097 Dresden, Germany (work done outside of Amazon)<sup>6</sup>Big Data Analytics, KTS, GESIS-Leibniz Institute for the Social Sciences, 50667 Cologne, Germany

Corresponding author: Shimaa Ibrahim (ibrahim@cs.uni-bonn.de)

**ABSTRACT** The amount of multilingual data on the Web proliferates; therefore, developing ontologies in various natural languages is attracting considerable attention. In order to achieve semantic interoperability for the multilingual Web, cross-lingual ontology matching techniques are highly required. This paper proposes a Multilingual Ontology Matching (MoMatch) approach for matching ontologies in different natural languages. MoMatch uses machine translation and various string similarity techniques to identify correspondences across different ontologies. Furthermore, we propose a Quality Assessment Suite for Ontologies (QASO) that comprises 14 metrics, out of which seven metrics are used to assess the quality of the matching process and seven metrics are used to evaluate the quality of the ontology. We present an in-depth comparison of different string similarity techniques across various languages to get the most effective similarity measure(s) between multilingual terms. To illustrate the applicability of our approach and how it can be used in different domains, we present two use cases. MoMatch has been implemented using Scala and Apache Spark under an open-source license. We have compared our results with the results from the Ontology Alignment Evaluation Initiative (OAEI 2020). MoMatch has achieved significantly high precision, recall, and F-measure compared to five state-of-the-art matching approaches.

**INDEX TERMS** Cross-lingual matching, knowledge management, multilingual web, ontology engineering, ontology matching, string similarity.

## I. INTRODUCTION

Ontologies are being widely used in various fields of science other than computer science, including Biology [1], Engineering [2], and Medicine [3]. With the rapid expansion of multilingual data on the Semantic Web, more ontologies have become available in different natural languages. According to Linked Open Vocabularies (LOV),<sup>1</sup> English is by far the most prominent language, i.e., most ontologies in the Semantic Web are in English; however, many ontologies in other Indo-European languages (e.g., German) also exist. Specifically, out of a total of 782 vocabularies found in LOV, 584 are in

English, 65 in French, and 41 in German. Few ontologies registered in LOV are in non-Indo-European languages, e.g., ten are in Chinese, and seven are in Arabic. In fact, monolingual ontologies that include labels or local names presented in a particular language are not understandable to people who speak other languages, hindering semantic interoperability. Therefore, to improve semantic interoperability between monolingual ontologies, innovative algorithms need to be developed to digest multilingual data and match ontologies in different natural languages [4]. Identifying correspondences between ontologies in different natural languages is called *Cross-lingual ontology matching*.

Cross-lingual ontology matching plays a crucial role in various research areas involving ontology enrichment/merging, data warehouses, linked data, peer-to-peer information

The associate editor coordinating the review of this manuscript and approving it for publication was Mansoor Ahmed<sup>id</sup>.

<sup>1</sup><https://lov.linkeddata.es/dataset/lov/vocabs>

sharing, and web service composition [5]. The main idea is that all ontologies should be mapped to a core set of related ontologies in advance when a system is attempting to automatically search for a specific piece of information from or exchange knowledge with other systems.

To date, there is no clear winner in solving the cross-lingual matching problem. Further research is required to advance cross-lingual ontology matching techniques to obtain better results compared to monolingual approaches. Therefore, we can conclude that there is a need to create a flexible framework for matching ontologies in different languages without relying on a specific language.

In this paper, we propose a Multilingual Ontology Matching (MoMatch) approach for matching ontologies in different natural languages. MoMatch uses different string similarity techniques and machine translation to match classes and properties across ontologies. MoMatch comprises four phases: 1) *resource extraction*: extracting all resources (classes and properties) from the input ontologies, 2) *pre-processing*: preparing the extracted resource for the subsequent phases, 3) *translation*: translating the extracted resources, and 4) *matching*: identifying potential matches between the input ontologies.

The quality of the matching process greatly depends on the quality of the input ontologies to be matched. Existing approaches defined in the literature have addressed the quality evaluation of ontology from different dimensions, such as syntactic, semantic, pragmatic, social, structural, functional, content, schema, and usage [6], [7], [8], [9]. Each dimension has its own set of criteria, and each criterion has its own set of metrics to measure the characteristics of an ontology that can be represented formally [10]. A vast number of metrics have been addressed from different aspects. As a result, determining what quality metrics that affect the quality of the ontology matching process is complex. In our previous work MULON [11], an approach for merging monolingual ontologies in different natural languages, we proposed seven quality metrics to assess the quality of ontologies on the schema level. In this paper, we extend the quality metrics, by proposing the Quality Assessment Suite for Ontologies (QASO). QASO comprises 14 quality metrics for assessing the quality of the matching process in addition to the quality of the input ontologies' schema that influences the matching process (more details in subsection V-A). Out of the 14 metrics, seven are used to assess the quality of the input ontologies and seven to evaluate the quality of the matching process. We address the following research questions: *RQ1) What is the most effective string similarity measure(s) for matching multilingual terms across different ontologies? RQ2) How strong is the performance of the similarity measure when the language of the input ontologies is changed? and RQ3) How can the quality of the ontology matching process be measured using a set of metrics?*

The contributions of this work can be summarized in the following points:

- MoMatch can efficiently match ontologies in any natural language compared to state-of-the-art (cf. subsection VII-C),
- Ten language pairs, including Indo and Non-Indo European languages, have been tested in this approach,
- We present a comparative analysis of 13 different string similarity measures,
- A metric suit (QASO) has been designed for assessing the quality of the matching process, which also can be used to assess the quality of the ontology,
- Two use cases demonstrating the usability of the MoMatch in matching multilingual as well as monolingual ontologies from different domains are presented, and
- MoMatch empirically showed significantly better performance when compared to five state-of-the-art approaches.

We believe that MoMatch is a crucial step towards realizing the multilingual Semantic Web as it supports the integration of ontologies in different languages. MoMatch<sup>2</sup> and QASO<sup>3</sup> are available in two separate public repositories in GitHub, in which the source code is documented, describing each configurable parameter and function.

The rest of this paper is organized as follows: In section II, we provide background information about our research topic as well as formal definitions of ontology and the ontology matching process. In section III, we provide an overview of related work. The proposed approach is described in detail in section IV. The quality metric suite is presented in section V. Two use cases are presented in section VI to demonstrate possible applications of MoMatch. The results of experiments and evaluations are presented in section VII. Finally, in section VIII, we summarize the main conclusions with an outline of future work.

## II. PRELIMINARIES AND DEFINITIONS

Ontology is considered one of the significant cornerstones of representing information more meaningfully, providing machine-understandable semantics of knowledge. It is a container for capturing semantic information of a particular domain, which allows sharing and reusing knowledge in this domain [17], [18].

*Definition 1 (Ontology)*: Formally, an ontology can be represented as a tuple of five components [19]:  $O = \langle C, R, P, I, X \rangle$ , where  $C$  is the set of classes/concepts,  $R$  is the set of relationships between classes (object properties),  $P$  is the set of data properties (a specific type of relation whose domain is a class and the range is a data type),  $I$  is the set of class instances (concrete objects), and  $X$  is the set of axioms and rules used for checking and verifying ontology consistency and new knowledge inference.

<sup>2</sup><https://github.com/SmartDataAnalytics/MoMatch>

<sup>3</sup><https://github.com/SmartDataAnalytics/QASO>

**TABLE 1.** An overview of eight sequence-based string similarity measures utilized in MoMatch, as well as their formulas.

Name	Description	Formula
Hamming distance	Measures the minimum number of substitutions required to change one string $x$ into the other $y$ [12].	The hamming distance normalized by the length of the longest string is given by [5]: $Hamm(x, y) = \frac{(\sum_{i=1}^{\min( x ,  y )} x[i] \neq y[i]) +   x  -  y  }{\max( x ,  y )}$ (1)
Jaro	Counts the number of common characters $c$ between two strings $x, y$ and the number of transpositions $t$ required to make these common characters have the same sequence [5], [13].	$Jaro(x, y) = \frac{1}{3} \times (\frac{c}{ x } + \frac{c}{ y } + \frac{c - \frac{t}{2}}{c})$ (2)
Jaro Winkler	Captures cases where two strings $x, y$ have a low Jaro score but share a prefix. It enhances Jaro by including the length $l$ of the longest common prefix and a weight $w$ given to that prefix [5], [13].	$JW(x, y) = (1 - l \times w) \times Jaro(x, y) + l \times w$ (3)
Levenshtein	Counts the minimum number of single-character edits (insertions, deletions, or substitutions) required to transform one string $x$ into the other $y$ [5], [14]. Also known as Edit Distance.	$Lev(x, y) = 1 - \frac{\delta(x, y)}{\max(\text{length}(x), \text{length}(y))}$ (4) where $\delta(x, y)$ is the edit distance between strings $x, y$ .
Ratio	Calculates the ratio of the number of matching characters $m$ in the two strings $x, y$ to the total number of characters $t$ in both strings [15].	$R = \text{round}(2.0 \times \frac{m}{t} \times 100)$ (5)
Partial ratio	Finds the similarity between the shorter string $x$ and every sub-string $y_i$ of length $m$ of the longer string $y$ , and returns the maximum of that similarity measures [15].	$PR = \max(\sum_{i=1}^n \text{ratio}(x, y_i))$ (6) where $n$ is the number of sub-strings.
Token sort	Splits the two strings $x, y$ into tokens, sorting the tokens, and calculates the ratio similarity of the transformed strings [16].	
Partial token sort	Splits the two strings $x, y$ into tokens, sorting the tokens, and calculates the partial ratio similarity of the transformed strings [16].	

Ontology matching is a complicated procedure to bridge the semantic gap between multiple representations of the same domain [20]. It identifies correspondences between the entities (classes and properties) of two or more ontologies that satisfy specific conditions.

**Definition 2 (Ontology Matching):** “a function  $f$  which, from a pair of ontologies to match  $O_1$  and  $O_2$ , an input alignment  $A$ , a set of parameters  $p$  and a set of resources  $r$ , returns an alignment  $A'$  between these ontologies:  $A' = f(O_1, O_2, A, p, r)$ ” [5].

**Definition 3 (Alignment):** the alignment between two ontologies,  $O_1$  and  $O_2$ , is a set of correspondences between pairs of entities belonging to  $O_1$  and  $O_2$ , respectively [5]. It is the output of the matching process.

**Definition 4 (Correspondence):** given two ontologies  $O_1$  and  $O_2$ , the correspondence between two entities  $e_{11}$  and  $e_{12}$ , where  $e_{11} \in O_1$  and  $e_{12} \in O_2$ , is the relation that produces according to a matching algorithm between  $e_{11}$  and  $e_{12}$  [5]. It can be represented by the triple:  $\langle e_{11}, e_{12}, r \rangle$ , where  $r$  is the relation between the two entities  $e_{11}$  and  $e_{12}$ .

**Definition 5 (Monolingual ontology matching):** the process of matching ontologies in the same natural language, i.e.,  $L_1 = L_2$ , where  $L_1$  is the natural language of  $O_1$ , and  $L_2$  is the natural language of  $O_2$  [4].

The growing amount of multilingual data on the Web and the resulting development of ontologies in different natural languages has increased the demand for cross-lingual ontology matching.

**Definition 6 (Cross-lingual ontology matching):** the process of matching ontologies in different natural languages, i.e.,  $L_1 \neq L_2$  [4].

Ontology matching techniques are used to identify the correspondence between two ontologies’ entities, including the analysis of subsumption between classes and the similarity between the entity names. Different ontology matching techniques have been proposed, which can be classified as [20]: a) Element-level matching techniques: identify the correspondences by analyzing entities in the ontologies in isolation, ignoring the structure, i.e., ignoring their relations with other entities or instances. b) Structure-level matching techniques: identify the correspondences by analyzing the structure of entities in the ontology, i.e., considering the relations between entities and their instances. In this paper, we propose an element-level matching technique.

String-based techniques (as element-level techniques) are utilized to match names and name descriptions of ontology entities [5]. These techniques consider strings as alphabetical sequences of letters. The more similar the strings are, the

**TABLE 2.** An overview of five set-based string similarity measures utilized in MoMatch, as well as their formulas. The input is two sets  $X$  and  $Y$ , each containing a set of tokens for the two strings  $x$  and  $y$ , respectively. All the resulting values are in the range of  $[0,1]$ .

Name	Description	Formula
Cosine	Computes a variant of cosine measure known as Ochiai coefficient, which calculates the intersection between two sets $X, Y$ <sup>4</sup> [16].	$\text{Cosine}(X, Y) = \frac{ X \cap Y }{\sqrt{ X  \cdot  Y }} \quad (7)$
Dice	Twice the common terms (intersection) between two strings over the total number of terms in both strings <sup>4</sup> [16].	$\text{Dice}(X, Y) = \frac{2 \times  X \cap Y }{ X  +  Y } \quad (8)$
Jaccard	The size of the common terms between two strings divided by the size of the union of all terms in the two strings <sup>4</sup> [16].	$\text{Jaccard}(X, Y) = \frac{ X \cap Y }{ X \cup Y } \quad (9)$
Overlap coefficient	Measures the overlap between two sets by dividing the intersection between them by the smaller sizes of the two sets <sup>4</sup> [21]. If one string is a subset of another, it is considered a full match.	$\text{OC}(X, Y) = \frac{ X \cap Y }{\min( X ,  Y )} \quad (10)$
Tversky Index	The size of the common tokens between two sets divided by: the sum of intersection between sets, the number of items only available on the first token set multiplied by a coefficient $\alpha$ , and the number of elements only available in the second token sets multiplied by a coefficient $\beta$ [22].	$T(X, Y) = \frac{ X \cap Y }{ X \cap Y  + \alpha X - Y  + \beta Y - X } \quad (11)$ where $\alpha$ and $\beta > 0$ .

more likely they represent the same concepts. Inspired by `py_stringmatching` library<sup>4</sup> and Doan et al. [13] string-based techniques are categorized as follows:

- **Sequence-based:** Input strings are considered as a sequence of characters. Such as Hamming distance, Jaro, Jaro Winkler, Levenshtein, Partial ratio, Partial token sort, Ratio, and Token sort.
- **Set-based:** Input strings are considered as sets or multi-sets of tokens. Such as Cosine, Dice, Jaccard, Overlap coefficient, and Tversky Index.
- **Bag-based:** Input strings are considered as bags, i.e., collections of tokens, in which a token appears multiple times. Such as TF/IDF.
- **Hybrid-based:** It combines sequence-based and token-based similarity measures. Such as generalized Jaccard, soft TF/IDF, and Monge-Elkan.
- **Phonetic-based:** It matches strings based on their sound instead of appearances. This similarity measure effectively matches names, which often appear in different ways that sound the same, such as Meyer, Meier, and Mire; Smith, Smithe, and Smythe. Such as Soundex.

MoMatch utilizes a set of string similarity measures to search for the most effective string similarity measures that can be used to find identical matches between entities. The utilized string-based techniques are described in Table 1 and Table 2 with their formulas.

### III. RELATED WORK

According to a recent review of the multilingual Web of Data literature, fewer researchers have addressed cross-lingual

ontology matching [23]. Cross-lingual ontology matching techniques are mainly used for matching linguistic information across ontologies in different natural languages [4], [23].

#### A. CROSS-LINGUAL MATCHING APPROACHES

Abu Helou and Palmonari [24] proposed a cross-lingual lexical matching technique to map lexically-rich language resources such as WordNet. The results of word translations are used as evidence to map concepts lexicalized in different languages. Google Translate and BabelNet are used as external resources for translation. Four language versions of WordNet (Arabic, Italian, Slovene, and Spanish) are mapped to the English WordNet. Musyaffa et al. [25] proposed a framework for interlinking heterogeneous multilingual open fiscal data. Machine translation and similarity measures are used to map concepts across different languages.

Fu and Brennan [26], [27] proposed an algorithm for matching English and Chinese ontologies that takes into account the target ontology's semantics, mapping intent, operating domain, time and resource constraints, and user feedback. Hertling and Paulheim [28] proposed an approach for finding corresponding ontology elements that makes use of Wikipedia's interlanguage links. Lin and Krizhanovsky [29] proposed an approach that uses Wiktionary<sup>5</sup> as a resource for background information to match English and French ontologies.

Trojahn et al. [30] proposed a multi-agent architecture for cross-lingual ontology matching. The translation agent translates the source ontology into the target language using a dictionary. Lexical databases and thesaurus have been used

<sup>4</sup>[http://anhaidgroup.github.io/py\\_stringmatching/v0.4.x/index.html](http://anhaidgroup.github.io/py_stringmatching/v0.4.x/index.html)

<sup>5</sup><https://www.wiktionary.org/>



for identifying mappings. Tigrine et al. [31] presented an approach for matching several ontologies in different natural languages that uses the multilingual semantic network BabelNet<sup>6</sup> as a source of background information.

Ali et al. [32] proposed a multi-agent architecture-based cross-lingual ontology enrichment approach to enrich ontologies from multilingual text or ontologies. Ibrahim et al. [33], [34] proposed a fully automated ontology enrichment approach based on cross-lingual matching that creates a multilingual ontology by enriching a monolingual one from another in a different language. They used lexical similarity (Jaccard) and semantic similarity (based on WordNet) to filter the equivalent classed. All translations produced by Google Translate for each class are considered during the matching process. They proposed another approach for merging monolingual ontologies in various natural languages to produce a multilingual ontology [11]. First, the alignments between input ontologies are identified using the cross-lingual matching techniques, then adds them to the merged multilingual ontology by adding `rdfs:label` for each language (using language-tagged strings). Offline dictionaries have been built using Yandex translate API for concept translations, and lexical (Jaccard) and semantic similarity are used to find the alignment between input ontologies.

SimCat [35] and Crolom [36] proposed a lexical matching technique that uses Yandex translator and WordNet to compute the semantic similarity between concepts. They first apply NLP techniques to normalize ontology entities, translate all entities, and compute the similarity between them. Such order affects the quality of the translation because NLP techniques eliminate and normalize words, which can greatly affect the translation quality and significantly reduce the alignment quality.

#### B. CROSS-LINGUAL MATCHING TOOLS IN OAEI 2020

In the context of OAEI 2020<sup>7</sup> results of the ontology alignment evaluation initiative 2020 campaign for evaluating ontology matching technologies, VeeAlign [37], AML [38], LogMap [39], and Wiktionary Matcher [40] provide high-quality alignments for the cross-lingual matching task. These approaches heavily rely on the lexical matching technique, except VeeAlign [37], which discovers alignments using a supervised deep learning approach. VeeAlign [37] proposes a two-step model which uses contextualized representations of concepts to discover alignments based on semantic and structural aspects of an ontology. AML [38] is based on lexical and structural matching algorithms. It utilizes background knowledge and machine translation tools, such as Microsoft Translator, before starting the matching process. LogMap [39] implements optimized data structures for lexically and structurally indexing the input ontologies for the matching process. It is an iterative process that begins with initial mappings (i.e., almost exact lexical correspondences) and proceeds to

the discovery of new mappings. Therefore, LogMap cannot find matches between ontologies that lack sufficient lexical information. LogMapLt [39] is a “lightweight” version of LogMap, which only applies string matching techniques. Wiktionary Matcher [40] is an element-level, label-based matcher which uses multiple language versions of Wiktionary as an external background knowledge source. Meilicke et al. [41] developed a benchmark dataset (MultiFarm) based on manual translations of a set of ontologies from the conference domain into eight natural languages. This dataset is widely utilized to evaluate cross-lingual matching approaches [37], [38], [39], [40]. Good literature on the state-of-the-art approaches in cross-lingual ontology matching is provided in [4].

Despite ongoing efforts to develop various techniques, no clear winner has emerged in the solution to the cross-lingual matching problem. Further investigations are still needed to advance cross-language ontology matching techniques not only to obtain good results but also to assess them.

### IV. MOMATCH: THE PROPOSED APPROACH

In this section, we will describe in detail the four phases of our approach (cf. Figure 1). The input consists of two ontologies  $O_1$  and  $O_2$ , which can be in two different natural languages (for example,  $L_1 = fr$  and  $L_2 = de$ ) or in the same language. The output is the alignment between the input ontologies in addition to the assessment sheet for the input ontologies and the resultant alignment. In the following subsections, we describe each of these phases in detail.

#### A. RESOURCE EXTRACTION

This phase aims to extract all resources (including both classes and properties) from the two input ontologies and store them in the resources matrix  $R$ .  $R$  is a two-dimensional matrix ( $n \times 6$ ), where  $n$  is the number of extracted resources. Each row in  $R$  is represented as a tuple of  $\langle \text{resource, type, source, language, translation, pre-processed translation} \rangle$ , which contains the resource label, the type of the resource ('C' for a class and 'P' for a property), the source ontology of this resource, the language tag, the translation of this resource, and the pre-processed translation. Currently, the translation and the pre-processed translation for each resource will be NULL and will be assigned later in the following phases. For more illustration, consider the tuple  $\langle \text{Comité de programme, C, } O_1, fr, \text{ NULL, NULL} \rangle$ , in which the resource “Comité de programme” is a class extracted from  $O_1$  and its language is French. Similarly, the tuple  $\langle \text{ist geschrieben von, P, } O_2, de, \text{ NULL, NULL} \rangle$ , in which the resource “ist geschrieben von” is a property extracted from  $O_2$  and its language is German. The output of this phase is  $R$ .

#### B. TRANSLATION

In order to match resources from different languages, all of them should be translated into a common natural language.

<sup>6</sup><https://babelnet.org/>

<sup>7</sup><http://oaei.ontologymatching.org/2020/results/multifarm/index.html>

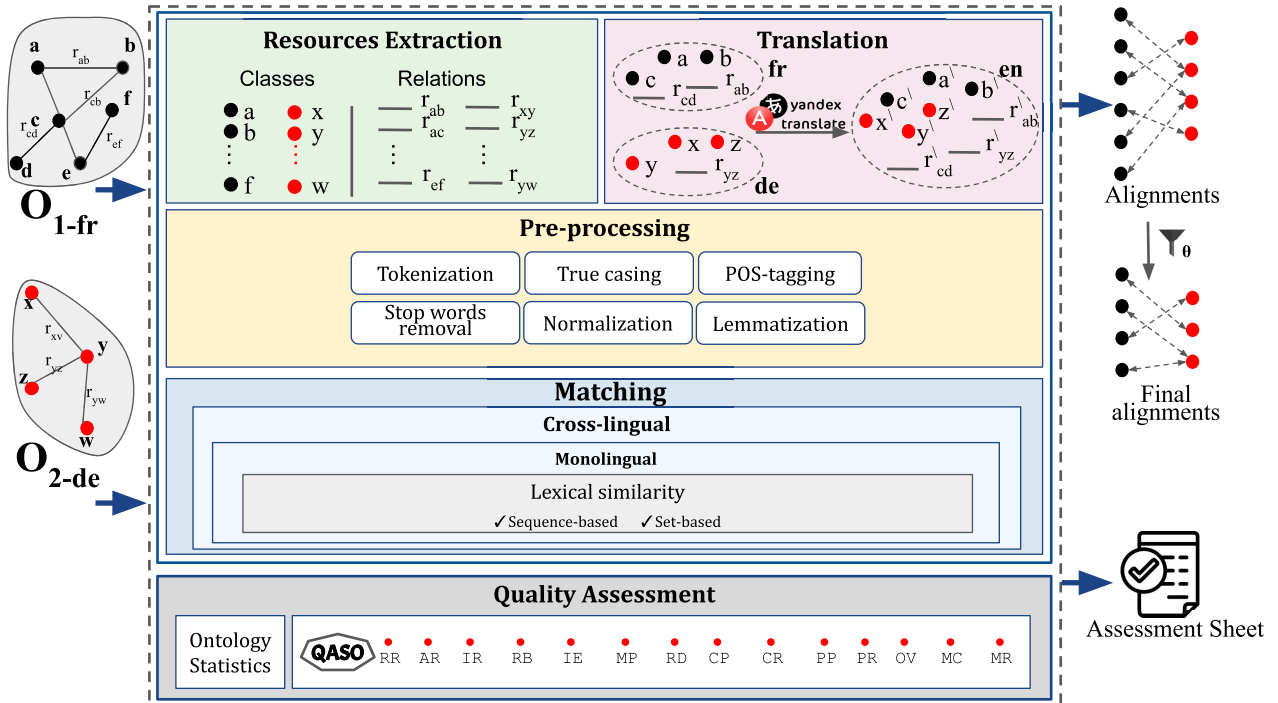


FIGURE 1. MoMatch architecture.

Two translation paths could be followed. The first path is to translate one ontology’s resources into the other’s language. For example, if the input ontologies are in German and French, then either the resources of the German ontology are translated to French or the resources of the French ontology are translated to German. The second path is to translate both of the input ontologies into a chosen pivot language, like English, for example. In our approach, we chose the second path because we found that most of the Natural Language Processing (NLP) techniques for English text outperform others in other languages. Furthermore, most of the external knowledge sources used in the matching process, such as dictionaries and thesaurus, are available in English.

According to Abu Helou et al. [42], machine translation tools can return proper translations for a very large number of resources in the cross-lingual matching task. We use Yandex translate API<sup>8</sup> to translate each resource in the resource matrix  $R$  and add the resource’s translation to  $R$ . For example, consider the same tuple from the previous example in subsection IV-A (Comité de programme,  $C$ ,  $O_1$ , fr, Program Committee, NULL), in which the English translation “Program Committee” is attached to the French class “Comité de programme”. The output of this phase is  $R$  with the translations of all resources.

C. PRE-PROCESSING

This phase aims to clean and prepare the translated resources for the matching phase by employing the following NLP techniques.

<sup>8</sup><https://tech.yandex.com/translate/>

- *Tokenization*: divides the resource name into a set of tokens. Tokens are separated by delimiters such as whitespace characters.
- *True casing*: recognizes resources with camel cases and adds a space between lower-case and upper-case letters such as “BestPosterAward” and “isSponsorOf” became “Best Poster Award” and “is Sponsor Of” respectively.
- *POS-tagging*: classifies tokens into their parts of speech, depending on their definition and context, then assigns them a special label or tag (such as Adjective (ADJ), Conjunction (CONJ), Verb (V), and Preposition (PREP)) accordingly.
- *Stop words removal*: removes tokens with high frequencies of occurrence and have no contribution to the subject of a text, such as pronouns, prepositions, and conjunctions.
- *Normalization and regular expressions*: transforms a token into a standard form by removing non-alphanumeric characters and additional white spaces.
- *Lemmatization*: removes inflections in a token and maps it to its root form. For example, “learning” is mapped to “learn” and “played” to “play”.

The output of this phase is  $R$  with the pre-processed translated resources.

D. MATCHING

This phase aims to identify correspondences (matched resources) between the two input ontologies. We perform a pairwise lexical similarity between the translated pre-

**TABLE 3. Ontology statistics metrics in MoMatch.**

Metric	Description
$getNumTrip(O)$	Get the number of triples in the ontology.
$getNumRes(O)$	Get the number of all resources (classes and properties) in the ontology.
$getNumbClass(O)$	Get the number of classes in the ontology.
$getNumProp(O)$	Get the number of all properties in the ontology.
$getNumObjProp(O)$	Get the number of object properties in the ontology.
$getNumDataProp(O)$	Get the number of data properties in the ontology.
$getNumAnnProp(O)$	Get the number of annotation properties in the ontology.
$getNumSubClassOf(O)$	Get the number of <code>subClassOf</code> relations in the ontology.
$getNumHRD(O)$	Get the number of human-readable descriptions (HRD) such as comment, label, and description.

processed resources of the two input ontologies. We use 13 string matching measures described in Table 1 and Table 2. We take different threshold values to select the top-matched resources in the matching process. MoMatch matches not only ontologies in different natural languages (i.e., cross-lingual matching) but also ontologies in the same natural language (i.e., monolingual matching). In monolingual matching, the translation phase is skipped, and the matching occurs between the pre-processed resources of the two ontologies.

The output of this phase is the matched resources stored in a matrix  $M$ .  $M$  is a two-dimensional matrix ( $m \times 4$ ), where  $m$  is the number of matched resources. Each row in  $M$  is represented as a tuple of  $(resource1, resource2, type, simScore)$ , which contains `resource1` from  $O_1$ , `resource2` from  $O_2$ , `type` of both resources, and the similarity score between them. For example,  $(Car, , C, 1.00)$ , i.e., “Car” in English and “” in Arabic are two resources of a type class, from  $O_1$  and  $O_2$ , respectively, with a similarity score of 1.00.

## V. QUALITY ASSESSMENT

We propose the metric suite QASO for assessing the quality of the matching process. These metrics need to have prior statistical information about the ontologies. We calculate the statistics of the input ontologies using a distributed in-memory approach for statistical computing calculations of large-scale RDF datasets using Apache Spark [43]. A sample for the statistics metrics in QASO is described in Table 3.

### A. QASO SUITE

QASO comprises seven metrics for assessing the quality of the ontologies and seven for evaluating the quality of the matching process. We adapt and reformulate the metrics defined in [6], [7], and [8].

- *Relationship richness (RR)* [6]: refers to the diversity of relations and their position in the ontology. The more relations the ontology has (except `rdfs:subClassOf` relation), the richer it is. The quality score function  $f_{RR} : O \rightarrow \mathbb{R}$  for an input ontology  $O$  is defined as follows:

$$f_{RR}(O) = \frac{|P_{obj}|}{|P_{subClassOf}| + |P_{obj}|} \in [0, 1] \quad (12)$$

where  $P_{obj}$  represents the relationships (i.e., object properties) and  $P_{subClassOf}$  represents the `rdfs:subClassOf` relations in  $O$ .

- *Attribute Richness (AR)* [6]: refers to how much knowledge about classes is in the schema. The more attributes are defined, the more knowledge the ontology provides. The quality score function  $f_{AR} : O \rightarrow \mathbb{R}$  for an input ontology  $O$  is defined as follows:

$$f_{AR}(O) = \frac{|P_{attr}|}{|C|} \in [0, +\infty] \quad (13)$$

where  $C$  represents the ontology classes and  $P_{attr}$  represents all classes' attributes (i.e., data properties).

- *Inheritance Richness (IR)* [6]: refers to how well knowledge is distributed across different levels in the ontology. The more `rdfs:subClassOf` relations, the broader range of general knowledge the ontology provides. The quality score function  $f_{IR} : O \rightarrow \mathbb{R}$  for an input ontology  $O$  is defined as follows:

$$f_{IR}(O) = \frac{|P_{subClassOf}|}{|C|} \in [0, +\infty] \quad (14)$$

- *Readability (RB)* [6]: refers to the existence of human-readable descriptions (HRD) in the ontology, such as comments, labels, or descriptions. The more HRD exists, the more readable the ontology is. The quality score function  $f_{RB} : O \rightarrow \mathbb{R}$  for an input ontology  $O$  is defined as follows:

$$f_{RB}(O) = \frac{|HRD|}{|R|} \in [0, +\infty] \quad (15)$$

where  $HRD \in \{label, comment, description\}$  and  $R$  represents the ontology resources.

- *Isolated Elements (IE)* [7]: refers to classes and properties which are defined but not connected to the rest of the ontology, i.e., not used. The quality score function  $f_{IE} : O \rightarrow \mathbb{R}$  for an input ontology  $O$  is defined as follows:

$$f_{IE}(O) = \frac{|R_{isolated}|}{|R|} \in [0, 1] \quad (16)$$

where  $R_{isolated}$  represents resources defined but not used in  $O$ .

- *Missing Domain or Range in Properties (MP)* [7]: refers to missing information about properties. The less information about properties is missing, the more complete the ontology. The quality score function  $f_{MP} : O \rightarrow \mathbb{R}$  for an input ontology  $O$  is defined as follows:

$$f_{MP}(O) = \frac{|P_{incomplete}|}{|P|} \in [0, 1] \quad (17)$$

where  $P_{incomplete}$  represents properties that do not have domain or range.

- **Redundancy (RD)** [8]: refers to how many redundant resources exist. Resources which are syntactically (e.g. “isMemberOf” and “is\_member\_of”) or semantically (e.g. “Chair” and “Chairman”) close are considered as redundant resources. The quality score function  $f_{RD} : O \rightarrow \mathbb{R}$  for an input ontology  $O$  is defined as follows:

$$f_{RD}(O) = \frac{|R_r|}{|R|} \in [0, +\infty] \quad (18)$$

where  $R_r$  represents the redundant resources in  $O$ .

All the previous metrics can be used to assess the quality of any ontology. To assess the quality and effectiveness of a matching process, we need to verify whether all relevant correspondences have been retrieved (high recall) and correct (high precision). The following metrics are adapted to assess the quality of the matching process not only by using the reference alignment (*Ref*) but also without it.

- **Class Precision (CP)**: refers to the fraction of relevant matched classes among the retrieved ones. The more relevant results retrieved, the more precision the matching process has. The quality score function  $f_{CP} : O \rightarrow \mathbb{R}$  for the matching process  $M$  is defined as follows:

$$f_{CP}(M) = \frac{|Ref_c| \cap |C_{match}|}{|C_{match}|} \in [0, 1] \quad (19)$$

where  $C_{match}$  is the retrieved matched classes by MoMatch, and  $Ref_c$  is the matched classes in the reference alignment.

- **Class Recall (CR)**: refers to the fraction of relevant matched classes retrieved by the system. The more relevant results retrieved, the more recall the matching process has. The quality score function  $f_{CR} : O \rightarrow \mathbb{R}$  for the matching process  $M$  is defined as follows:

$$f_{CR}(M) = \frac{|Ref_c| \cap |C_{match}|}{|Ref_c|} \in [0, 1] \quad (20)$$

- **Property Precision (PP)**: refers to the fraction of relevant matched properties among the retrieved ones. The more relevant results retrieved, the more precision the matching process has. The quality score function  $f_{PP} : O \rightarrow \mathbb{R}$  for the matching process  $M$  is defined as follows:

$$f_{PP}(M) = \frac{|Ref_p| \cap |P_{match}|}{|P_{match}|} \in [0, 1] \quad (21)$$

where  $P_{match}$  is the retrieved matched properties by MoMatch, and  $Ref_p$  is the matched properties in the reference alignment.

- **Property Recall (PR)**: refers to the fraction of relevant matched properties retrieved by the system. The more relevant results retrieved, the more recall the matching process has. The quality score function  $f_{PR} : O \rightarrow \mathbb{R}$  for the matching process  $M$  is defined as follows:

$$f_{PR}(M) = \frac{|Ref_p| \cap |P_{match}|}{|Ref_p|} \in [0, 1] \quad (22)$$

- **Degree of Overlap (OV)**: refers to how many common resources exist between the input ontologies. Resources that are syntactically or semantically close are considered common resources. The quality score function  $f_{OV} : O \rightarrow \mathbb{R}$  for the matching process  $M$  is defined as follows:

$$f_{OV}(M) = \frac{|R_{match}|}{|R_1| \cup |R_2|} \in [0, 1] \quad (23)$$

where  $R_{match}$  represents the set of found correspondences in a match result produced when matching  $O_1$  and  $O_2$  ontologies.  $R_1$  and  $R_2$  represent the resources of the input ontologies  $O_1$  and  $O_2$  respectively.

If the reference alignment is unavailable, we use rough approximations for recall and precision based on the relative quality of the obtained matching results produced by MoMatch [44].

- **Match Coverage (MC)** [44]: estimation for recall refers to the fraction of resources that exist in at least one correspondence in the matching results in comparison to the total number of resources in the input ontologies. The quality score function  $f_{MC} : O \rightarrow \mathbb{R}$  for the matching process  $M$  is defined as follows:

$$f_{MC}(M) = \frac{|R_{O_1-match}| \cup |R_{O_2-match}|}{|R_1| \cup |R_2|} \in [0, 1] \quad (24)$$

where  $R_{O_1-match}$  and  $R_{O_2-match}$  represents the set of matched resources of ontologies  $O_1$  and  $O_2$  respectively.

- **Match Ration (MR)** [44]: estimation for precision refers to the ratio between the number of found correspondences and the number of matched resources in the input ontologies. The closer the ratio is to 1.00, the better precision the match result has. In other words, when a match result is not loosely matched to many other resources but only the most similar ones, the match result is better. The quality score function  $f_{MR} : O \rightarrow \mathbb{R}$  for the matching process  $M$  is defined as follows:

$$f_{MR}(M) = \frac{2 \times |R_{match}|}{|R_{O_1-match}| \cup |R_{O_2-match}|} \in [1, +\infty] \quad (25)$$

## VI. USE CASES

In this section, we show the applicability of MoMatch in monolingual and cross-lingual ontology matching in different domains.

### A. USE CASE 1: CROSS-LINGUAL MATCHING IN SCHOLARLY COMMUNICATION DOMAIN

In this use case, we use ontologies from the scholarly communication domain in various natural languages. We use an example scenario to match the *SEO<sub>en</sub>*<sup>9</sup> ontology (with 106 classes and 88 properties), in English, with *Conference<sub>de</sub>* ontology (60 classes and 64 properties), in German, from the

<sup>9</sup><https://w3id.org/seo>



**TABLE 4.** Use case 1: a sample output of each phase for matching  $SEO_{en}$  with  $Conference_{de}$ , starting from resource extraction to matching results.

Phase	Output
Resource extraction	$\langle$ Konferenzbeteiligung, C, $O_1$ , de) $\langle$ Vorsitzender, C, $O_1$ , de) $\langle$ hat Beitrag eingreicht, P, $O_1$ , de) $\langle$ student registration, C, $O_2$ , en)
Translation	$\langle$ Tutorium, C, $O_1$ , de, Tutorial) $\langle$ Thema, C, $O_1$ , de, Topic) $\langle$ hat Beitrag eingreicht, P, $O_1$ , de, has contribution submitted)
Pre-processing	SizeOrDuration $\rightarrow$ size duration WorkshopProposals $\rightarrow$ workshop proposal InvitedSpeaker $\rightarrow$ invite speaker In-useTrack $\rightarrow$ use track
Matching	$\langle$ Gutachter, reviewer, C, 1.00) $\langle$ eingeladener Referent, invited speakers, C, 1.00) $\langle$ eingeladener Referent, keynote speaker, C, 0.57) $\langle$ person, person, C, 1.00) $\langle$ Tutorium, tutorial proposals, C, 0.78) $\langle$ Koferenzdokument, license document, C, 0.61) $\langle$ Gutachten, reviewer, C, 0.75) $\langle$ proceedings, proceedings, C, 1.00) $\langle$ Organisator, organizer, C, 1.00) $\langle$ wichtiges Datum, important dates, C, 1.00) $\langle$ Herausgeber, publisher, C, 1.00) $\langle$ Leiter der Workshops, workshop chair, C, 1.00) $\langle$ hat wichtiges Datum, has important dates, P, 1.00) $\langle$ hat Abfolge, hasTrack, P, 1.00)

MultiFarm dataset (see section VII). The goal of this use case is to demonstrate the entire process, from submitting the input ontologies to producing the alignment. Here,  $O_1$  is the German ontology  $Conference_{de}$  and  $O_2$  is the English ontology  $SEO_{en}$ . Table 4 demonstrates the matching process and shows each phase's sample output. The relevant matching results are identified in the matching phase using Jaccard similarity with a threshold  $\theta \geq 0.90$ . MoMatch identified eight matched classes and two matched properties.

## B. USE CASE 2: MONOLINGUAL MATCHING IN BIOMEDICAL DOMAIN

In this use case, we use ontologies from the biomedical domain. We match monolingual ontologies from BioPortal<sup>10</sup>— a web-based application for accessing and sharing biomedical ontologies and providing alignments between them. We match the Sample Processing and Separation Techniques Ontology (SEP)<sup>11</sup> with several ontologies from the Bio-portal such as Plant Experimental Conditions (PECO),<sup>12</sup>

**TABLE 5.** Statistics of BioPortal ontologies ( $R_{match}$  = number of matched resources, i.e., correspondences) and degree of overlap (OV) between SEP ontology and other ontologies.

Ontology Name	Classes	Properties	$R_{match}$ in BioPortal	$R_{match}$ by MoMatch	OV <sup>16</sup> by MoMatch
PECO	3,119	124	11	46	1.17%
PO	2,019	111	1	39	1.38%
PTO	1,659	1	0	37	2%
UO	634	2	448	434	34%
SEP	658	2	–	–	–

Plant (PO),<sup>13</sup> Plant Trait (PTO),<sup>14</sup> and Units of Measurement (UO)<sup>15</sup> ontologies. Table 5 shows the statistics of the Bioportal ontologies and the degree of overlap between SEP with each ontology. The degree of overlap refers to how many common resources exist between ontology pairs as defined in subsection V-A. MoMatch identified 434 out of 448 correspondences found in Bioportal for matching SEP  $\times$  UO with a 34% degree of overlap. Surprisingly, MoMatch found new correspondences which were missing in BioPortal. In matching SEP  $\times$  PECO, MoMatch found 35 new correspondences. Similarly, in matching SEP  $\times$  PO, MoMatch found 38 new correspondences. In matching SEP  $\times$  PTO, MoMatch found 37 new correspondences.

## VII. EVALUATION

In this section, we show the results of two experiments to provide an in-depth analysis for comparing similarity measures across different languages. In addition, we evaluate the quality of the matching process using QASO.

### A. DATASET AND EXPERIMENT SETUP

#### 1) DATASET

We use MultiFarm benchmark<sup>17</sup> from the T-Box/Schema matching track of OAEI 2020.<sup>18</sup> The OAEI 2020 competition is an annual international ontology matching competition. MultiFarm is a cross-lingual ontology matching system evaluation benchmark. It consists of seven ontologies (*Cmt*, *Conference*, *ConfOf*, *Edas*, *Ekaw*, *Iasted*, *Sigkdd*) derived from the Conference benchmark of OAEI, their translation into nine languages (Chinese, Czech, Dutch, French, German, Portuguese, Russian, Spanish and Arabic), and the corresponding cross-lingual alignments between them. Statistics for Multifarm ontologies are presented in Table 6. Classification of the nine languages according to Wikipedia<sup>19</sup> is presented in Table 7. We could not use the dataset from OAEI 2021 [45] because the results are not available in detail (for each language pair and each ontology pair) as in OAEI 2020.

<sup>13</sup><https://bioportal.bioontology.org/ontologies/PO>

<sup>14</sup><https://bioportal.bioontology.org/ontologies/PTO>

<sup>15</sup><https://bioportal.bioontology.org/ontologies/UO>

<sup>17</sup><https://www.irit.fr/recherches/MELODI/multifarm/>

<sup>18</sup><http://oaei.ontologymatching.org/2020/>

<sup>19</sup>[https://en.wikipedia.org/wiki/Language\\_family](https://en.wikipedia.org/wiki/Language_family)

<sup>10</sup><https://bioportal.bioontology.org/>

<sup>11</sup><https://bioportal.bioontology.org/ontologies/SEP>

<sup>12</sup><https://bioportal.bioontology.org/ontologies/PECO>

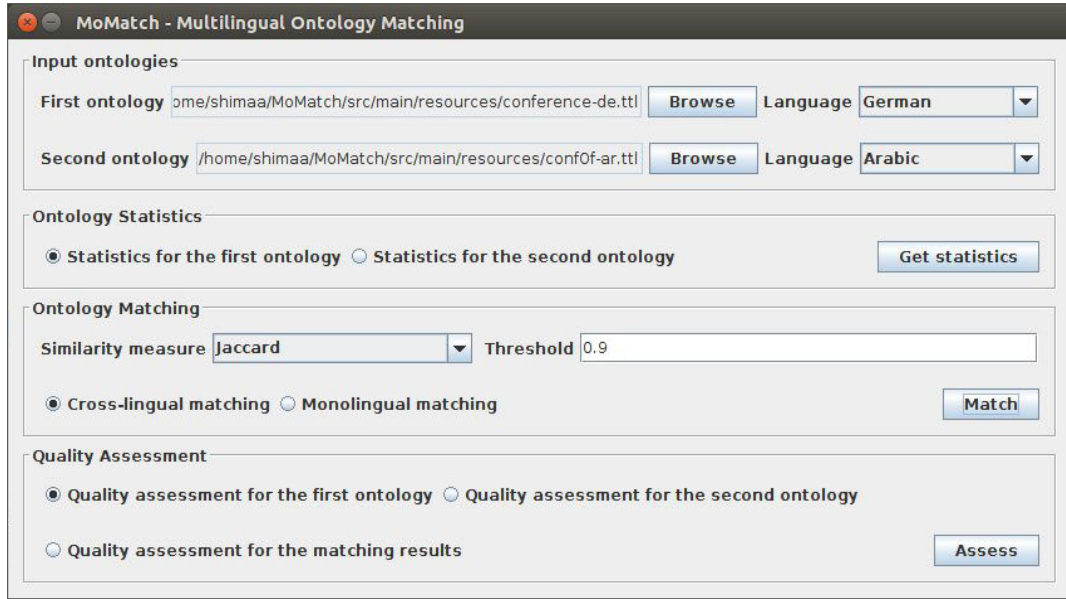


FIGURE 2. MoMatch GUI.

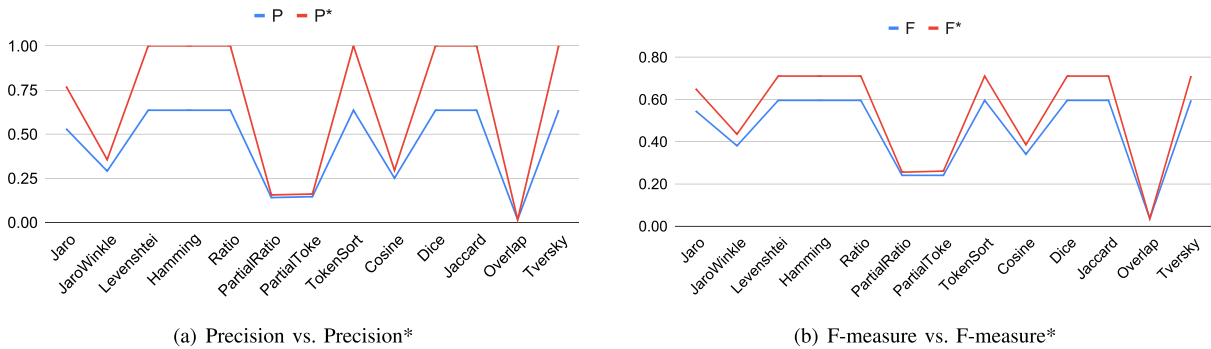


FIGURE 3. Matching French–German ontologies where  $\theta \geq 0.90$ . P\* and F\* represent the improvement in precision and F-measure results when considering the new correspondences.

TABLE 6. Statistics of MultiFarm ontologies.

Ontology Name	Classes	Object Properties	Data Properties
<i>Ekaw</i>	74	33	0
<i>Edas</i>	104	30	20
<i>Conference</i>	60	46	18
<i>ConfOf</i>	38	13	23
<i>Cmt</i>	36	49	10
<i>Iasted</i>	140	38	3
<i>Sigkdd</i>	49	17	11

2) EXPERIMENT SETUP

Scala and Apache Spark<sup>20</sup> were used to implement all phases of MoMatch. A graphical user interface of MoMatch is created (cf. Figure 2). To parse and manipulate the input ontologies (as RDF triples), SANSARDF library<sup>21</sup> [46] with

TABLE 7. Classification of languages in Multifarm according to Wikipedia and number of LOV in each category.

Language	# of LOV	Category	
English (en)	584	Indo-European	
German (de)	41		
Dutch (nl)	20		
French (fr)	65		
Portuguese (pt)	19		
Spanish (es)	49		
Russian (ru)	17		
Czech (cz)	9		
Chinese (cn)	10		Tai-Kadai
Arabic (ar)	7		Afro-Asiatic

Apache Jena framework<sup>22</sup> are used. To process the resource labels, the Stanford CoreNLP<sup>23</sup> [47] is used. All experiments

<sup>20</sup><https://spark.apache.org/>

<sup>21</sup><https://github.com/SANSARDF/SANSARDF>

<sup>22</sup><https://jena.apache.org/>

<sup>23</sup><https://stanfordnlp.github.io/CoreNLP/>

TABLE 8. Average values (for all ontology pairs) for precision (P), recall (R), and F-measure (F) of matching multilingual ontologies. P\* and F\* describe the results with adjusted precision and F-measure. Red, green, and blue entries are the top scores for precision, recall, and F-measure for each similarity measure per row. Similar values are merged.

Similarity measure	I						0.95						0.90						0.85						0.80					
	P	P*	R	R*	F	F*	P	P*	R	R*	F	F*	P	P*	R	R*	F	F*	P	P*	R	R*	F	F*	P	P*	R	R*	F	F*
Jaro	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>
JaroWinkler	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>
Levenshtein	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>
Hamming	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>
Ratio	0.15	0.16	<b>0.73</b>	0.24	0.26	0.26	0.15	0.16	<b>0.73</b>	0.24	0.26	0.26	0.15	0.16	<b>0.73</b>	0.24	0.26	0.26	0.15	0.16	<b>0.73</b>	0.24	0.26	0.26	0.15	0.16	<b>0.73</b>	0.24	0.26	0.26
PartialRatio	0.15	0.16	<b>0.73</b>	0.24	0.26	0.26	0.15	0.16	<b>0.73</b>	0.24	0.26	0.26	0.15	0.16	<b>0.73</b>	0.24	0.26	0.26	0.15	0.16	<b>0.73</b>	0.24	0.26	0.26	0.15	0.16	<b>0.73</b>	0.24	0.26	0.26
PartialTokenSort	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>
TokenSort	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>
Cosine	0.02	0.02	0.73	0.05	0.05	0.05	0.02	0.02	0.73	0.04	0.04	0.04	0.02	0.02	0.73	0.04	0.04	0.04	0.02	0.02	0.73	0.04	0.04	0.04	0.02	0.02	0.73	0.04	0.04	0.04
Dice	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>
Jaccard	0.02	0.02	0.73	0.05	0.05	0.05	0.02	0.02	0.73	0.04	0.04	0.04	0.02	0.02	0.73	0.04	0.04	0.04	0.02	0.02	0.73	0.04	0.04	0.04	0.02	0.02	0.73	0.04	0.04	0.04
Overlap	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>
Tversky	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>	0.64	<b>1.00</b>	0.56	0.60	<b>0.71</b>	<b>0.71</b>
Similarity measure	P	P*	R	R*	F	F*	P	P*	R	R*	F	F*	P	P*	R	R*	F	F*	P	P*	R	R*	F	F*	P	P*	R	R*	F	F*

are run on Ubuntu 16.04 LTS with an Intel Core i7-4600U CPU @ 2.10GHz x 4 CPU and 10 GB of memory.

### 3) EVALUATION METRICS

The precision, recall, and F-measure metrics, inspired by the information retrieval community, can be used to evaluate the effectiveness of the matching process. We use the gold standard alignments between each pair of ontologies in Multifarm to compute precision, recall, and F-measure. Precision is the fraction of retrieved resources that are relevant, while recall is the fraction of relevant matched resources retrieved by MoMatch. The F-measure is the harmonic mean of precision and recall. Formally, precision is defined as  $TP/(TP + FP)$ , and recall is defined as  $TP/(TP + FN)$ , where  $TP$  (true positive) is a matching result that MoMatch retrieves and exists in the gold standard,  $FP$  (false positive) is a matching result that MoMatch retrieves and does not exist in the gold standard, and  $FN$  (false negative) is a matching result that exists in the gold standard, and MoMatch could not retrieve it.

### 4) EXPERIMENTAL CONFIGURATION

In our experiments, we choose one natural language for each language category in Table 7. Regarding categories with more than one language, such as Italic, we select the most widespread language according to Linked Open Vocabularies (LOV)<sup>1</sup>. For example, the Italic family has three languages where 65 vocabularies are in French, 19 in Portuguese, and 49 in Spanish, i.e., French is selected, which has the most extensive vocabulary in the Italic family. We did not include English in our experiments because it is the most prominent language, and a lot of work has been done in this language. Therefore, we used German, French, Russian, Chinese, and Arabic in the experiments.

The Multifarm benchmark is composed of 55 pairs of languages, with 49 matching tasks for each of them, taking into account the alignment direction (e.g.,  $Cmt_{en} \rightarrow Edas_{de}$  and  $Cmt_{de} \rightarrow Edas_{en}$  are distinct matching tasks), i.e., 2,695 matching task [48]. In MoMatch, we used 13 similarity measures, so we needed to perform a 35,035 matching task. To decrease the matching tasks without losing their value, we design our experiments to 1) evaluate the effectiveness of the cross-lingual matching process in MoMatch using different similarity measures compared to the reference alignment provided in the MultiFarm benchmark, 2) compare MoMatch matching results with five state-of-the-art approaches, and 3) assess the quality of the matching process using QASO.

### B. EFFECTIVENESS OF MOMATCH

In this experiment, we use all similarity measures listed in both Table 1 and Table 2. Different threshold values ( $\theta = \{1.00, 0.95, 0.90, 0.85, 0.80\}$ ) are used to select the top-matched results according to the resulting similarity scores between every two resources' labels. We select two ontologies (*Iasted* and *ConfOf*) from the Multifarm benchmark and two language pairs from different families: French–German

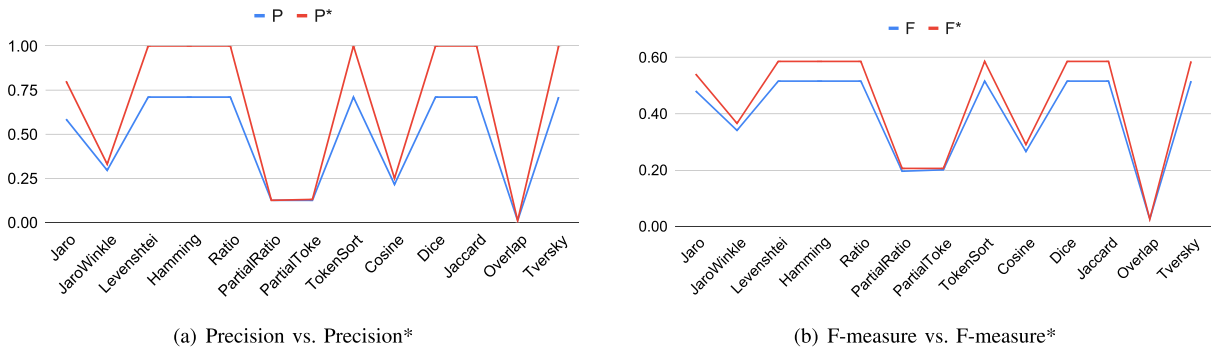
and German–Arabic. In matching French–German ontologies, we match the resources' labels of the French version of *Iasted<sub>fr</sub>* ontology with the resources' labels of the German version of *ConfOf<sub>de</sub>* (i.e.,  $Iasted_{fr} \times ConfOf_{de}$ ). Similarly, we match the resources' labels of the French version of *ConfOf<sub>fr</sub>* ontology with the resources' labels of the German version of *Iasted<sub>de</sub>*. The same procedure has been followed in matching German–Arabic ontologies. The resulting alignments are compared with the reference alignments as a gold standard provided in the benchmark for each pair of ontologies. Table 8(a) and Table 8(b) show the average values (for all ontology pairs) for precision, recall and F-measure for matching French–German and German–Arabic ontologies respectively. Surprisingly, we found new correspondences missing in the gold standard alignments, such as the correspondence  $\langle \text{écrit, schreibt, } \equiv, 1.00 \rangle$ , which implies that the French and German properties “écrit” and “schreibt” are identical with a similarity score = 1.00.  $P^*$  and  $F^*$  represent adjusted precision and F-measure results when considering the new correspondences. Therefore,  $P^*$  and  $F^*$  represent results that are not false positives in practice. Figure 3(a) and Figure 3(b) present the significant improvement of  $P^*$  and  $F^*$  over  $P$  and  $F$ , respectively, in all similarity measures for matching French–German ontologies, where  $\theta \geq 0.90$ . The precision and F-measure results are improved by an average of 23% and 8%, respectively. Similarly, Figure 4(a) and Figure 4(b) present a significant improvement in matching German–Arabic ontologies. The precision and F-measure results are improved by an average of 18% and 5%, respectively. To address our research questions, we study these results in terms of two dimensions: 1) similarity measures (RQ1) and 2) language pairs (RQ2).

### 1) SIMILARITY MEASURES

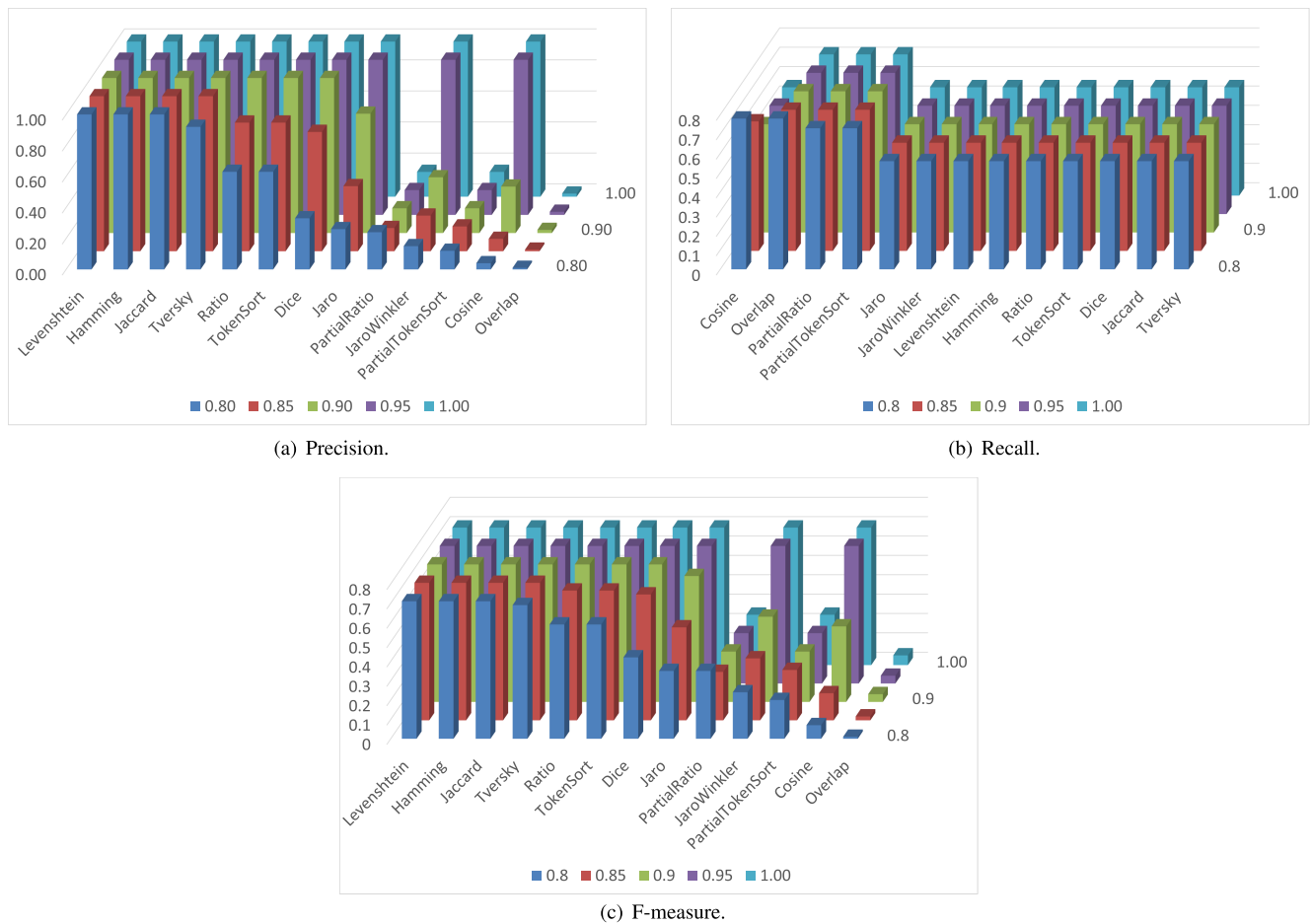
As shown in Table 8, the precision is directly proportional to the threshold values, which means the precision increases when the similarity measure value increases, especially with  $\theta \geq 0.90$ . While the recall is inversely proportional to the threshold values, e.g., the highest recall is achieved with  $\theta = 0.80$ .

In matching French–German ontologies, Levenshtein, Hamming, Jaccard, and Tversky have achieved the best precision of 100% for all thresholds except for  $\theta = 0.80$  Tversky performs 92% (see Figure 5(a)). Similarly, Levenshtein, Hamming, Jaccard, and Tversky have achieved the best F-measure of 71% for all thresholds except for  $\theta = 0.80$  Tversky reaches 69% (see Figure 5(c)). Cosine, Overlap coefficient, Partial ratio, and Partial token sort have achieved the best recall of 78%, 78%, 73%, and 73%, respectively, where  $\theta = 0.80$  (see Figure 5(b)). In matching German–Arabic ontologies, Levenshtein, Hamming, Jaccard, and Tversky have achieved the best precision of 100% for all thresholds (except for  $\theta = 0.80$ , Tversky gains 78%) (see Figure 6(a)). Levenshtein, Hamming, Jaccard, and Tversky have achieved the best F-measure of 59% for all thresholds (except for  $\theta = 0.80$ , Tversky performs 54%) (see Figure 6(c)). Cosine,





**FIGURE 4.** Matching German–Arabic ontologies where  $\theta \geq 0.90$ . The improvement of  $P^*$  and  $F^*$  have similar patterns with Figure 3(a) and Figure 3(b).

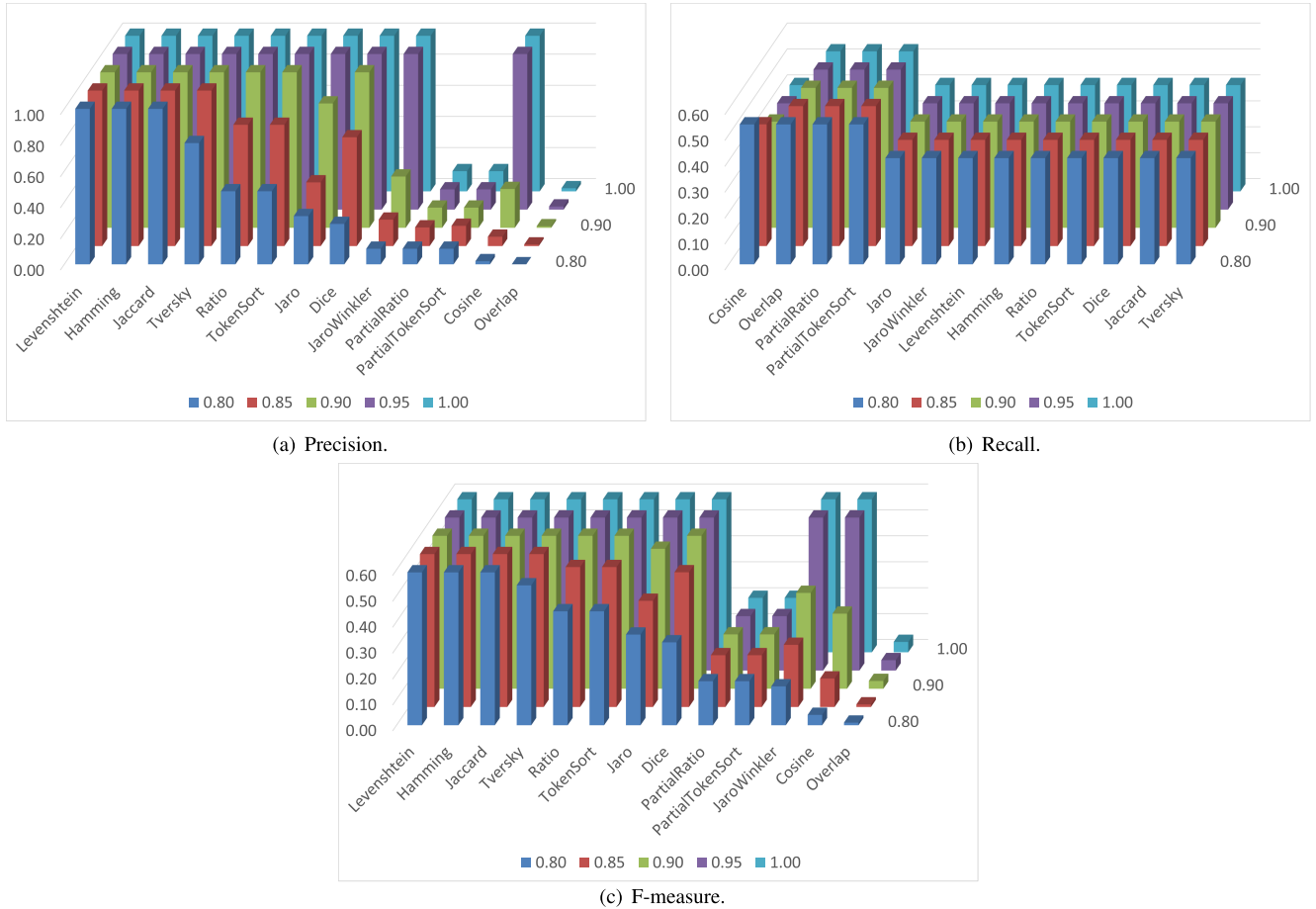


**FIGURE 5.** Comparing precision, recall, and F-measure of different similarity measures over all thresholds for matching French–German ontologies, as shown in Table 8. Similarity measures are sorted from left (the highest value) to right (the lowest value) according to precision, recall, and F-measure. The more threshold values, the more precision and F-measure scores are achieved. In contrast, the lower the threshold values, the more recall scores are achieved.

Overlap coefficient, Partial ratio, and partial token sort have achieved the best recall of 54% where  $\theta = 0.80$  (see Figure 6(b)).

Across the two language pairs, French  $\times$  German and German  $\times$  Arabic, precision and F-measure in sequence-based similarity measures are sensitive to variations in the

threshold (see Figure 7(a) and Figure 7(e)). Levenshtein, Hamming, Token sort, and Ratio provide the best precision and F-measure scores. In contrast, recall remains constant across all threshold values (see Figure 7(c)). Partial ratio and Partial token sort provide the highest recall of 64% across all threshold values. Similarly, precision and F-measure in



**FIGURE 6.** Precision, recall, and F-measure charts of different similarity measures over all thresholds for matching German-Arabic ontologies, as shown in Table 8(b). There is no significant difference between these charts and the charts for matching French-German ontologies (Figure 5).

set-based similarity measures are also sensitive to variations in the threshold (see Figure 7(b) and Figure 7(f)). Jaccard, Tversky, and Dice have the best precision and F-measure scores. The recall is relatively stable against threshold changes (see Figure 7(d)). The overlap coefficient provides the best recall of 64% and 66% for  $\theta \geq 0.80$ .

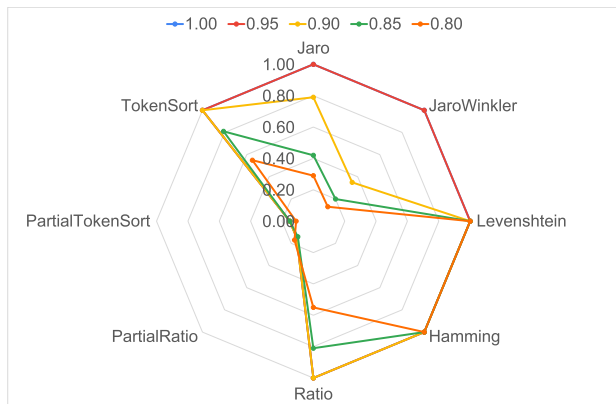
The choice of string similarity measure greatly influences the precision and recall of the matching process. When selecting a similarity measure to be used in ontology matching, it is crucial to examine the features of the ontologies being matched and whether the precision or the recall is more significant to the matching technique. Levenshtein, Hamming, and Jaccard achieve the best precision. If one wants to use more than one similarity measure, then one can choose from the following list, sorted in descending order of precision: 1) Levenshtein, Hamming, and Jaccard, 2) Tversky, 3) Token sort and Ratio, 4) Dice, 5) Jaro, 6) Jaro Winkler, 7) Cosine, 8) Partial ratio, 9) Partial token sort, 10) Overlap coefficient. On the other hand, the Overlap coefficient achieves the best recall. Similarly, if one wants to use more than one similarity measure, then one can select from the following list sorted in descending order of recall: 1) Overlap coefficient, 2) Partial

ratio and Partial token sort, 3) Cosine, 4) Jaro, Jaro Winkler, Levenshtein, Hamming, Ratio, Token sort, Dice, Jaccard, and Tversky.

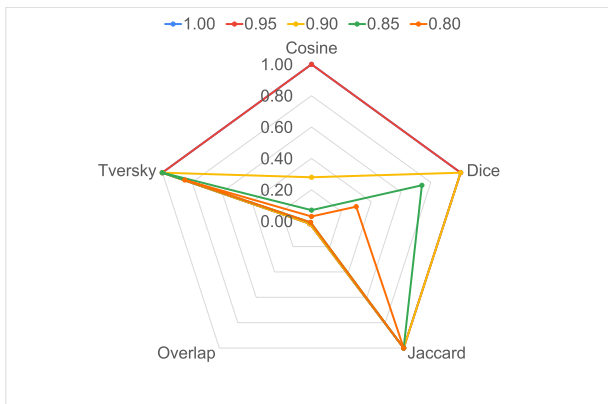
## 2) LANGUAGE PAIRS

The language pair French-German has better results than German-Arabic (see Figure 8). Two native speakers of Arabic found that the reason behind that is the linguistic mistakes found in the Arabic ontologies, which negatively affect the translation and the matching results. We correct these mistakes and make them available at the MoMatch GitHub repository<sup>2</sup>.

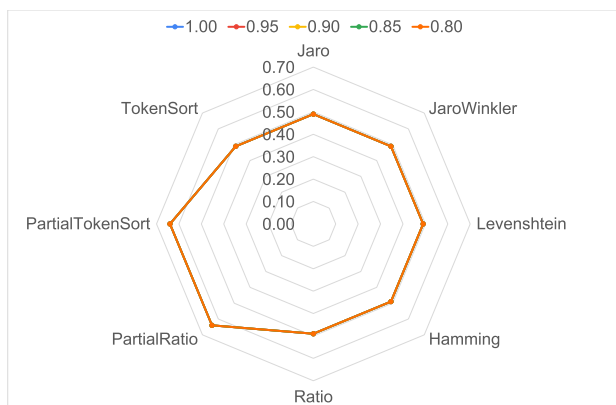
In order to answer the research question *RQ2*, we calculate the Spearman correlation between the two language pairs for precision, recall, and F-measure where  $\theta = \{1.00, 0.95, 0.90, 0.85, 0.80\}$ . All threshold values have the highest correlation of 1.00 except for  $\theta = 0.80$ ; the correlation values are 0.99, 0.80, and 0.98 for precision, recall, and F-measure, respectively. As a result, the two language pairs are positively and strongly correlated to each other, i.e., the higher similarity measure is ranked in French-German, the higher similarity measure is ranked in German-Arabic, and



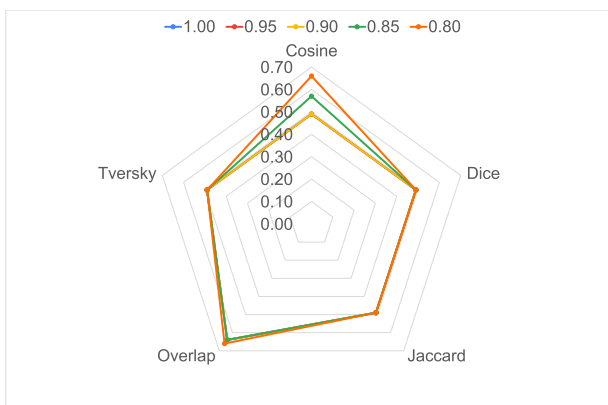
(a) Precision – Sequence-based



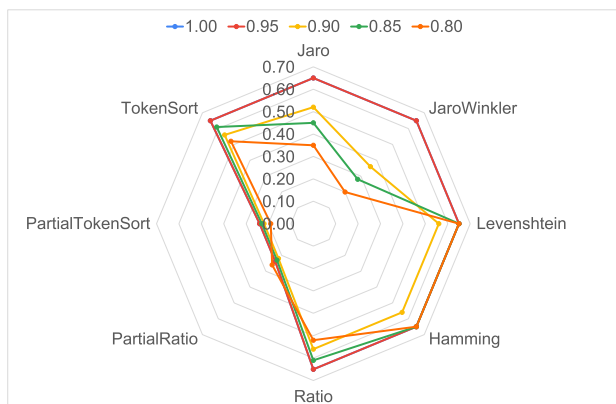
(b) Precision – Set-Based



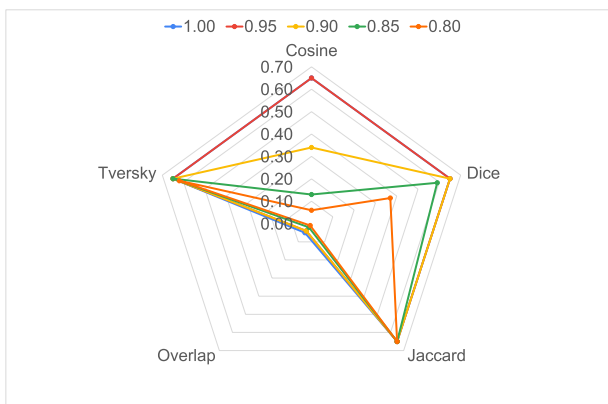
(c) Recall – Sequence-Based



(d) Recall – Set-Based



(e) F-measure – Sequence-Based



(f) F-measure – Set-Based

**FIGURE 7.** Average precision, recall, and F-measure on different similarity measures and thresholds across the two language pairs French–German and German–Arabic. The greater the distance between the threshold line and the polygon’s center, the higher the precision, recall, and F-measure scores.

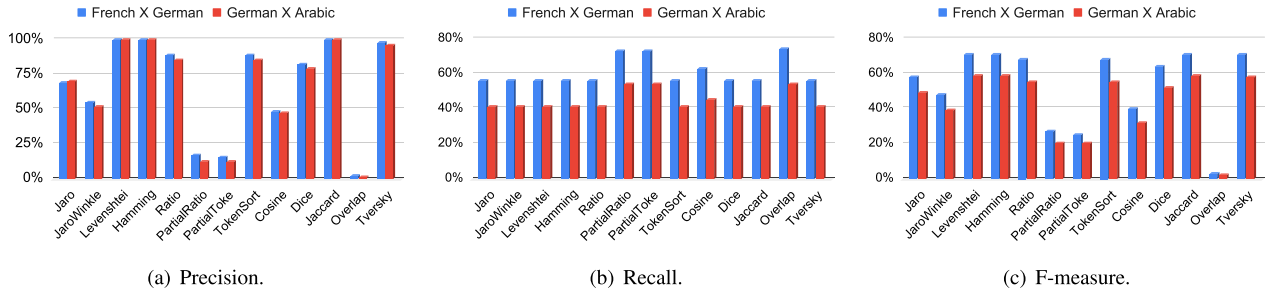


FIGURE 8. Average values (for all thresholds) for precision, recall, and F-measure for matching French–German and German–Arabic ontologies.

TABLE 9. State-of-the-art comparison results using Jaccard, Levenshtein, and Overlap coefficient similarity measures. Red, green, and blue entries are the top scores for precision, recall, and F-measure for each language pair per row.

(a) Using Jaccard similarity.

Language pairs	AML [38]			LogMap [39]			LogMapLt [39]			VeeAlign [37]			Wiktionary [40]			MoMatch			MoMatch*	
	P	F	R	P	F	R	P	F	R	P	F	R	P	F	R	P	F	R	P*	F*
German × French	0.59	0.38	0.29	0.67	0.38	0.27	1.00	0.05	0.03	0.25	0.00	0.00	0.61	0.19	0.11	0.60	0.47	0.39	0.97	0.55
German × Arabic	0.56	0.30	0.21	0.62	0.33	0.23	1.00	0.00	0.00	0.54	0.17	0.10	1.00	0.05	0.03	0.62	0.47	0.39	0.94	0.55
French × Arabic	0.51	0.26	0.18	0.42	0.17	0.11	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.58	0.42	0.33	0.96	0.49
Russian × Arabic	0.41	0.18	0.12	0.67	0.35	0.24	1.00	0.00	0.00	0.18	0.05	0.03	1.00	0.05	0.03	0.52	0.38	0.31	0.79	0.44
German × Russian	0.50	0.29	0.21	0.81	0.35	0.23	1.00	0.00	0.00	0.38	0.03	0.02	0.82	0.25	0.16	0.54	0.43	0.37	0.80	0.50
French × Russian	0.61	0.37	0.27	0.67	0.35	0.25	1.00	0.00	0.00	0.31	0.09	0.06	0.58	0.19	0.12	0.50	0.38	0.31	0.83	0.44
Chinese × Arabic	0.50	0.21	0.14	0.39	0.13	0.08	1.00	0.00	0.00	0.75	0.00	0.00	1.00	0.00	0.00	0.48	0.31	0.24	0.91	0.37
French × Chinese	0.59	0.34	0.24	0.53	0.20	0.13	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.46	0.31	0.24	0.95	0.37
German × Chinese	0.50	0.26	0.18	0.63	0.17	0.10	1.00	0.00	0.00	0.25	0.00	0.00	0.64	0.27	0.17	0.50	0.36	0.30	0.92	0.43
Russian × Chinese	0.65	0.40	0.29	0.72	0.32	0.21	1.00	0.00	0.00	0.75	0.02	0.01	1.00	0.00	0.00	0.40	0.26	0.21	0.84	0.31

(b) Using Levenshtein similarity.

Language pairs	AML [38]			LogMap [39]			LogMapLt [39]			VeeAlign [37]			Wiktionary [40]			MoMatch			MoMatch*	
	P	F	R	P	F	R	P	F	R	P	F	R	P	F	R	P	F	R	P*	F*
German × French	0.59	0.38	0.29	0.67	0.38	0.27	1.00	0.05	0.03	0.25	0.00	0.00	0.61	0.19	0.11	0.55	0.44	0.38	1.00	0.55
German × Arabic	0.56	0.30	0.21	0.62	0.33	0.23	1.00	0.00	0.00	0.54	0.17	0.10	1.00	0.05	0.03	0.58	0.50	0.44	1.00	0.61
French × Arabic	0.51	0.26	0.18	0.42	0.17	0.11	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.58	0.41	0.32	1.00	0.48
Russian × Arabic	0.41	0.18	0.12	0.67	0.35	0.24	1.00	0.00	0.00	0.18	0.05	0.03	1.00	0.05	0.03	0.52	0.36	0.29	0.83	0.42
German × Russian	0.50	0.29	0.21	0.81	0.35	0.23	1.00	0.00	0.00	0.38	0.03	0.02	0.82	0.25	0.16	0.55	0.43	0.36	0.80	0.49
French × Russian	0.61	0.37	0.27	0.67	0.35	0.25	1.00	0.00	0.00	0.31	0.09	0.06	0.58	0.19	0.12	0.50	0.37	0.30	0.83	0.43
Chinese × Arabic	0.50	0.21	0.14	0.39	0.13	0.08	1.00	0.00	0.00	0.75	0.00	0.00	1.00	0.00	0.00	0.50	0.30	0.23	1.00	0.36
French × Chinese	0.59	0.34	0.24	0.53	0.20	0.13	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.48	0.31	0.24	1.00	0.37
German × Chinese	0.50	0.26	0.18	0.63	0.17	0.10	1.00	0.00	0.00	0.25	0.00	0.00	0.64	0.27	0.17	0.53	0.37	0.30	0.97	0.43
Russian × Chinese	0.65	0.40	0.29	0.72	0.32	0.21	1.00	0.00	0.00	0.75	0.02	0.01	1.00	0.00	0.00	0.41	0.26	0.21	0.83	0.31

(c) Using Overlap coefficient similarity.

Language pairs	AML [38]			LogMap [39]			LogMapLt [39]			VeeAlign [37]			Wiktionary [40]			MoMatch			MoMatch*	
	P	F	R	P	F	R	P	F	R	P	F	R	P	F	R	P	F	R	P*	F*
German × French	0.59	0.38	0.29	0.67	0.38	0.27	1.00	0.05	0.03	0.25	0.00	0.00	0.61	0.19	0.11	0.02	0.03	0.56	0.02	0.03
German × Arabic	0.56	0.30	0.21	0.62	0.33	0.23	1.00	0.00	0.00	0.54	0.17	0.10	1.00	0.05	0.03	0.03	0.05	0.69	0.03	0.05
French × Arabic	0.51	0.26	0.18	0.42	0.17	0.11	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.03	0.06	0.52	0.03	0.06
Russian × Arabic	0.41	0.18	0.12	0.67	0.35	0.24	1.00	0.00	0.00	0.18	0.05	0.03	1.00	0.05	0.03	0.03	0.06	0.48	0.03	0.06
German × Russian	0.50	0.29	0.21	0.81	0.35	0.23	1.00	0.00	0.00	0.38	0.03	0.02	0.82	0.25	0.16	0.03	0.06	0.49	0.03	0.06
French × Russian	0.61	0.37	0.27	0.67	0.35	0.25	1.00	0.00	0.00	0.31	0.09	0.06	0.58	0.19	0.12	0.03	0.05	0.44	0.03	0.05
Chinese × Arabic	0.50	0.21	0.14	0.39	0.13	0.08	1.00	0.00	0.00	0.75	0.00	0.00	1.00	0.00	0.00	0.03	0.06	0.42	0.03	0.06
French × Chinese	0.59	0.34	0.24	0.53	0.20	0.13	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.02	0.04	0.38	0.02	0.04
German × Chinese	0.50	0.26	0.18	0.63	0.17	0.10	1.00	0.00	0.00	0.25	0.00	0.00	0.64	0.27	0.17	0.03	0.05	0.43	0.03	0.05
Russian × Chinese	0.65	0.40	0.29	0.72	0.32	0.21	1.00	0.00	0.00	0.75	0.02	0.01	1.00	0.00	0.00	0.03	0.05	0.34	0.03	0.05

vice versa. This result indicates that regardless of language pairs, the resulting scores for precision, recall, and F-measure obtained from the combination of different similarity measures and threshold values stay highly correlated.

C. COMPARISON WITH THE STATE-OF-THE-ART

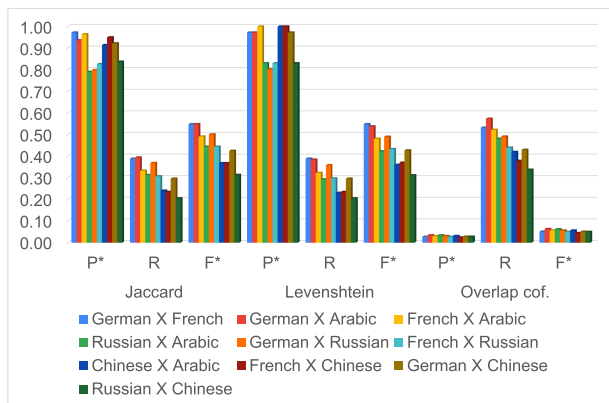
In this experiment, we identified five of the related approaches (AML [38], LogMap [39], LogMapLt [39],

VeeAlign [37], Wiktionary [40]) to be included in our evaluation. The other related works neither publish their code nor their evaluation datasets [26], [27], [31]. In order to show the applicability of MoMatch, we use different language pairs. We select the broadest language from each category in Table 7, i.e., German (Germanic), French (Italic), Russian (Balto-Slavic), Chinese (Tai-Kadai), and Arabic (Afro-Asiatic). Therefore, we have ten pairs of language as follows:



**TABLE 10. Quality assessment results for input ontologies and the matching results. Bold entries are the top scores.**

Input Ontologies	RR <sup>12</sup>	AR <sup>13</sup>	IR <sup>14</sup>	RB <sup>15</sup>	IE <sup>16</sup>	MP <sup>17</sup>	RD <sup>18</sup>
<i>Conference<sub>de</sub></i>	<b>0.48</b>	<b>0.78</b>	0.83	1.17	0.11	0.02	0.00
<i>Edas<sub>fr</sub></i>	0.26	0.29	0.82	0.08	<b>0.77</b>	0.00	0.00
<i>Conference<sub>ru</sub></i>	<b>0.48</b>	<b>0.78</b>	0.83	1.17	0.11	0.02	0.00
<i>Ekaw<sub>ar</sub></i>	0.32	0.45	<b>0.97</b>	0.02	0.02	<b>0.21</b>	0.00
<i>ConfOf<sub>de</sub></i>	0.28	0.34	0.87	<b>1.45</b>	0.04	0.03	0.00
<i>Iasted<sub>ar</sub></i>	0.22	0.27	0.94	1.02	0.08	0.10	0.00
Ontology pairs	CP <sup>19</sup>	CR <sup>20</sup>	PP <sup>21</sup>	PR <sup>22</sup>	OV <sup>23</sup>	MC <sup>24</sup>	MR <sup>25</sup>
<i>Conference<sub>de</sub> × Edas<sub>fr</sub></i>	0.86	<b>0.46</b>	<b>0.50</b>	<b>0.67</b>	0.05	0.09	1.00
<i>Conference<sub>ru</sub> × Ekaw<sub>ar</sub></i>	0.44	0.17	0.33	0.50	<b>0.06</b>	<b>0.11</b>	<b>1.04</b>
<i>ConfOf<sub>de</sub> × Iasted<sub>ar</sub></i>	<b>1.00</b>	0.44	0.00	N/A	0.03	0.05	1.00

**FIGURE 9. Precision, recall, and F-measure for matching ontologies in ten language pairs using Jaccard vs. Levenshtein vs. Overlap coefficient similarity measures.**

German × French, German × Russian, German × Chinese, German–Arabic, French × Russian, French × Chinese, French × Arabic, Russian × Chinese, Russian × Arabic, and Chinese × Arabic. We choose the most effective similarity measures in terms of precision, recall, and F-measure from the previous experiment. Therefore, we select Jaccard and Levenshtein, which has achieved the best precision and F-measure, and Overlap coefficient, which has achieved the best recall. In order to compare our results with the state-of-the-art, we match *Conference* (the middle ontology in Table 6) with *Edas* and *Ekaw* ontologies as mentioned in the results of OAEI 2020<sup>7</sup>. Therefore, there are four ontology pairs in each language pair. For example, in matching French × Chinese, the ontology pairs are: *Conference<sub>fr</sub> × Edas<sub>cn</sub>*, *Conference<sub>cn</sub> × Edas<sub>fr</sub>*, *Conference<sub>fr</sub> × Ekaw<sub>cn</sub>*, and *Conference<sub>cn</sub> × Ekaw<sub>fr</sub>*. We evaluate the quality of the matching process by calculating precision, recall, and F-measure as in the previous experiment. Table 9(a), Table 9(b), and Table 9(c) show a comparison between MoMatch’s results for matching ontologies in ten pairs of languages against five state-of-the-art systems using Jaccard, Levenshtein, and Overlap coefficient respectively. We found new correspondences for all ontology pairs which were missing in the gold standard alignments. MoMatch\* represents results when considering the new correspondences. It provides the matching results with the adjusted precision and F-measure. LogMapLt

achieves the highest precision of 100% but the lowest recall and F-measure of 0% for all language pairs. Similarly, Wiktionary achieves the highest precision of 100% but the lowest recall and F-measure of 0% for all language pairs with Arabic or Chinese, except for German × Chinese. MoMatch outperforms most other systems in terms of precision, recall, and F-measure when using Jaccard and Levenshtein similarities and not considering the new correspondences as false positives. For instance, in matching German × French ontologies, MoMatch outperforms AML, the highest precision (after LogMapLt), recall, and F-measure among the others in matching German × French ontologies, by 38%, 10%, and 17% in terms of precision, recall, and F-measure respectively when using Jaccard similarity. Similarly, MoMatch outperforms AML by 41%, 9%, and 17% in precision, recall, and F-measure, respectively, when using Levenshtein similarity.

The use of Jaccard and Levenshtein similarity measures give relatively similar results in precision, recall, and F-measure (see Figure 9). While the use of Overlap coefficient similarity in MoMatch achieves the highest recall among the other systems for matching all language pairs. These results confirm our findings from the previous experiment where Jaccard and Levenshtein similarity measures can achieve the best precision and F-measure while the Overlap coefficient can accomplish the best recall.

We calculate the Spearman correlation between F-measure values for Jaccard, Levenshtein, and Overlap coefficient produced by MoMatch, and the ten language pairs. All language pairs achieve the highest correlation of 1.00 except the two language pairs, French × Chinese and Russian × Chinese achieve 0.87. In addition, the correlation between German × Chinese and French × Chinese achieve 0.87 as well. As a result, the ten language pairs are positively and strongly correlated to each other. These results confirm our results from the previous experiment (see subsection VII-B), where the effects of different similarity measures are independent of the language pair.

#### D. EVALUATING THE QUALITY OF THE MATCHING PROCESS USING QASO

In this experiment, we randomly choose three pairs of ontologies from the previous experiments (*Conference<sub>de</sub> × Edas<sub>fr</sub>*, *Conference<sub>ru</sub> × Ekaw<sub>ar</sub>*, and *ConfOf<sub>de</sub> × Iasted<sub>ar</sub>*). We assess

the quality of the matching process using QASO described in subsection V-A. Table 10 shows the assessment results for each input ontology in addition to the assessment for the matching results. The quality metric results for *Conference<sub>de</sub>* and *Conference<sub>ru</sub>* are identical because it is the same ontology but in two different natural languages (German and Russian). *Conference* has the highest results of 48% and 78% in terms of relationship richness and attribute richness, respectively. In terms of inheritance richness, *Ekaw<sub>ar</sub>* achieves the highest results of 97%. *Ekaw<sub>ar</sub>* is the most ontology that suffers from missing properties information by 21%.

MoMatch achieves the highest class precision of 100% for matching *ConfOf<sub>de</sub>*  $\times$  *Iasted<sub>ar</sub>* and the highest class recall of 46% for matching *Conference<sub>de</sub>*  $\times$  *Edas<sub>fr</sub>*. In terms of property precision and recall, MoMatch achieves the highest results of 50% and 67%, respectively, in matching *Conference<sub>de</sub>*  $\times$  *Edas<sub>fr</sub>*. The property recall for matching *ConfOf<sub>de</sub>*  $\times$  *Iasted<sub>ar</sub>* is N/A because there are no properties retrieved in the reference alignment ( $Ref_p = 0$ ). MoMatch identifies the highest degree of overlap of 6% in matching *Conference<sub>ru</sub>*  $\times$  *Ekaw<sub>ar</sub>*, where it has the highest results for match coverage and match ratio of 11% and 104% respectively.

## VIII. CONCLUSION

We propose the MoMatch approach, which matches ontologies in different natural languages. We show a comparative analysis of 13 different string similarity measures. Additionally, we present QASO – a metrics suite for assessing the quality of any ontology and the quality of the matching process. We test the performance of MoMatch over different ontologies in different natural languages, including Indo and non-Indo-European languages. We find that Levenshtein, Hamming, and Jaccard similarity measures have the highest precision and F-measure, while partial ratio, partial token sort, and overlap coefficient have the highest recall for matching multilingual ontologies (RQ1). Accordingly, we sorted the 13 similarity measures into two lists according to precision and recall to support choosing the most appropriate one for the matching process. The correlation between language pairs is consistently high and positive (RQ2). The results of the cross-lingual matching process in MoMatch are found to be promising compared to five state-of-the-art approaches. We assess the quality of the matching process using QASO (RQ3). We show the usability of MoMatch by presenting two use cases in scholarly communication and biomedical domains for both cross-lingual and monolingual ontology matching. MoMatch can be easily adapted for other use cases and domains. In conclusion, MoMatch established the first step toward a multilingual Semantic Web.

In the future, we intend to further; 1) consider individuals in the matching process, 2) include other similarity measures, such as string-based structural measures which consider the entity's neighbors in the matching process, and 3) develop scalable approaches to match large-scale ontologies and knowledge graphs efficiently.

## REFERENCES

- [1] G. O. Consortium, "The gene ontology resource: 20 years and still going strong," *Nucleic Acids Res.*, vol. 47, no. 1, pp. 330–338, 2019.
- [2] S. El Kadiri, W. Terkaj, E. N. Urwin, C. Palmer, D. Kiritsis, and R. Young, "Ontology in engineering applications," in *Proc. Int. Workshop Formal Ontologies Meet Industries*. Cham, Switzerland: Springer, 2015, pp. 126–137.
- [3] E. Ong, L. L. Wang, J. Schaub, J. F. O'Toole, B. Steck, A. Z. Rosenberg, F. Dowd, J. Hansen, L. Barisoni, S. Jain, and I. H. De Boer, "Modelling kidney disease using ontology: Insights from the kidney precision medicine project," *Nature Rev. Nephrology*, vol. 16, no. 11, pp. 686–696, 2020.
- [4] C. Trojahn, B. Fu, O. Zamazal, and D. Ritze, "State-of-the-art in multilingual and cross-lingual ontology matching," in *Towards the Multilingual Semantic Web*. Berlin, Germany: Springer, 2014, pp. 119–135.
- [5] J. Euzenat and P. Shvaiko, *Ontology Matching*, 2nd ed. Heidelberg, Germany: Springer-Verlag, 2013.
- [6] S. Tartir, I. B. Arpinar, M. Moore, A. P. Sheth, and B. Aleman-Meza, "OntoQA: Metric-based ontology quality analysis," in *Proc. IEEE Workshop Knowl. Acquisition From Distrib., Auton., Semantically Heterogeneous Data Knowl. Sources (ICDM)*, Boston, MA, USA, 2005, pp. 45–53.
- [7] S. Mc Gurk, C. Abela, and J. Debattista, "Towards ontology quality assessment," in *Proc. MEPDaW/LDQ ESWC*, 2017, pp. 94–106.
- [8] S. Raunich and E. Rahm, "Towards a benchmark for ontology merging," in *Proc. OTM Confederated Int. Conf. Move Meaningful Internet Syst.* Cham, Switzerland: Springer, 2012, pp. 124–133.
- [9] R. S. I. Wilson, J. S. Goonetillake, A. Ginige, and W. A. Indika, "Ontology quality evaluation methodology," in *Proc. Int. Conf. Comput. Sci. Appl.* Cham, Switzerland: Springer, 2022, pp. 509–528.
- [10] R. Wilson, J. S. Goonetillake, W. Indika, and A. Ginige, "Analysis of ontology quality dimensions, criteria and metrics," in *Proc. Int. Conf. Comput. Sci. Appl.* Cham, Switzerland: Springer, 2021, pp. 320–337.
- [11] S. Ibrahim, S. Fathalla, J. Lehmann, and H. Jabeen, "Multilingual ontology merging using cross-lingual matching," in *Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. Intell. Agent Technol. (WI-IAT)*, Dec. 2020, pp. 113–120.
- [12] A. Bookstein, V. A. Kulyukin, and T. Raita, "Generalized Hamming distance," *Inf. Retr.*, vol. 5, no. 4, pp. 353–375, 2002.
- [13] A. Doan, A. Halevy, and Z. Ives, *Principles of Data Integration*. Amsterdam, The Netherlands: Elsevier, 2012.
- [14] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Sov. Phys.-Dokl.*, vol. 10, no. 8, pp. 707–710, 1966.
- [15] G. A. Rao, G. Srinivas, K. VenkataRao, and P. P. Reddy, "A partial ratio and ratio based fuzzy-wuzzy procedure for characteristic mining of mathematical formulas from documents," *IJSC-ICTACT J. Soft. Comput.*, vol. 8, no. 4, pp. 1728–1732, 2018.
- [16] L. Bulygin, "Combining lexical and semantic similarity measures with machine learning approach for ontology and schema matching problem," in *Proc. Int. Conf. Data Anal. Manag. Data Intensive Domains (DAM-DID/RCDL)*, 2018, pp. 245–249.
- [17] G. Petasis, V. Karkaletsis, G. Paliouras, A. Krithara, and E. Zavitsanos, "Ontology population and enrichment: State of the art," in *Knowledge-Driven Multimedia Information Extraction and Ontology Evolution*. Berlin, Germany: Springer-Verlag, 2011, pp. 134–166.
- [18] M. Florence, "Building a multilingual ontology for education domain using monto method," *Comput. Sci. Inf. Technol.*, vol. 1, no. 2, pp. 47–53, Jul. 2020.
- [19] A. Guedes, F. Baiao, and K. Revoredo, "On the identification and representation of ontology correspondence antipatterns," in *Proc. 8th Int. Workshop Modular Ontologies (WoMO)*, vol. 1248, 2014, pp. 38–48.
- [20] L. Otero-Cerdeira, F. J. Rodríguez-Martínez, and A. Gómez-Rodríguez, "Ontology matching: A literature review," *Exp. Syst. Appl.*, vol. 42, no. 2, pp. 949–971, Feb. 2015.
- [21] M. K. Vijaymeena and K. Kavitha, "A survey on similarity measures in text mining," *Mach. Learn. Appl. Int. J.*, vol. 3, no. 2, pp. 19–28, 2016.
- [22] A. Tversky, "Features of similarity," *Psychol. Rev.*, vol. 84, no. 4, p. 327, Jul. 1977.
- [23] J. Gracia, E. Montiel-Ponsoda, P. Cimiano, A. Gómez-Pérez, P. Buitelaar, and J. McCrae, "Challenges for the multilingual web of data," *J. Web Semantics*, vol. 11, pp. 63–71, Mar. 2012.
- [24] M. A. Helou and M. Palmonari, "Cross-lingual lexical matching with word translation and local similarity optimization," in *Proc. 11th Int. Conf. Semantic Syst.*, Sep. 2015, pp. 97–104.
- [25] F. A. Musyaffa, M.-E. Vidal, F. Orlandi, J. Lehmann, and H. Jabeen, "IOTA: Interlinking of heterogeneous multilingual open fiscal DaTA," *Exp. Syst. Appl.*, vol. 147, Jun. 2020, Art. no. 113135.

- [26] B. Fu and R. Brennan, "Cross-lingual ontology mapping and its use on the multilingual semantic web," in *Proc. 1st Workshop Multilingual Semantic Web*, vol. 571, Jan. 2010, pp. 13–20.
- [27] B. Fu, R. Brennan, and D. O'Sullivan, "A configurable translation-based cross-lingual ontology mapping system to adjust mapping outcomes," *J. Web Semantics*, vol. 15, pp. 15–36, Sep. 2012.
- [28] S. Hertling and H. Paulheim, "WikiMatch: Using Wikipedia for ontology matching," *Ontology Matching*, vol. 946, pp. 1–47, Nov. 2012.
- [29] F. Lin and A. Krizhanovsky, "Multilingual ontology matching based on wiktionary data accessible via SPARQL endpoint," in *Proc. RCDL*, 2011, pp. 1–8.
- [30] C. Trojahn, P. Quaresma, and R. Vieira, "A framework for multilingual ontology mapping," in *Proc. 6th Int. Lang. Resour. Eval. (LREC)*. Marrakesh, Morocco: European Language Resources Association, 2008.
- [31] A. N. Tigrine, Z. Bellahsene, and K. Todorov, "Light-weight cross-lingual ontology matching with LYAM++," in *Proc. OTM Confederated Int. Conf. Move Meaningful Internet Syst.* Cham, Switzerland: Springer, 2015, pp. 527–544.
- [32] M. Ali, S. Fathalla, S. Ibrahim, M. Kholief, and Y. F. Hassan, "CLOE: A cross-lingual ontology enrichment using multi-agent architecture," *Enterprise Inf. Syst.*, vol. 13, nos. 7–8, pp. 1002–1022, Sep. 2019.
- [33] S. Ibrahim, S. Fathalla, H. S. Yazdi, J. Lehmann, and H. Jabeen, "OECM: A cross-lingual approach for ontology enrichment," in *Proc. Eur. Semantic Web Conf.* Cham, Switzerland: Springer, 2019, pp. 100–104.
- [34] S. Ibrahim, S. Fathalla, H. S. Yazdi, J. Lehmann, and H. Jabeen, "From monolingual to multilingual ontologies: The role of cross-lingual ontology enrichment," in *Proc. Int. Conf. Semantic Syst.* Cham, Switzerland: Springer, 2019, pp. 215–230.
- [35] A. Khiat, E. A. Ouhiba, M. A. Belfedhal, and C. E. Zoua, "SimCat results for OAEI 2016," in *Proc. OM ISWC*, 2016, pp. 217–221.
- [36] A. Khiat, "Crolom: Cross-lingual ontology matching system," in *Proc. 15th Int. Semantic Web Conf. (ISWC)*, 2016, pp. 146–152.
- [37] V. Iyer, A. Agarwal, and H. Kumar, "VeeAlign: A supervised deep learning approach to ontology alignment," in *Proc. OM ISWC*, 2020, pp. 216–224.
- [38] B. Lima, D. Faria, F. M. Couto, I. F. Cruz, and C. Pesquita, "OAEI 2020 results for AML and AMLC," in *Proc. OM ISWC*, 2020, pp. 154–160.
- [39] E. Jimenez-Ruiz, "Logmap family participation in the OAEI 2020," in *Proc. 15th Int. Workshop Ontology Matching (OM)*, vol. 2788, 2020, pp. 201–203.
- [40] J. Portisch and H. Paulheim, "Wiktionary matcher results for OAEI 2020," in *Proc. CEUR Workshop*, vol. 2788. Aachen, Germany: RWTH, 2020, pp. 225–232.
- [41] C. Meilicke, R. García-Castro, F. Freitas, W. R. van Hage, E. Montiel-Ponsoda, R. R. De Azevedo, H. Stuckenschmidt, O. Sváb-Zamazal, V. Svátek, A. Taminin, C. Trojahn, and S. Wang, "MultiFarm: A benchmark for multilingual ontology matching," *J. Web Semantics*, vol. 15, pp. 62–68, Sep. 2012.
- [42] M. A. Helou, M. Palmonari, and M. Jarrar, "Effectiveness of automatic translations for cross-lingual ontology mapping," *J. Artif. Intell. Res.*, vol. 55, pp. 165–208, Jan. 2016.
- [43] G. Sejdiu, I. Ermilov, J. Lehmann, and M. N. Mami, "DistLODStats: Distributed computation of RDF dataset statistics," in *Proc. Int. Semantic Web Conf.* Cham, Switzerland: Springer, 2018, pp. 206–222.
- [44] X. Xue, Y. Wang, W. Hao, and J. Hou, "Optimizing ontology alignments through NSGA-II without using reference alignment," *Comput. Informat.*, vol. 33, no. 4, pp. 857–876, 2014.
- [45] M. A. N. Pour, "Results of the ontology alignment evaluation initiative 2021," in *Proc. CEUR Workshop*, vol. 3063, 2021, pp. 62–108.
- [46] J. Lehmann, G. Sejdiu, L. Buhmann, P. Westphal, C. Stadler, I. Ermilov, S. Bin, N. Chakraborty, M. Saleem, A. C. N. Ngomo, and H. Jabeen, "Distributed semantic analytics using the SANSa stack," in *Proc. Int. Semantic Web Conf.* Cham, Switzerland: Springer, 2017, pp. 147–155.
- [47] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky, "The Stanford CoreNLP natural language processing toolkit," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics, Syst. Demonstrations*, 2014, pp. 55–60. [Online]. Available: <http://www.aclweb.org/anthology/P14/P14-5010>
- [48] N. Pour, A. Algergawy, R. Amini, D. Faria, I. Fundulaki, I. Harrow, S. Hertling, E. Jimenez-Ruiz, C. Jonquet, N. Karam, and A. Khiat, "Results of the ontology alignment evaluation initiative 2020," in *Proc. 15th Int. Workshop Ontology Matching (OM)*, vol. 2788, 2020, pp. 92–138.



**SHIMAA IBRAHIM** received the B.Sc. and M.Sc. degrees from the University of Alexandria, Egypt. She is currently pursuing the Ph.D. degree with the Smart Data Analysis Group, University of Bonn, Germany. Her current research interests include semantic web technologies, natural language processing, machine learning, and knowledge graphs.



**SAID FATHALLA** received the M.Sc. degree in computer science from the University of Alexandria, Egypt, and the Ph.D. degree in information systems from the University of Bonn, Germany, in 2021. He has several conferences and journals publications in the semantic web, text classification, bioinformatics, scholarly communication, and knowledge engineering. He has published over 40 papers in peer-reviewed journals and international conferences. His current research interests include semantic web technologies, scholarly communication, knowledge graph construction, metadata management, and natural language processing.



**JENS LEHMANN** received the Ph.D. degree (summa cum laude) from the University of Leipzig and the joint master's degree in computer science from the Technical University of Dresden (TU Dresden) and the University of Bristol. He is currently a Principal Scientist with Amazon, working at Alexa AI on conversational AI, knowledge graphs, and generalized intelligence. He is also an Honorary Professor with TU Dresden. Previously, he was a Full Professor at the University of Bonn and Fraunhofer IAIS. His academic activities at TU Dresden and InfAI support the Smart Data Analytics Research Group. He has authored more than 300 publications, which were cited more than 25000 times. His research interests include semantic web technologies, question answering, machine learning, and knowledge graph analysis. He was selected as a fellow of ELLIS. He is a member of InfAI. He has received 16 international best papers awards, including multiple "Test of Time" awards.



**HAJIRA JABEEN** received the Ph.D. degree in computer science from the National University of Computing and Emerging Sciences, Islamabad, Pakistan. She is currently a Team Leader in big data analytics with the GESIS-Leibniz Institute for the Social Sciences and an Associate Researcher with the University of Bonn. Her research interests include artificial intelligence, data analytics, knowledge graphs, and scalable machine learning. She has authored many journals articles and conference papers in these areas.