

## PAPER

# Ontology Matching and Repair Based on Semantic Association and Probabilistic Logic

Nan WU<sup>†</sup>, Xiaocong LAI<sup>†</sup>, Mei CHEN<sup>†</sup>, and Ying PAN<sup>†a)</sup>, *Nonmembers*

**SUMMARY** With the development of the Semantic Web, an increasing number of researchers are utilizing ontology technology to construct domain ontology. Since there is no unified construction standard, ontology heterogeneity occurs. The ontology matching method can fuse heterogeneous ontologies, which realizes the interoperability between knowledge and associates to more relevant semantic information. In the case of differences between ontologies, how to reduce false matching and unsuccessful matching is a critical problem to be solved. Moreover, as the number of ontologies increases, the semantic relationship between ontologies becomes increasingly complex. Nevertheless, the current methods that solely find the similarity of names between concepts are no longer sufficient. Consequently, this paper proposes an ontology matching method based on semantic association. Accurate matching pairs are discovered by existing semantic knowledge, and then the potential semantic associations between concepts are mined according to the characteristics of the contextual structure. The matching method can better carry out matching work based on reliable knowledge. In addition, this paper introduces a probabilistic logic repair method, which can detect and repair the conflict of matching results, to enhance the availability and reliability of matching results. The experimental results show that the proposed method effectively improves the quality of matching between ontologies and saves time on repairing incorrect matching pairs. Besides, compared with the existing ontology matching systems, the proposed method has better stability.

**key words:** *semantic relevance, semantic similarity, ontology matching, matching and repair*

## 1. Introduction

In the development of the Semantic Web, ontology technology has always attracted the attention of researchers, which makes the research of ontology have a significant effect. With the continuous improvement of ontology construction platform functions, more and more researchers in different industries can construct their ontologies according to requirements [1]. Nevertheless, there are various complicated semantic relationships between ontology concepts, which lead to the differences between information and semantics contained in the ontology. Therefore, Ontology Matching (a.k.a. Ontology Alignment (OA)) technology is a method proposed to solve the problem of ontology heterogeneity and implement interoperability between ontologies. Ontology matching technology has been used extensively in the fields of ontology engineering, biomedical, Peer-to-Peer (P2P) information sharing, Web service composition, and semantic

Internet of Things [2].

The core of ontology matching technology is to semantic association concepts with similar meanings, where each pair of associated concepts is called a matching pair. Usually, more semantic associations can be found to enhance the matching effect based on the similarity of strings and structures between concepts or the use of external knowledge. Simultaneously, the combination of different similarity calculation methods has its advantages. How to decrease matching mistakes and unsuccessful matching and how to perform high-quality and high-efficiency ontology matching have been the goal of research in this field [3], [4].

Hence, this paper will concentrate on ontology matching to carry out the following work:

- (1) An Ontology Matching Method Based on Semantic Association (OMSA) is proposed. First, this approach introduces a synonyms dictionary to calculate the semantic similarity between concepts, which obtains a batch of accurate and reliable matching pairs. Then, the structural similarity between entities is calculated based on the internal structural relationships, which deduce more matching pairs containing potential semantic associations. The aforementioned strategies facilitate matching based on reliable knowledge to discover more semantic associations between ontologies, surpassing singular similarity calculation methods.
- (2) To enhance the usability of the matching results, introduce a Probabilistic Description Logic (PDL) method, which can detect and repair the logic conflicts of matching results. Thus, an Ontology Matching System Based on Semantic Association and Probabilistic Logic (OMS-SAAPL) is proposed. This approach can provide better matching pair detection and repair process, further improving matching results.

The remains of the arrangement are as follows. Section 2 introduces the current status of research on ontology matching and alignment repair at home and abroad. Section 3 describes the proposed ontology matching and repair method based on semantic association and probabilistic logic. Section 4 conducts comparative experiments on the experimental dataset provided by the Ontology Evaluation Alignment Initiative (OAEI) competition, which will demonstrate the superiority of the proposed method. The last part summarizes the research content and provides an outlook on future research directions.

Manuscript received February 1, 2024.

Manuscript revised May 14, 2024.

Manuscript publicized July 11, 2024.

<sup>†</sup>Guangxi Key Lab of Human-machine Interaction and Intelligent Decision, Nanning Normal University, Nanning 530001, China.

a) E-mail: panying@nnnu.edu.cn (Corresponding author)

DOI: 10.1587/transinf.2024EDP7028

## 2. Related Work

### 2.1 Research Status of Ontology Matching

With the continuous development of ontology technology, pursuing a mature matching system can effectively solve the difficulties of heterogeneity ontology fusion. Due to the difference in data volume and structure of the ontology, researchers have proposed quantities of ontology matching systems from different knowledge perspectives. Existing ontology matching methods work from a single or multi-angle, where multi-angle matching methods are greatly affected by weight assignment.

In addition to the most essential string similarity calculation methods, utilizing reliable knowledge improves the matching quality. Faria et al. proposed an ontology matching system based on lexical matching. This system utilized external resources such as WordNet to assist in ontology matching [5]. In practical applications, different ontologies represent the same domain knowledge. Slater et al. proposed an ontology-based vocabulary expansion method to achieve knowledge fusion [6]. The synonym expansion algorithm included two matching methods and steps, such as pruning candidate synonyms. The cross-ontology synonym expansion method could supplement the available vocabulary and improve the matching efficiency.

Researchers also addressed the ontology matching problem from the perspective of representation learning. Hertling et al. proposed an instance-based class matching approach. Initially, the Doc2vec approach was utilized to acquire vector representations of concepts. Subsequently, cosine distance was computed from the vectors to determine concept similarity [7]. Furthermore, Lütke utilized the Graph Walk method to create a corpus. Subsequently, the SkipGram model was used to learn the embedding representations of the corpus. Lastly, the stable marriage algorithm was utilized to identify optimal matches [8]. Nevertheless, the representations of entities are usually not accurate enough, neglecting the rich information of the ontology. Consequently, Li et al. proposed TransO, a knowledge representation learning model based on ontology information constraints, which could integrate rich ontology information into knowledge graphs to enhance the performance of the model [9].

With the increasing numbers of multilingualism on the Web, the ontology should have richer multilingual annotations to facilitate multilingual and cross-lingual ontology matching tasks. Accordingly, Ibrahim et al. proposed a Multilingual Ontology Matching (MoMatch) approach to matching different natural language ontologies [10]. The method utilized a group of string similarity strategies to discover the same matches between entities, while machine translation was used to identify the correspondence between different ontologies.

Besides the above methods, adding external resources is also a way to enhance the quality of matching. Xue et

al. formally defined the entity matching problem of ontology and proposed an ontology matching approach based on Interactive Compact Genetic Algorithm (ICGA) [11]. Compact encoding mechanisms and expert interaction mechanisms were utilized, which could enhance algorithm performance and entity alignment quality.

Currently, an ontology matching system was to match a couple of ontologies in the same domain. When confronted with multi-domain problems, the existing ontology matching system created scalability problems while disregarding multi-domain specificities. Consequently, Silva et al. proposed an ontology alignment approach, which measured the semantic overlap of ontology based on a high-confidence fast matching technique [12]. Two alignment strategies, pair-based and increment-based, were utilized in the alignment process.

Presently, a majority of ontology matching work is based on concept names, ontology structures, and external resources to match features. However, these traditional methods disregard the semantic information between concepts, which leads to low matching accuracy [4]. Beyond that, ontology matching methods still maintain many challenges [13]. In different fields, how to select an adequately matching method and cope with large ontologies more comprehensively and efficiently would be a question worth thinking about in future work.

### 2.2 Research Status of Alignment Repair

Usually, when at the fewest one unsatisfiable concept is inferred from semantic information (i.e., two contradictory conclusions are obtained), the problem of logical conflict arises, which leads to ontology inconsistency. Consequently, the matching results need to be detected and repaired after matching. Alignment repair is a method to solve the problem of logical conflict by removing some incoherent mappings. In recent years, many researchers have proposed solutions to the alignment repair problem, and the related researches are as follows.

Frequently, the alignment of large ontologies was logically inconsistent. Accordingly, Jiménez-Ruiz et al. introduced an ontology matching tool with inferential and diagnostic capabilities. During reconciliation, it dynamically detected unsatisfiable concepts and employed a greedy algorithm to automatically rectify matching results [14]. Husein et al. proposed a heuristic alignment repair approach based on dynamic weighting [15]. First, looked for unsatisfiable classes and kept the conflict mapping in the conflict set. Then, two minimization focus diagnostic approaches had employed in the repair stage, which minimized the number and confidence values of removed mapping to accomplish conflict-free alignments.

Some researchers have found that the inclusion relationship between concepts in ontology was similar to the connectivity relationship between directed graph nodes. Li et al. proposed a graph-based structure to resolve incoherent mapping pairs [16]. Transformed the DL-Lite ontology

into a graph structure, which would detect inconsistencies in mapping pairs and found the max set of unrelated mapping pairs. In addition, an influence function based on mapping weights and graphs was designed to reduce the number of expert decisions.

Generally, when agents used different ontologies to represent knowledge, ontology alignment was used for communication. Consequently, Van den Berg et al. proposed an ontology alignment repair approach based on dynamic epistemic logic [17]. When a communication failure occurred, the Agent could communicate and apply adaptation operators to detect and repair errors in ontology alignment.

If the ontology or data source schema changes, which would require modifying the mapping associations between ontologies. Consequently, Lembo et al. proposed a method to repair mapping pairs based on the principle of minimum change [18]. Furthermore, two concepts, deletion-based and implicit-based mapping repair, were defined. Li et al. proposed an approach to repair the mapping results of ontology by using probabilistic reasoning and belief revision technology [19]. First, the mapping weights were converted into probability intervals using a probabilistic description logic approach. Then, the probability intervals for incoherent mapping pairs were relaxed, which could alleviate conflicts until probabilistic coherence.

However, existing repair approaches usually utilize feature engineering or non-contextual word embeddings to repair but do not outperform rule-based systems. He et al. proposed an ontology alignment system based on the Bidirectional Encoder Representation from Transformer (BERT) model [20]. First, constructing a BERT fine-tuning task to learn the meaning of the concept, which could apply the resulting classifier to mapping predictions. Then, the mapping was extended and repaired with ontology structure and logic.

Moreover, in many cases matching results were invalid and required manual repair and review by domain experts. Therefore, making full use of some inspection tools could assist the work of experts. Santos et al. introduced an ontology alignment repair and inspection tool, which displayed the alignment between ontologies and helped experts visually perform alignment repair tasks [21]. During the ontology matching process, professionals could repair the ontology in time, effectively avoiding a series of errors, but it took quite a little time and effort. Alrabbaa et al. provided an interactive view tool for Enhanced visual ontology navigation and emendation (Evonne) [22]. This tool could display the diagnosed defects and the modular structure of the ontology described by atomic decomposition, which assisted evaluate the effect of removing an axiom.

While the above approaches have achieved significant results on the alignment repair problem, there are still some deficiencies. Presently, alignment repair methods generally collect candidate error mapping based on unsatisfiable concepts. While such approaches ensure logical coherence, repair does not necessarily guarantee the quality of the obtained mapping, leading to inefficient repair [19], [23].

### 3. Ontology Matching and Repair Method Based on Semantic Association and Probabilistic Logic

#### 3.1 Related Concepts

Due to the complexity of ontology structure, researchers frequently utilize Web Ontology Language (OWL) to describe ontology. Presently, as there is no unified definition of ontology, the most common form definition is as follows.

**Definition 3.1 (Ontology) [10].** The ontology  $O$  consists of a quintuple, which can describe as  $O = \langle C, HC, R, HR, I \rangle$ , where  $O$  consists of concept set  $C$ , concept hierarchy set  $HC$ , relation set  $R$ , relational architecture  $HR$ , and instance set  $I$ . This paper focuses on the matching relationships between concepts in ontology. To ensure the standardization of the paper description,  $O_1$  and  $O_2$  are used to represent the two ontologies to be matched, and  $C_1$  and  $C_2$  represent the two concepts to be matched in the ontologies  $O_1$  and  $O_2$ , respectively.

**Definition 3.2 (Ontology matching)** The matching pair obtained by ontology matching consists of a quintuple  $\langle id, C_1, C_2, R, w \rangle$ . Where  $id$  represents the unique serial number of matching pairs,  $C_1$  and  $C_2$  describe the concepts in the ontology  $O_1$  and  $O_2$ , respectively. The  $R$  indicates the relationship between  $C_1$  and  $C_2$ , which contains a series of relationships such as equality, inequality, inclusion, intersection, etc. The  $w$  expresses the similarity value, and  $w \in [0, 1]$ . The higher the  $w$  value, the more likely it is that  $C_1$  and  $C_2$  represent the same thing, and vice versa, the less likely it is to be similar.

#### 3.2 Ontology Matching Method Based on Semantic Association

##### 3.2.1 Semantic Similarity

The calculation of semantic similarity usually requires the assistance of the WordNet synonymy set. WordNet is a large-scale English word knowledge base system, which depicts the lexical ontology in the form of a semantic network, and plenty of synonymy sets are used to organize various vocabularies. Usually, the semantic relationship between words contains synonymy, antonymy, hypernym and hyponym, meronym and holonym, entailment, etc.

There are three main steps to matching. First, the synonym table is imported, extracts two candidate matching concepts from the to be matched ontology  $O_1$  and  $O_2$ , respectively, while denoted as  $C_1$  and  $C_2$ , and maps these two concepts into WordNet knowledge. Then, look up its corresponding synonym sets  $Synonyms(C_1)$  and  $Synonyms(C_2)$  from the vocabulary, where  $Synonyms(C_1) = \{syn_{11}, syn_{12}, \dots, syn_{1n}\}$ ,  $Synonyms(C_2) = \{syn_{21}, syn_{22}, \dots, syn_{2m}\}$ . Ultimately, through the two synonym sets, the similarity of the string is calculated as follows.

$$sim_{WordNet}(C_1, C_2) =$$

$$\begin{cases} 1, & \text{syn}_{1i} = \text{syn}_{2j} \\ \max_{\substack{i=1,2,\dots,n; \\ j=1,2,\dots,m}} (sim(\text{syn}_{1i}, \text{syn}_{2j}), sim(C_1, C_2)), & \text{others} \end{cases} \quad (1)$$

Where  $i \in [1, n]$ ,  $j \in [1, m]$ , where  $n$  and  $m$  are the number of synonyms of  $C_1$  and  $C_2$ , respectively. In conformity with formula (1), when concepts  $C_1$  and  $C_2$  have the synonym, their semantic similarity value  $sim_{WordNet}$  is 1. Otherwise, calculate the similarity of the synonym of concepts  $C_1$  and  $C_2$ , while the maximum value is the value of  $sim_{WordNet}$ . Where  $sim(\text{syn}_{1i}, \text{syn}_{2j})$  and  $sim(C_1, C_2)$  are both calculated cosine distances between two concepts or synonyms of the concept, this cosine value can represent the degree of similarity between the two concepts.

### 3.2.2 Structural Similarity

Typically, ontology is constructed based on the hierarchy of concepts. The researchers discovered that the structure composed of three semantic relationships, such as subClassOf, is-a, and part-of closely resembled the shape of a directed acyclic graph. Generally, the calculation method of structural similarity between concepts is based on the contextual semantic relationship. There are two main structure-based matching methods:

- (1) If the ancestor nodes of the two concepts to be matched match each other, it follows that the two concepts are more likely to match each other;
- (2) The similarity of two concepts can transfer the matching relationship to adjacent nodes. When two entities can match each other, the surrounding entities have a high probability of matching.

The most common approach is to match the ancestor or descendant nodes of two concepts through the contextual structure. The similarity of the two nodes to be matched can calculate by comparing the similarity of the ancestor node or descendant node. Figure 1 shows the calculation method of similarity.

The premise of the above two methods is to obtain some matching pairs as known semantic knowledge, which

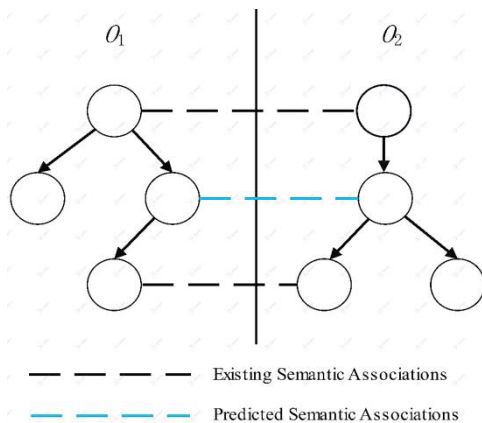


Fig. 1 Calculation of structure similarity.

assists in the inference calculation of the contextual structure. Therefore, when the correctness of the matching pair is higher, the matching effect will not be extensively affected. In this paper, the structural similarity is calculated by merging the ancestor node and descendant node, which can obtain a more reliable similarity value.

For a given concept  $C_1$  and  $C_2$ , the matching pairs corresponding to their adjacent ancestor nodes are  $M_{parent} = \{P_1, P_2, \dots, P_k\}$ , and the matching pairs corresponding to their adjacent descendant nodes are  $M_{son} = \{S_1, S_2, \dots, S_t\}$ . Among these,  $k$  and  $t$  represent the total number of existing ancestor node matching pairs and descendant node matching pairs, respectively. Where  $P_i$  ( $i \in [1, k]$ ) and  $S_j$  ( $j \in [1, t]$ ) are both described as serial numbers of matching pairs, through which the two concepts that constitute the matching pair and their corresponding similarity values can be acquired. Formula (2) defines the similarity calculation approach based on the contextual structure of concepts  $C_1$  and  $C_2$ .

$$\begin{aligned} & sim_{structure}(C_1, C_2) \\ &= \omega_1 * \frac{\sum_{i=1}^k weight_{P_i}}{k} + \omega_2 * \frac{\sum_{j=1}^t weight_{S_j}}{t} \end{aligned} \quad (2)$$

Among them,  $\omega_1$  and  $\omega_2$  define the weights of ancestor and descendant mapping in structural similarity calculation. Moreover,  $weight_{P_i}$  and  $weight_{S_j}$  represent the similarity of  $P_i$  and  $S_j$ , respectively.

Due to the particularity of the contextual structure,  $\omega_1$  and  $\omega_2$  are defined by formula (3).

$$f(\omega_1, \omega_2) = \begin{cases} \omega_1 = \omega_2 = 0.5, & \text{When } k > 0 \text{ and } t > 0 \\ \omega_1 = 1, & \text{When } k > 0 \text{ and } t = 0 \\ \omega_2 = 1, & \text{When } k = 0 \text{ and } t > 0 \end{cases} \quad (3)$$

In accordance with Eq. (3), when there is no ancestor matching pair between  $C_1$  and  $C_2$ , set  $\omega_2 = 1$ . When there is no descendant matching pair between  $C_1$  and  $C_2$ , set  $\omega_1 = 1$ . When both are present, set  $\omega_1 = \omega_2 = 0.5$ . Through structural similarity calculation, reliable potential matching pairs can be retained.

### 3.2.3 Ontology Matching Process

The design idea of the OMSA method is as follows: First, the semantic similarity of the ontology  $O_1$  and  $O_2$  to be matched is calculated, and the matching pairs below the threshold  $\tau$  are filtered to obtain the preliminary matching results. Then, calculate the structural similarity of the concepts that failed to match in ontology  $O_1$  and  $O_2$ , filter the matching pairs below the threshold  $\tau$ , and more matching results containing potential semantic associations are obtained. Table 1 shows the pseudocode of the ontology matching process. The specific processing steps are as follows.

- (1) The concepts  $C_1$  and  $C_2$  are extracted from ontology  $O_1$  and  $O_2$  to be matched, respectively.



**Table 1** Pseudocode for the ontology matching process.

---

**Algorithm 1** OMSA( $O_1, O_2, \tau$ )

---

**Input:** Ontology  $O_1$  and  $O_2$

**Output:** Matching results  $M$

1. *for*  $i \in [1, m]$  *in*  $O_1$ :
2.   *for*  $j \in [1, n]$  *in*  $O_2$ :
3.      $sim_{WordNet}(C_i, C_j) \leftarrow$  Calculated using Equation (1)
4.     **if**  $sim_{WordNet}(C_i, C_j) > \tau$  **then**
5.       List1 = Max( $sim_{WordNet}(C_i, C_j)$ )
6.     **end if**
7.   **end for**
8.  $M = (C_i, C_j, \text{Max}(sim_{WordNet}(C_i, C_j)))$
9. Delete  $C_j$  in  $O_2$
10. **end for**
11. *for*  $(C_i, C_j, sim_{WordNet}(C_i, C_j))$  in  $M$ :
12.   Delete  $C_i$  in  $O_1$
13. **end for**
14. *for*  $i \in [1, m]$  *in*  $O_1$ :
15.   *for*  $j \in [1, n]$  *in*  $O_2$ :
16.      $sim_{structure}(C_i, C_j) \leftarrow$  Calculated using Equation (2)
17.     **if**  $sim_{structure}(C_i, C_j) > \tau$  **then**
18.       List2 = Max( $sim_{structure}(C_i, C_j)$ )
19.     **end if**
20.   **end for**
21.  $M = (C_i, C_j, \text{Max}(sim_{structure}(C_i, C_j)))$
22. Delete  $C_j$  in  $O_2$
23. **end for**
24. **return**  $M$

---

- (2) Calculate the semantic similarity of concepts  $C_1$  and  $C_2$ , denoted as  $sim_{WordNet}(C_1, C_2)$ . As long as all the concepts in  $O_2$  have traversed, then filter. The principle of filtering is to store them in the candidate list List1 when  $sim_{WordNet}(C_1, C_2) > \tau$ . Otherwise, continue to select the next concept from  $O_2$  and perform the iterative calculation in step (2).
- (3) The concept with the maximum semantic similarity value is selected from List1 and stored with  $C_1$  in the matching result  $M$ , then deletes this concept from ontology  $O_2$ . When all concepts in ontology  $O_1$  have traversed, go to the next step.
- (4) The preliminary matching result  $M$  is obtained through semantic similarity, and all successfully matched concepts are deleted from  $O_1$ .
- (5) Two unmatched concepts,  $C_1$  and  $C_2$ , are extracted from ontology  $O_1$  and  $O_2$ , respectively.
- (6) Calculate the structural similarity of concepts  $C_1$  and  $C_2$ , denoted as  $sim_{structure}(C_1, C_2)$ . When all the concepts in  $O_2$  have traversed, the next step is filtering. The criterion of filtering is to store them in the candidate list List2 when  $sim_{structure}(C_1, C_2) > \tau$ . On the contrary,

the next concept is selected from  $O_2$  to perform the iterative calculation in step (6).

- (7) The concept with the maximum structural similarity value is determined from List2 and stored with  $C_1$  in the matching result  $M$ , then deletes this concept from ontology  $O_2$ . When all concepts in ontology  $O_1$  have traversed, go to the next step.
- (8) Output the final matching result  $M$ .

### 3.3 Alignment Repair Method Based on Probabilistic Logic

#### 3.3.1 Related Concepts

Probabilistic description logic is an extension of description logic, which utilizes probability to infer uncertain knowledge in the objective world, thereby providing a rationale for the occurrence of uncertain knowledge. The related concepts of the PDL method are defined while demonstrating the superiority of the repair method [19], [24].

In the knowledge base, there is an interpretation  $I = (\Delta^I, \bullet^I)$ , in which  $I_C$  is defined as the set of worlds related to the concept set  $C$ . Among them,  $\Delta^I$  represents the interpretation domain,  $\bullet^I$  represents the interpretation function. In probabilistic logic, as long as there is a probabilistic interpretation that  $\text{Pr}$  is a function on  $I_C$ , it can be defined as  $\text{Pr} : I_C \rightarrow [0, 1]$ , and  $\sum_{W \in I_C} \text{Pr}(W) = 1$ . In the real world, it can be comprehended that the sum of all possible probabilities of an event is 1. In the ontology knowledge base, if there is a probabilistic interpretation that  $\text{Pr}$  satisfies a TBox (Terminological Axioms), it can be expressed as  $\text{Pr} \models \text{TBox}$ . Consequently, solving the inconsistency problem in TBox can be converted into cracking the inconsistency problem of  $\text{Pr}$ . The probabilistic knowledge base consists of two parts, one is PTBox set, which is a classical (descriptive logic) knowledge base, including the term probabilistic knowledge. Another is the PABox set, which contains the assertion of probabilistic knowledge about instances. This paper merely analyzes the term set, so it solely introduces PTBox.

**Definition 3.3** (Probabilistic Satisfiability of Concepts) [18]. Given a knowledge base  $T$ , concept  $C$  is satisfied in  $T$  when  $T$  possesses a model that satisfies  $\text{Pr}(\text{Pr} \models T)$  and  $\text{Pr}(C) > 0$ .

When applying this constraint to the repair process of matching results, it is essential to map the weights of matching pairs into a probability interval. The mapping transformation rules are defined as formulas (4)–(6), which facilitate the repair work utilizing the probabilistic logic approach.

$$(A_i, B_j, \sqsubseteq, n) \mapsto (B_j | A_i)[n, 1] \quad (4)$$

$$(A_i, B_j, \supseteq, n) \mapsto (A_i | B_j)[n, 1] \quad (5)$$

$$(A_i, B_j, \equiv, n) \mapsto (A_i \sqcap B_j | A_i \sqcap B_j)[n, 1] \quad (6)$$

Where  $A_i$  and  $B_j$  represent the matching pair to be repaired,  $n$  is the weight of the matching. In addition,  $\sqsubseteq$  indicates the inclusion relationship,  $\supseteq$  indicates its inverse, and  $\equiv$  indicates the equivalence relationship. Equation (6) represents

the conversion of the matching pair into two conditional intervals,  $(B_j|A_i)[n, 1]$  and  $(A_i|B_j)[n, 1]$ .

**Definition 3.4** (Probabilistic incoherent alignment) [19]. Suppose there are two ontologies  $O_i$  and  $O_j$ , and  $M$  is their matching result. Assuming that  $M$  is probabilistically incoherent, it means that in the semantic knowledge before ontology matching, then there is at least one concept  $C$  that is probabilistically satisfiable in  $O_i$  or  $O_j$  but probabilistically is not satisfiable in  $O_i \cup O_j \cup MC$ . In the formula,  $MC$  is the constraint condition for converting  $M$  according to the mapping rules (4)–(6). If there is no such concept  $C$ ,  $M$  is probabilistically coherent.

From Eqs. (4)–(6), when the concept  $C$  is probabilistically unsatisfiable at  $O_i \cup O_j \cup M$ , then the concept  $C$  is also probabilistically unsatisfiable after transformation through the transformation rule.

### 3.3.2 Alignment Repair Process for Matching Results

The probabilistic logic and standard correction methods are used to detect and repair the matching results. Probabilistic logic is primarily utilized to reason inconsistent knowledge, and the probability interval is adjusted appropriately to alleviate the inconsistency.

In the alignment repair method based on probabilistic logic, to decrease the computational complexity, a threshold  $\epsilon$  is designated for the string size of the two concepts in the matching pair to filter. There are two steps to determining the threshold: First, the longest string  $\text{maxlength}(C_1)$  and  $\text{maxlength}(C_2)$  and the shortest string  $\text{minlength}(C_1)$  and  $\text{minlength}(C_2)$  are extracted from the two concept sets  $C_1$  and  $C_2$ , respectively. Then, 80% of the difference between the maximum and the minimum number of strings in the two concept sets is the size of  $\epsilon$ , calculated by the formula (7). Among them, the determination of the  $\epsilon$  value as 0.8 primarily relies on empirical rules [19]. The threshold can be set to dynamically adapt to concept sets of varying lengths, enhancing the algorithm's flexibility.

$$\delta = (\max(\text{maxlength}(C_1), \text{maxlength}(C_2)) - \min(\text{minlength}(C_1), \text{minlength}(C_2)) + 1) \times 0.8 \quad (7)$$

If there is a string in the matching pair that exceeds the threshold  $\delta$ , the satisfiability of its probability is not detected, i.e., the matching pair is judged as an irrelevant matching pair and repaired directly.

A probabilistic logic repair approach is introduced under the traditional framework to alleviate logic conflicts. The specific implementation steps are as follows. The pseudocode for step 3 is shown in Table 2.

- (1) Collect potentially false matching pairs. The semantics between two ontologies and their matching results are associated, which can discover the Minimal Incoherence-Preserving Subset (MIPS). Among them, MIPS contain potentially false matching pairs.
- (2) Select a candidate matching pair. Depending on the

**Table 2** Pseudocode for repair candidate matching pairs.

<b>Algorithm 2</b> ARPL(MIPS, $\delta$ )	
<b>Input:</b>	MIPS
<b>Output:</b>	Repair result $R$
1.	<b>while</b>  MIPS  > 0:
2.	<b>for</b> $(C_i, C_j)$ in MIPS:
3.	<b>if</b> $\text{length}(C_i, C_j) > \delta$ <b>then</b>
4.	$R = \text{Repair}(C_i, C_j)$
5.	Delete $(C_i, C_j)$ in MIPS
6.	<b>else</b>
7.	Transform $(C_i, C_j) \leftarrow$ Utilize equations (4)–(6) transformations
8.	<b>if</b> $(C_i, C_j)$ is probabilistically coherent <b>then</b>
9.	retain $(C_i, C_j)$
10.	<b>else</b>
11.	adjust probability interval $(C_i, C_j)$
12.	<b>if</b> $(C_i, C_j)$ is <b>not</b> probabilistically coherent <b>then</b>
13.	$R = \text{Repair}(C_i, C_j)$
14.	Delete $(C_i, C_j)$ in MIPS
15.	<b>end if</b>
16.	<b>end if</b>
17.	<b>end for</b>
18.	<b>end while</b>
19.	<b>return</b> $R$

frequency of false mapping, semantic relevance, and weight of matching pairs, the most appropriate matching pairs are selected from MIPS for repair.

- (3) Repair candidate matching pairs. As long as there is no potential implicit condition in the candidate matching pair, it is deleted directly without repair. Otherwise, it needs to be repaired. These two repair strategies are employed to reduce computational complexity. The first strategy is to repair candidate matching pairs directly. When the string lengths of the two concepts are too complicated, that is, greater than the threshold  $\delta$ . The second strategy is utilized to solve the case where the string length of the two concepts is less than the threshold  $\delta$  in the candidate matching pair. First, the matching pairs are converted by the Eqs. (4)–(6). Then, it is determined whether the matching pairs are probabilistically incoherent. If the probabilities are coherent, retain them. Otherwise, it is critical to adjust its probability interval to alleviate its incoherence. If the probabilistically incoherent remain, repair them.
- (4) Update MIPS. Remove MIPS associated with candidate matching pairs and continue to iterate steps (2) to (4) until no MIPS appear.

### 3.4 OMS-SAAPL System

The ontology matching method based on semantic associa-

tion and the alignment repair method based on probabilistic logic are combined to construct a matching system named OMS-SAAPL. The system primarily contains four steps, and its processing process is as follows.

- (1) Preprocessing. The ontology  $O_1$  and  $O_2$  to be matched are input in the form of OWL files, and then the ontologies are loaded and parsed using the Jena tool, and the extracted concepts are stored in a list. Simultaneously, irrelevant information is removed from the textual description of the concepts.
- (2) Matching processing. First, the WordNet synonym set is utilized to assist in calculating the semantic similarity of concepts. Then, the structural similarity of the unmatched concepts is calculated. The matching approach OMSA is described in detail in Sect. 3.2.
- (3) Repairing process for matching results. First, the matching results  $M$  is combined with the semantic knowledge of ontology. Then, the matching results are detected and repaired utilizing the ARPL approach. The specific repair process is introduced in Sect. 3.3.
- (4) Output the final result.

## 4. Experimental Results and Analysis

### 4.1 Experimental Environment and Dataset

The experimental running environment is a computer with Intel(R) Core(TM) i7-9700CPU @3.6 GHz and 16 GB memory. Written in Java language, the development platform is Eclipse, JDK1.8. The experimental process calls for some approaches of open-source toolkits, such as Jena, OWLAPI, and dictionary WordNet.

This paper adopts the experimental dataset provided by the OAEI, including the National Cancer Institute Thesaurus (NCI), Foundational Model of Anatomy (FMA), and the Systematized Nomenclature of Human and Veterinary Medicine (SNOMED) Clinical Terms. In addition, anatomical Adult Mouse Anatomy (MA) and part of the Describe Human Anatomy are utilized as experimental data. The number of concepts and standard matching contained in the experimental data are shown in Table 3.

### 4.2 Experiment Evaluation Index

OAEI provides the Unified Medical Language System (UMLS) as a reference standard for ontology matching and

repair. The evaluation indexes provided are the precision (P), recall (R), and F1-measure value (F1), which are calculated by the formula (8)–(10). In the logic repair stage, the number of unsatisfiable concepts (denoted as *Unsat.*) and the proportion of the number of unsatisfiable concepts (denoted as *Degree*) are used. The calculation formula is defined as formula (11).

$$P = \frac{|M \cap \text{Ref}|}{|M|} \quad (8)$$

$$R = \frac{|M \cap \text{Ref}|}{|\text{Ref}|} \quad (9)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (10)$$

$$\text{Degree} = \frac{\text{Unsat.}}{n + m} \quad (11)$$

Among them,  $M$  represents the matching result received after utilizing the matching approach or the repair result acquired after employing the matching or repair method.

## 4.3 Experimental Results and Analysis

### 4.3.1 Comparison and Analysis of Ontology Matching Methods

When experimented on the MA-Human dataset, the OMSA method is compared with the seven methods, which are LogMapLite [25], FCA-MAP-KG [26], DOME [7], AGM [8], Alin [27], Lily [28], and Wiktionary [29]. Meanwhile, in the three datasets of FMA-NCI, FMA-SNOMED, and SNOMED-NCI, the OMSA method is compared with the five methods of LogMapLite, FCA-Map-KG, DOME, AGM, and Wiktionary.

Since the DOME and AGM approaches have been introduced in Sect. 2, further elaboration is not repeated here. The remaining methods are described as follows:

- (1) LogMapLite is a lightweight matching system primarily employing string matching techniques, devoid of reasoning and repair operations.
- (2) The FCA-Map-KG approach is a knowledge graph matching system based on formal concept analysis, supporting the matching of large-scale and intricate biomedical ontologies.
- (3) The Alin approach utilizes WordNet and domain-specific ontologies to find synonyms between entities. Subsequently, methods like Jaccard are used to calculate character similarity between concepts.
- (4) The Lily method employs semantic subgraphs to grasp the true meaning of ontologies. It then measures character similarity between ontologies using a semantic description document matcher.
- (5) The Wiktionary method leverages synonym relationships between ontologies and incorporates Wiktionary as external background knowledge to assist in ontology matching.

The experimental results are shown in Tables 4 to 7.

From Tables 4 to 7, it can be seen that the OMSA approach achieves the best results in terms of recall and F1-measure value in the four datasets. Among them, the recall

**Table 3** Corresponding matching ontology and data information.

Source Ontology	Number of Concepts	Target Ontology	Number of Concepts	Standard match number
MA	2744	Human	3304	1489
FMA	3696	NCI	6488	2686
FMA	10157	SNOMED	13412	6026
SNOMED	51128	NCI	23958	17210

**Table 4** MA-Human dataset matching results.

	P	R	F1
LogMapLite	0.955	0.728	0.826
FCA-Map-KG	<b>0.984</b>	0.631	0.768
DOME	0.983	0.615	0.756
AGM	0.152	0.195	0.171
Alin	0.973	0.698	0.813
Lily	0.873	0.796	0.833
Wiktionary	0.958	0.730	0.829
OMSA	0.959	<b>0.924</b>	<b>0.941</b>

**Table 5** FMA-NCI dataset matching results.

	P	R	F1
LogMapLite	0.968	0.820	0.888
FCA-Map-KG	0.968	0.817	0.886
DOME	0.984	0.766	0.862
AGM	0.495	0.481	0.488
Wiktionary	<b>0.991</b>	0.608	0.754
OMSA	0.957	<b>0.894</b>	<b>0.924</b>

**Table 6** FMA-SNOMED dataset matching results.

	P	R	F1
LogMapLite	0.968	0.208	0.343
FCA-Map-KG	0.973	0.222	0.361
DOME	<b>0.988</b>	0.198	0.330
AGM	0.463	0.365	0.409
Wiktionary	0.965	0.170	0.289
OMSA	0.924	<b>0.727</b>	<b>0.813</b>

**Table 7** SNOMED-NCI dataset matching results.

	P	R	F1
LogMapLite	0.951	0.567	0.710
FCA-Map-KG	0.938	0.555	0.697
DOME	<b>0.978</b>	0.503	0.665
AGM	0.414	0.364	0.387
Wiktionary	0.967	0.471	0.633
OMSA	0.915	<b>0.707</b>	<b>0.797</b>

increased in the range of 7.4% to 72.9%; the F1-measure value improved in the range of 3.5% to 77%. As can be seen from the recall that the OMSA method can match more accurate matching pairs, which improves the matching efficiency. In terms of precision, the DOME method has certain advantages in the FMA-SNOMED and SNOMED-NCI

**Table 8** Unsatisfiable concept cases.

	Unsat.	Degree
MA-Human	33	0.546%
FMA-NCI	550	5.401%
FMA-SNOMED	2824	11.982%

**Table 9** Average consumption time (unit: seconds).

	MA-Human	FMA-NCI	FMA-SNOMED
OARPL	4.2	24.3	813.3
ARPL	<b>4.1</b>	<b>23.9</b>	<b>748.5</b>

datasets. While in the MA-Human and FMA-NCI datasets, the FCA-Map-KG and Wiktionary methods have the highest precision, respectively.

In this paper, semantic similarity and structural similarity are combined. First, WordNet dictionaries contain large amounts of existing knowledge information, which helps calculate the semantic similarity between entities. Then, potential matching pairs can be mined utilizing structural similarity. In general, the proposed OMSA method aims to explore the degree of association between concepts with similar meanings but distinct names. This approach effectively addresses similarity issues among synonyms, thereby enhancing the efficiency of ontology matching.

#### 4.3.2 Comparison and Analysis of Alignment Repair Methods

In the Alignment repair method based on the probabilistic logic (ARPL), the  $\epsilon$  value that limits the string length of matching pairs is added. In the original alignment repair method based on probability logic (original ARPL, OARPL [19]), the value of  $\epsilon$  is fixed to 25. Accordingly, this paper will compare the repair time of ARPL and OARPL methods.

Since the larger the number of unsatisfiable concepts, the higher the degree of conflict and the longer the repair time. The number and proportion of unsatisfiable concepts are shown in Table 8.

Table 9 shows the average time consumed by ARPL and OAPRL methods, where the ARPL method saves more time. In particular, on MA-Human, FMA-NCI, and FMA-SNOMED datasets, the proposed ARPL approach saves 0.1 s, 0.4 s, and 64.8 s, respectively. The experimental results show that with the increase in the number of unsatisfiable concepts in the ontology, the ARPL approach saves more time.

Figures 2 to 4 illustrate the specific time it takes for OAPRL and APRL methods to be repaired 10 times on the MA-Human, FMA-NCI, and FMA-SNOMET datasets. Figures 2 to 4 make the conclusions in Table 9 more accurate, reliable and persuasive.

From the experimental results, it can be seen that for the



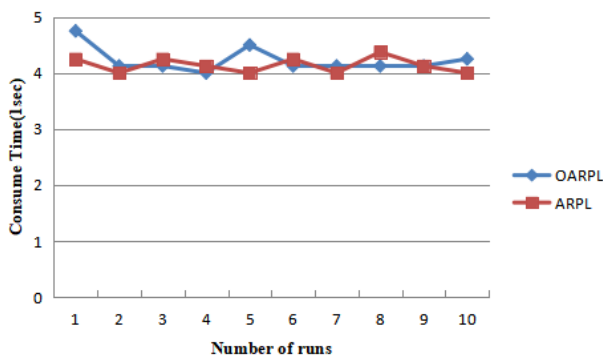


Fig. 2 MA-Human dataset repair time.

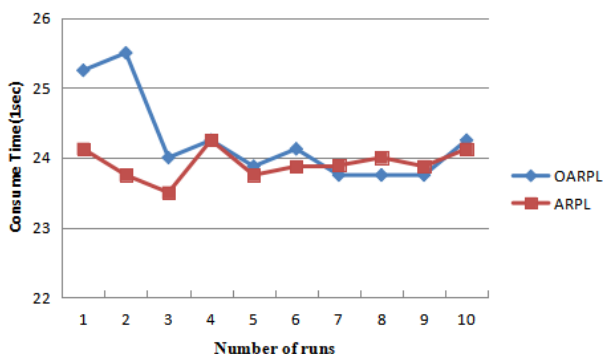


Fig. 3 FMA-NCI dataset repair time.

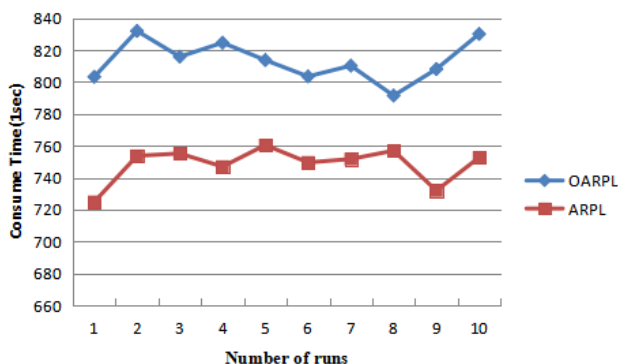


Fig. 4 FMA-SNOMED dataset repair time.

matching results with a low degree of conflict, the time consumption of the two methods of ARPL and OARPL are the same. Compared to the MA-Human dataset, as the number of unsatisfiable concepts in the FMA-NCI and FMA-SNOMED datasets increases, utilizing the ARPL method can save time. Specifically, as illustrated in Fig. 2, the OARPL method exhibits less repair time than the ARPL method proposed in this paper after the 8th run. Consequently, the average repair time values provided in Table 9 can better reflect the superiority of the ARPL method. Overall, the ARPL method proposed in this paper has certain advantages in repair.

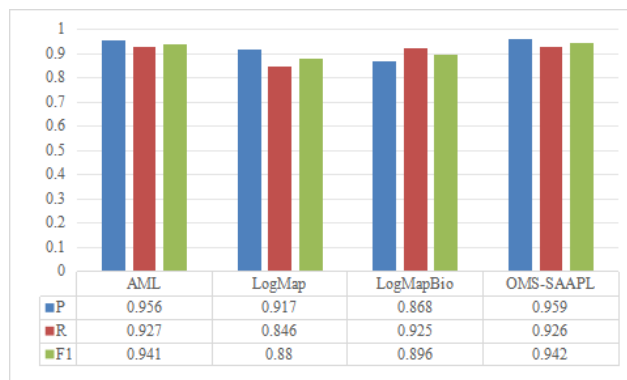


Fig. 5 MA-Human dataset matching and repair result.

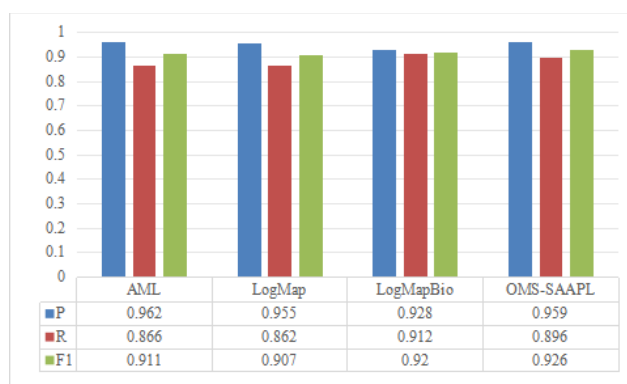


Fig. 6 FMA-NCI dataset matching and repair results.

### 4.3.3 Comparison and Analysis of Ontology Matching Systems

Consequently, in the comparative experiment of the matching system based on logical repair, the proposed OMS-SAAPL system and AML [5], LogMap [14], and LogMapBio [25] systems are compared on MA-Human, FMA-NCI, and FMA-SNOMED datasets.

Since the AML and LogMap approaches have been introduced in Sect. 2, further elaboration is not repeated here. The LogMapBio method is an extension of the LogMap system that utilizes BioPortal as an intermediate ontology to assist in matching. The experimental results are shown in Figs. 5 to 8.

According to the analysis of Figs. 5 to 8, the precision of the proposed OMS-SAAPL system on the MA-Human dataset is improved by 0.3%, 4.2%, and 9.1%, respectively, compared with the AML, LogMap, and LogMapBio systems. At the same time, F1-measure value also achieves the best results. In the FMA-NCI dataset, the OMS-SAAPL system increased the F1-measure value by 1.5%, 1.9%, and 0.6%, respectively, compared with the other three systems. In addition, the recall and F1-measure value ranked second in both FMA-SNOMED and SNOMED-NCI datasets. It can be seen that the OMS-SAAPL system has better stability than other systems.

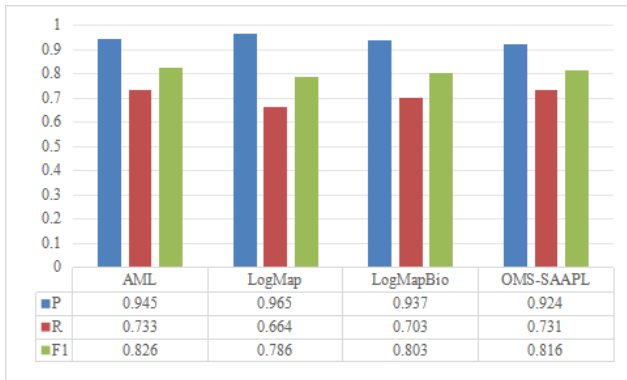


Fig. 7 FMA-SNOMED dataset matching and repair results.

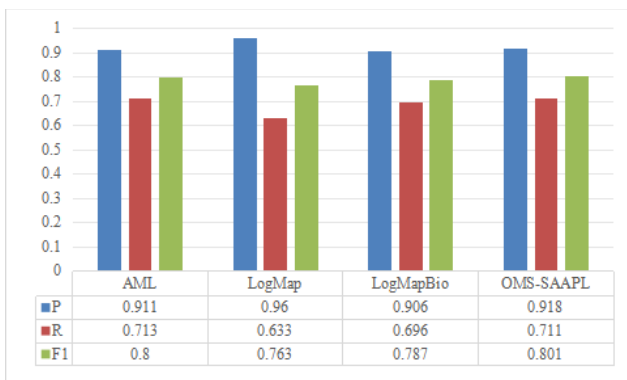


Fig. 8 SNOMED-NCI dataset matching and repair results.

## 5. Conclusions

If the existing knowledge conclusions can be utilized to reason and judge again, this can obtain more matching pairs containing potential semantic associations. Based on this idea, an ontology matching and repair method based on semantic association and probabilistic logic is proposed. This method combines semantics and contextual structure to discover more reliable matching pairs. Besides, to check the logical correctness of its matching results, a probabilistic logic method is introduced. Meanwhile, the parameter constraint is added to the concept name of the matching pair, which improves the reliability of the matching result.

In the future, related research will be explored in the following directions. (1) Currently, ontology matching studies mainly focus on the semantic similarity of terms and contextual structure, while ignoring the knowledge of attributes and instances. (2) Compared with the local matching method, the matching method that considers the global will obtain a higher number of matches. (3) Exclusively modifying or deleting matching pairs cannot solve the conflict problem. It is crucial to diagnose the semantic errors of the original knowledge and make adjustments and modifications.

## Acknowledgments

This research was supported by the National Natural Science Foundation of China under Grant No. 62267005; Chinese Guangxi Natural Science Foundation under Grant No. 2023GXNSFAA026493; Guangxi Collaborative Innovation Center of Multi-Source Information Integration and Intelligent Processing; Innovation Project of Guangxi Graduate Education No. YCSW2023437.

## References

- [1] W. Zhao, Z. Fu, T. Fan, and J. Wang, "Ontology construction and mapping of multi-source heterogeneous data based on hybrid neural network and autoencoder," *Neural Computing and Applications*, vol.35, no.36, pp.25131–25141, 2023.
- [2] L. Zhu, G. Hua, and W. Gao, "Mapping ontology vertices to a line using hypergraph framework," *Int. J. Cognitive Computing in Engineering*, vol.1, pp.1–8, 2020.
- [3] N. Ferranti, S.S.R.F. Soares, and J.F. de Souza, "Metaheuristics-based ontology meta-matching approaches," *Expert Systems with Applications*, vol.173, p.114578, 2021.
- [4] C. Trojahn, R. Vieira, D. Schmidt, A. Pease, and G. Guizzardi, "Foundational ontologies meet ontology matching: A survey," *Semantic Web*, vol.13, no.4, pp.685–704, 2022.
- [5] D. Faria, C. Pesquita, E. Santos, M. Palmonari, I.F. Cruz, and F.M. Couto, "The agreementmakerlight ontology matching system," *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*, pp.527–541, 2013.
- [6] L.T. Slater, W. Bradlow, S. Ball, R. Hoehndorf, and G.V. Gkoutos, "Improved characterisation of clinical text through ontology-based vocabulary expansion," *J. Biomedical Semantics*, Vol.12, no.7, 2021.
- [7] S. Hertling and H. Paulheim, "DOME results for OAEI 2019," *Proc. 18th Int. Semantic Web Conf.*, Auckland, New Zealand, 2019.
- [8] A. Lütke, "AnyGraphMatcher submission to the OAEI knowledge graph challenge 2019," *Proc. 18th Int. Semantic Web Conf.*, Auckland, New Zealand, 2019.
- [9] Z. Li, X. Liu, X. Wang, P. Liu, and Y. Shen, "Transo: A knowledge-driven representation learning method with ontology information constraints," *World Wide Web*, Vol.26, no.1, pp.297–319, 2023.
- [10] S. Ibrahim, S. Fathalla, J. Lehmann, and H. Jabeen, "Toward the multilingual semantic web: multilingual ontology matching and assessment," *IEEE Access*, vol.11, pp.8581–8599, 2023.
- [11] X. Xue, C. Yang, G. Mao, and H. Zhu, "Semi-automatic ontology matching based on interactive compact genetic algorithm," *Int. J. Pattern Recognition and Artificial Intelligence*, vol.36, no.05, 2257002, 2022.
- [12] M.C. Silva, D. Faria, and C. Pesquita, "Matching multiple ontologies to build a knowledge graph for personalized medicine," *European Semantic Web Conf., The Semantic Web*, pp.461–477, 2022.
- [13] H. Khan, M. Saqib, H.A. Khattak, S.I. Ali, and S. Lee, "Ontology alignment for accurate ontology matching: A survey," *Proc. 20th Int. Conf. Smart Homes and Health Telematics*, Wonju, South Korea, 2023.
- [14] E. Santos, D. Faria, C. Pesquita, and F.M. Couto, "Ontology alignment repair through modularization and confidence-based heuristics," *PLOS ONE*, vol.10, no.12, pp.1–19, 2015.
- [15] I.G. Husein, B. Sitohang, S. Akbar, and F.N. Azizah, "Heuristic based on dynamic weighting to support diagnosis with two minimization focus in alignment incoherence repair," *Int. J. Electrical Engineering and Informatics*, vol.12, no.1, pp.44–58, 2020.
- [16] W. Li, Q. Ji, S. Zhang, X. Fu, and G. Qi, "A graph-based method for interactive mapping revision in DL-Lite," *Expert Systems with Applications*, vol.211, p.118598, 2023.

- [17] L. van den Berg, M. Atencia, and J. Euzenat, "A logical model for the ontology alignment repair game," *Autonomous Agents and Multi-Agent Systems*, vol.35, no.2, pp.1–34, 2021.
- [18] D. Lembo, R. Rosati, V. Santarelli, D.F. Savo, and E. Thorstensen, "Mapping repair in ontology-based data access evolving systems," *Proc. Twenty-Sixth Int. Joint Conf. Artificial Intelligence*, 2017.
- [19] W. Li and S. Zhang, "Repairing mappings across biomedical ontologies by probabilistic reasoning and belief revision," *Knowledge-Based Systems*, vol.209, p.106436, 2020.
- [20] Y. He, J. Chen, D. Antonyrajah, and I. Horrocks, "BERTMap: A BERT-based ontology alignment system," *Proc. AAAI Conf. Artificial Intelligence*, vol.36, no.5, pp.5684–5691, 2022.
- [21] M.O. dos Santos, C.E.R. de Mello, and T.M. de Classe, "A useful tool to support the ontology alignment repair," *Brazilian Conf. Intelligent Systems*, pp.201–215, 2020.
- [22] C. Alrabbaa, F. Baader, R. Dachsel, T. Flemisch, and P. Koopmann, "Visualising proofs and the modular structure of ontologies to support ontology repair," *Description Logics*, Rhodes, Greece, 2020.
- [23] P. Lambrix, "Completing and debugging ontologies: State-of-the-art and challenges in repairing ontologies," *ACM J. Data and Information Quality*, vol.15, no.4, pp.1–38, 2023.
- [24] T. French, and T. Smoker, "An aleatoric description logic for probabilistic reasoning," *Proc. 34th International Workshop on Description Logics*, Bratislava, Slovakia, 2021.
- [25] E. Jiménez-Ruiz, B.C. Grau, A. Solimando, and V.V. Cross, "LogMap family results for OAEI 2015," *Proc. 14th Int. Semantic Web Conf.*, Bethlehem, PA, USA, 2015.
- [26] M. Zhao, and S. Zhang, "Identifying and validating ontology mappings by formal concept analysis," *Proc. 15th Int. Semantic Web Conf.*, Kobe, Japan, 2016.
- [27] J. Da Silva, K. Revoredo, F. Baião, and J. Euzenat, "Alin: improving interactive ontology matching by interactively revising mapping suggestions," *The Knowledge Engineering Review*, vol.35, 2020, e1.
- [28] Y. Hu, S. Bai, S. Zou, and P. Wang, "Lily results for OAEI 2020," *Proc. 19th Int. Semantic Web Conf.*, Athens, Greece, 2020.
- [29] J. Portisch and H. Paulheim, "Wiktionary matcher results for OAEI 2021," *Proc. 20th Int. Semantic Web Conf.*, Virtual conference, 2022.



**Mei Chen** is currently a graduate student at the School of Computer & Information Engineering, Nanning Normal University, Nanning, Guangxi, China. Her research interests include Knowledge Graphs and Graph Mining.



**Ying Pan** received her Ph.D. degree from Sun Yat-sen University in 2011. Currently, she is a professor at the School of Computer & Information Engineering, Nanning Normal University, Guangxi, China. Her research interests include Semantic Web, Graph Databases, and Intelligent Computing.



**Nan Wu** is currently a graduate student at the School of Computer & Information Engineering, Nanning Normal University, Nanning, Guangxi, China. Her research interests include Semantic Web and Graph Mining.



**Xiaocong Lai** is currently a graduate student at the School of Computer & Information Engineering, Nanning Normal University, Nanning, Guangxi, China. Her research interests include Semantic Web and Graph Data Management.