Pivot-based Candidate Retrieval for Cross-lingual Entity Linking

Qian Liu*
Australian Artificial Intelligence Institute (AAII)
University of Technology Sydney, Australia
Qian.Liu-9@student.uts.edu.au

Xiubo Geng†
STCA NLP Group, Microsoft
xigeng@microsoft.com

Jie Lu
Australian Artificial Intelligence Institute (AAII)
University of Technology Sydney, Australia
Jie.Lu@uts.edu.au

Daxin Jiang†
STCA NLP Group, Microsoft
djiang@microsoft.com

ABSTRACT

Entity candidate retrieval plays a critical role in cross-lingual entity linking (XEL). In XEL, entity candidate retrieval needs to retrieve a list of plausible candidate entities from a large knowledge graph in a target language given a piece of text in a sentence or question, namely a mention, in a source language. Existing works mainly fall into two categories: lexicon-based and semantic-based approaches. The lexicon-based approach usually creates cross-lingual and mention-entity lexicons, which is effective but relies heavily on bilingual resources (e.g. inter-language links in Wikipedia). The semantic-based approach maps mentions and entities in different languages to a unified embedding space, which reduces dependence on large-scale bilingual dictionaries. However, its effectiveness is limited by the representation capacity of fixed-length vectors. In this paper, we propose a pivot-based approach which inherits the advantages of the aforementioned two approaches while avoiding their limitations. It takes an intermediary set of plausible target-language mentions as pivots to bridge the two types of gaps: cross-lingual gap and mention-entity gap. Specifically, it first converts mentions in the source language into an intermediary set of plausible mentions in the target language by cross-lingual semantic retrieval and a selective mechanism, and then retrieves candidate entities based on the generated mentions by lexical retrieval. The proposed approach only relies on a small bilingual word dictionary, and fully exploits the benefits of both lexical and semantic matching. Experimental results on two challenging cross-lingual entity linking datasets spanning over 11 languages show that the pivot-based approach outperforms both the lexicon-based and semantic-based approach by a large margin.

CCS CONCEPTS

• **Information system** → **Information extraction**.

*Work done during internship at STCA NLP Group, Microsoft.
†Corresponding authors.

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-8312-7/21/04.
https://doi.org/10.1145/3442381.3449852

KEYWORDS

Information extraction, entity linking, cross-lingual retrieval

ACM Reference Format:
Qian Liu, Xiubo Geng, Jie Lu, and Daxin Jiang. 2021. Pivot-based Candidate Retrieval for Cross-lingual Entity Linking. In *Proceedings of the Web Conference 2021 (WWW '21), April 19–23, 2021, Ljubljana, Slovenia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3442381.3449852

1 INTRODUCTION

Entity Linking [12] is the task which associates mentions in a sentence with their corresponding entities in a knowledge base. Considering the diversity of languages used on the web, cross-lingual entity linking (XEL) [25, 39] where the sentences are in a source language different from the knowledge base language has attracted wide attention recently. XEL is an important component task for many downstream tasks, such as cross-lingual knowledge-based question answering [37], cross-lingual information extraction [38], etc.

Typically, XEL consists of two steps: (1) *candidate retrieval*, which retrieves a small subset (e.g. 1000) of plausible candidates from a large set of KB entries in the target language (e.g. 6 million English entities in DBPedia); and (2) *entity disambiguation*, which re-ranks the selected candidates and returns the most likely entities. Candidate retrieval plays a critical role for cross-lingual entity linking, since missing entities in this step will never be recovered by the downstream disambiguation step. Nevertheless, the quality of candidate retrieval under a cross-lingual setting is far from complete. For example, as illustrated in Zhou et al. [39], a recall of retrieved candidates can reach over 80% for English mentions with the help of a Wikipedia mention-entity dictionary, while that of the state-of-the-art method is only 40% for mentions in Telugu (a Dravidian language spoken in southeastern India). The low-quality of the candidate retrieval step is gradually becoming a key obstacle in the XEL task.

In general, candidate retrieval for monolingual entity linking suffers from *mention-entity gap*, because surface forms of entities often differ from mentions. For example, the mention *Einstein* is linked to the entity `Albert_Einstein`. For XEL tasks, candidate retrieval is also hindered by *cross-lingual gap*, since the source and target languages are in different scripts. For example, *Manhattan Bridge* refers to *pont de Manhattan* in Spanish. To fill these two gaps,

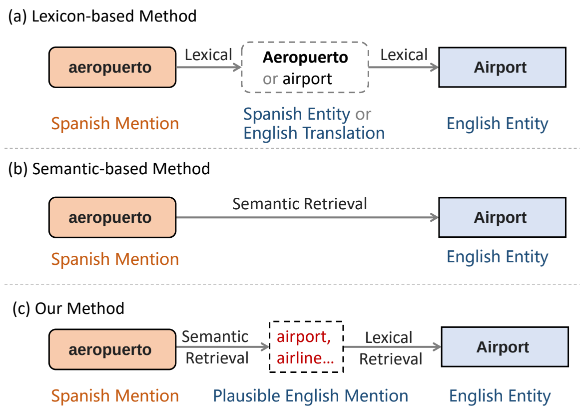

Figure 1: Comparison of lexicon-based method, semantic-based method, and our pivot-based method.

existing works mainly take two types of approaches: lexicon-based and semantic-based approaches.

The lexicon-based approach usually creates lexicons to bridge both gaps with Wikipedia resources. For example, Pan et al. [23] proposed to map source-language mentions to target-language ones using inter-language links, and then retrieved candidate entities from an English mention-entity dictionary. A major problem with such approach is that it heavily relies on Wikipedia inter-language links, which can only cover a small percentage of target-language entities, and this problem is especially severe for low-resource languages.

The semantic-based method generates candidate entities by leveraging the cross-lingual semantic retrieval. It usually builds an aligned embedding space between the source-language mentions and target-language entities, where synonyms of different languages have similar embeddings, and candidate entity retrieval can be undertaken by searching nearest neighbors of each mention in the embedding space of target entities. However, a single low-dimensional embedding has limited representation capacity for mentions or entities, and tends to lose lexical matching information which is critical to retrieval [11]. For example, for the French mention *pont de Manhattan* (*Manhattan Bridge* in English), the semantic-based approach tends to retrieve different kinds of bridges, such as *Belmont Avenue Bridge in Philadelphia, Bridges of the Merritt Parkway*, which successfully captures the keyword *bridge* while ignores the other one *Manhattan.*

In this work, we propose a pivot-based approach for the cross-lingual entity candidate retrieval task, which fully explores the advantages of both lexicon-based and semantic-based retrieval and avoids their limitations. In one aspect, it is usually difficult to derive an inter-language lexicon with high coverage. However, there is relatively large volume of monolingual data for both source and target languages, which can be fully leveraged by pre-trained models to map words into embeddings. Furthermore, with only a small set of bilingual word pairs, cross-lingual alignment can easily map word embeddings from one language to those in another language. Therefore, our approach first converts source-language mentions into an intermediary set of plausible target-language mentions with word-level cross-lingual semantic retrieval and a selective mechanism. In another aspect, there is usually rich lexicons such as alias or anchor texts to bridge the gap between entities and mentions in the target language. Therefore, our approach further conducts lexical retrieval with the generated intermediary target-language mentions.

We illustrate the difference among lexicon-based, semantic-based, and our pivot-based approach in Figure 1. Compared to lexicon-based approach, the proposed pivot-based method does not rely on Wikipedia inter-language links, and it fully leverages pre-trained word embeddings and only needs a small set of seed bilingual word pairs to learn cross-lingual alignment. Compared to semantic-based approach, our method converts a source-language mention into a diverse intermediary set of plausible target-language mentions with a flexible selective mechanism, and fully leverages the rich lexical resources of target-language knowledge base, and thus can retrieve more diverse and accurate candidates.

Merhav and Ash [20] suggested that most of the candidate generation methods perform well on the Wikipedia-based datasets but fail to generalize beyond Wikipedia because they rely heavily on Wikipedia resources (e.g. inter-language links). Therefore, we evaluate the proposed method on two XEL entity linking datasets, QALD which contains non-Wikipedia questions in 8 languages and WIKI-LRL which contains Wikipedia titles in 3 low-resource languages. Experimental results show that it outperforms both the lexicon-based and semantic-based approach by a large margin. The source code of our method is available in the GitHub[1].

The main contributions of this work are summarized as follows.

- We propose a pivot-based candidate retrieval framework for XEL, which jointly leverages semantic retrieval and lexical retrieval.
- We emphasize the importance of leveraging English pivots to bridge the cross-lingual and mention-entity gaps.
- We perform extensive experiments on both non-Wikipedia question-based and Wikipedia-related sentence-based XEL datasets. Experimental results demonstrate the effectiveness of our method for both cross-lingual entity candidate retrieval and the end-to-end entity linking task.

2 RELATED WORKS

In this section, we introduce representative candidate retrieval methods for XEL.

Lexicon-based methods. For the monolingual candidate retrieval task, candidate retrieval mainly relies on string matching or mention-entity lexicons [2, 10, 16, 36]. For a cross-lingual entity linking task, Wikipedia inter-language resources are employed to fill the cross-lingual gap, such as parallel Wikipedia titles, inter-language entity links. Several lexicon-based candidate retrieval methods have been widely-used in existing state-of-the-art XEL systems [28, 29, 32]. For example, Tsai and Roth [30] build a direct probabilistic mapping table using parallel Wikipedia titles and the anchor text mappings, between the source-language and English. It first extracts a source-language mention-entity map from anchor-text mapping in Wikipedia pages. Then, the source-language entity is redirected to its corresponding English entity using the Wikipedia

[1]https://github.com/qianliu0708/PivotsCR

inter-language links. Pan et al. [23] and Zhang et al. [38] proposed to induce word-by-word translations using parallel Wikipedia titles, and used the translated mention to retrieve candidate entities from an existing English mention-entity map. This improved method reduces reliance on source-language anchor-text mapping. Lexicon-based methods are effective for high-resource languages, such as Spanish, but they rely heavily on the coverage of Wikipedia resources, resulting in restrictions on low-resource languages.

Semantic-based methods. Word semantic representation methods [19, 21], which encode meanings of words to low dimensional vector spaces, have become very popular in natural language processing and information retrieval, such as query expansion [18] and text classification [17]. Recently, pre-trained multilingual word representations [1, 4, 15] have been employed to bridge the cross-lingual gap. These methods learn a mapping function to align the source and target embedding space, where synonyms of different languages have similar embeddings. The mentions and entities are represented as fixed-length vectors. Candidate entities retrieval can be undertaken by searching the nearest neighbors of each mention in the embedding space. However, a single low-dimensional embedding has limited representation capacity for mentions or entities [11]. Moreover, powerful pre-trained language models (e.g., Multilingual-BERT) have powerful representation capacities, but they are cost-prohibitive for the candidate retrieval step.

Pivoting language methods. These methods improve the performance of candidate retrieval for low-resource languages (LRL) using a closely related high-resource language (HRL) as an intermediate pivot. For example, *Poland* in Marathi and Hindi are written similarly, and Hindi can be used as a pivoting language for Marathi. Rijhwani et al. [25] train a neural character level string matching model to encode the LRL mentions by leveraging HLR training data. Zhou et al. [39] show that the character-level string matching can be further improved with character n-gram information [34] and extending entity-entity pairs with mention-entity pairs in the training process.

Transliteration methods. These methods are employed when the source-language and English word pairs have similar pronunciation. For example, Upadhyay et al. [33] use a sequence-to-sequence model and a bootstrapping method to transliterate low-resource entity mentions using extremely limited training data. Tsai and Roth [31] combine the standard translation method for candidate retrieval with a transliteration score to improve candidate recall.

Different from the previous methods, our method jointly leverages semantic retrieval and lexical retrieval to search candidate entities for source-language mentions. We learn an intermediary collection with several plausible English mentions to fill the cross-lingual gap and mention-entity gap.

3 TASK DESCRIPTION

Cross-lingual entity linking aims to link mentions in a source language to entities in a knowledge base which is written in a target language. It usually consists of two steps: candidate retrieval and entity disambiguation. In this work, we mainly focus on the candidate retrieval component, which plays a critical role in cross-lingual entity linking. For a better understanding, we elaborate on the terminology and corresponding examples in Table 1. Formally, given

Table 1: Terminology and the corresponding description and examples used in the cross-lingual candidate retrieval task.

Term	Description	Examples
Source language	the language of the text to be linked to KB	French
Target language	the language of the used structural KB	English
Mention	a piece of text in a sentence/question to be linked to KB	pont de Manhattan
Gold Entity	the correct entity in KB for the mention	Manhattan_Bridge
Candidate Entity	retrieved entity from KB for the mention	Bridges_of_Dee

a set of source-language mentions $\mathcal{M} = \{m_1, m_2, \cdots, m_{|\mathcal{M}|}\}$ and a target-language knowledge base $\mathbf{K}$ which contains millions of entities, the goal of candidate retrieval is to retrieve a list of candidate entities $\mathcal{E}_i = \{e_{i1}, e_{i2}, \cdots, e_{iN}\}$ from $\mathbf{K}$ for each mention $m_i \in \mathcal{M}$, where N is the size of each candidate list.

As the final results of XEL are only generated from candidate entities in $\mathcal{E}$, the candidate list should be as comprehensive as possible to ensure that gold entities are included. Therefore, candidate retrieval methods are measured by *recall*, which is the percentage of retrieved candidate lists that contain corresponding gold entities. Suppose the gold entity of mention m_i is $\hat{e}_i$, *Recall@N* is defined as,

$$Recall@N = \frac{\sum_{i=1}^{|\mathcal{M}|} I(\hat{e}_i \in \mathcal{E}_i)}{|\mathcal{M}|}, \quad (1)$$

where $I(\cdot)$ is the indicator function which is set to 1 if true else 0, $|\mathcal{M}|$ is the number of mentions, and N is the number of candidate entities in the retrieved list $\mathcal{E}_i$.

4 METHODOLOGY

To bridge these two gaps, the key idea of our method is to learn an intermediary collection of target-language words which are semantically similar to the source language mention and lexically similar to the target-language gold entity. Figure 2 illustrates our method. The proposed method consists of three stages.

- First, we generate an initial intermediary collection of target-language words using cross-lingual semantic representations. It fills the cross-lingual gap and does not rely on Wikipedia bilingual resources [23, 29], such as anchor-text links and inter-language links. In addition, high-quality and publicly available multilingual word representations, such as MUSE [15], have a better ability than bilingual lexicons to find a comprehensive collection of related words.
- Second, we design a selective mechanism to refine the initial intermediary collection. The goal is to alleviate the duplication and coverage issue, and thus empower the following lexical search to retrieve a more comprehensive set of candidates.
- Third, we fill the mention-entity gap using lexical retrieval. Each mention is represented as target-language string queries based on the intermediary collection, and the lexical retrieval model uses string overlap information to score mention-entity pairs.

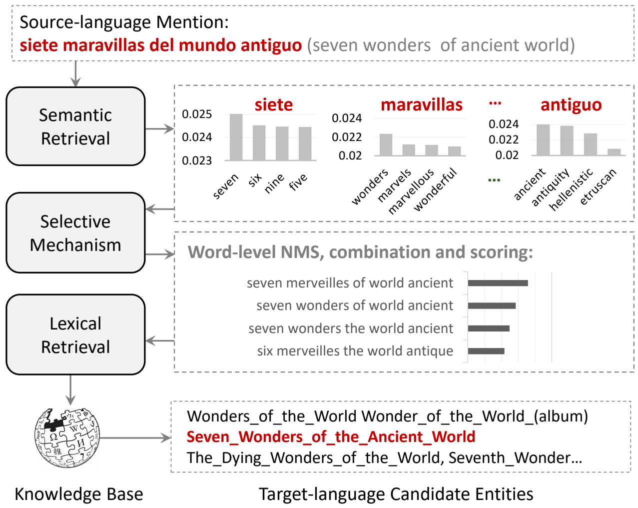

Figure 2: An example to illustrate our pivot-based approach.

The main contribution of this work lies in the framework which effectively combines the advantage of semantic-based and lexical-based retrieval, and a flexible selective mechanism in the framework which can alleviate the duplication and coverage issues.

For the sake of convenience, we assume the source language is Spanish and the target language is English to illustrate our method in the following section.

4.1 Filling the Cross-lingual Gap

Given a Spanish mention $m = \{x_1, x_2, \cdots, x_k\}$ which contains k words, we first generate a set of English words as the intermediary collection $\mathcal{P}$, by searching the English vocabulary. The collection $\mathcal{P}$ aims to represent the semantics of m as comprehensively and accurately as possible to bridge the gap between source and target languages.

Inspired by Lample et al. [15], we employ bilingual word-by-word induction with the help of cross-lingual word embeddings. This process involves (1) aligning source and target embedding spaces and (2) retrieving English words for each Spanish word x_i in m.

Let $\mathcal{X}$ and $\mathcal{Y}$ be the Spanish and English embedding spaces[2], respectively. We learn a mapping $\mathbf{W} \in \mathbb{R}^{d \times d}$ from $\mathcal{X}$ to $\mathcal{Y}$ to align the two spaces, with the objective that synonyms have similar representations. Concretely, we use a seed dictionary of l pairs of words $\{x_i, y_i\}_{i \in \{1,l\}}$, and learn the linear mapping by optimizing,

$$\mathbf{W}^* = \underset{\mathbf{W} \in \mathbb{R}^{d \times d}}{\arg\min} \ \|\mathbf{WX} - \mathbf{Y}\|_{\mathrm{F}}, \tag{2}$$

where d is the dimension of the embeddings, $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{d \times l}$ are corresponding word embeddings of word pairs in the seed dictionary, and $\|\cdot\|_{\mathrm{F}}$ indicates the Frobenius norm. To improve the performance, following Xing et al. [35], we impose an orthogonality constraint on $\mathbf{W}$, i.e., $\mathbf{W}\mathbf{W}^\top = \mathbf{W}^\top\mathbf{W} = \mathbf{I}$. The optimization of $\mathbf{W}$ corresponds to the singular value decomposition (SVD) of $\mathbf{Y}\mathbf{X}^\top$,

$$\begin{aligned} \mathbf{W}^* &= \mathbf{U}\mathbf{V}^\top, \\ \text{with } \mathbf{U}\Sigma\mathbf{V}^\top &= \mathbf{SVD}\,(\mathbf{Y}\mathbf{X}^\top). \end{aligned} \tag{3}$$

Then, we retrieve English words for each Spanish x_i in mention m. Specially, x_i is represented by applying projection matrix $\mathbf{W}$ on its Spanish embedding $\mathbf{x}_i$, as $\mathbf{x}_i^* = \mathbf{W}\mathbf{x}_i$. Next, we explore the nearest English words to $\mathbf{x}_i^*$ in $\mathcal{Y}$. To measure the similarity between Spanish word x_i and each English word y we use the cross-domain similarity local scaling metric (CSLS),

$$\mathrm{CSLS}(\mathbf{W}\mathbf{x}_i, \mathbf{y}) = 2\cos(\mathbf{W}\mathbf{x}_i, \mathbf{y}) - r_{\mathcal{Y}}(\mathbf{W}\mathbf{x}_i) - r_{\mathcal{X}}(\mathbf{y}). \tag{4}$$

Here $\mathbf{y} \in \mathbb{R}^d$ denotes the embedding of word y in $\mathcal{Y}$. $r_{\mathcal{Y}}(\mathbf{W}\mathbf{x}_i)$ is the mean similarity of $\mathbf{x}_i$ to its K neighborhoods in $\mathcal{Y}$,

$$r_{\mathcal{Y}}(\mathbf{W}\mathbf{x}_i) = \frac{1}{K} \sum_{\mathbf{y}' \in \mathcal{N}_{\mathcal{Y}}(\mathbf{W}\mathbf{x}_i)} \cos(\mathbf{W}\mathbf{x}_i, \mathbf{y}'), \tag{5}$$

where $\cos(\cdot)$ denotes the cosine similarity, $\mathcal{N}_{\mathcal{Y}}(\mathbf{W}\mathbf{x}_i)$ is the K neighborhoods associated with $\mathbf{W}\mathbf{x}_i$ in $\mathcal{Y}$. Similarly, $r_{\mathcal{X}}(\mathbf{y})$ denotes the mean similarity of a target word y to its neighborhoods. We refer readers to Johnson et al. [14] and Lample et al. [15] for more details. We employ CSLS here because it significantly increases the accuracy of word retrieval and does not require any parameter tuning.

We select K English words for each Spanish word x_i in mention m, and combine them as the intermediary collection, i.e., $\mathcal{P}(m) = \{y_{1,1}, y_{1,2}, \cdots, y_{1,K}, \cdots, y_{k,1}, y_{k,2}, \cdots, y_{k,K}\}$. Each English word $y_{i,j} \in \mathcal{P}(m)$ is assigned with a score, i.e., $\mathrm{CSLS}(\mathbf{W}\mathbf{x}_\mathbf{i}, \mathbf{y}_\mathbf{i,j})$.

Moreover, in order to alleviate the out-of-vocabulary problem for Spanish word embedding, we also employ multilingual character embeddings [25] to estimate the similarity between x_i and each English word y_j, and retrieve x_i's K most similar English words. We detail the multilingual character embedding training and retrieval, and evaluate its effectiveness in Section 5.4.1.

Compared to the lexicon-based approach, our method only relies on a small bilingual dictionary (around 5K word pairs) to align the source and target embedding spaces.

4.2 Selective Mechanism

The initial intermediary collection $\mathcal{P}$ suffers from duplication and coverage issues in its role to connect the Spanish mention and English candidate entities. For example, the top-5 retrieved English words for the Spanish word *maravillas* (*wonders* in English) are {*miracle*, *miracles*, *miraculous*, *miraculously*, *wonderful*}. The duplication issue arises because multiple words have the same meaning with different morphologies, leading to a large number of the same candidate entities appearing repeatedly in the downstream retrieval. The coverage issue arises because some important words with lower similarity are ignored, e.g., the word *wonders* is excluded in $\mathcal{P}$(*marvillas*). The low diversity of intermediate sets may result in incomplete candidate entities.

To alleviate these issues, we employ a selective mechanism to refine the intermediary collection. Inspired by the non-maximum suppression (NMS) algorithm [27] that is used to prune redundant bounding boxes in object detection [24] and candidate answer spans in machine reading comprehension [13], we design a word-level

[2] In our method and experiments, we employ the fastText to train monolingual word embeddings: https://fasttext.cc/docs/en/crawl-vectors.html

NMS to prune morphological variations and improve diversity. Given the initial intermediary collection $\mathcal{P}(x_i) = \{y_1, y_2, \cdots, y_k\}$, after selecting the word y_a which possesses the maximum score, we remove it from the set $\mathcal{P}(x_i)$ and add it to $\mathcal{P}_{NMS}(x_i)$, and delete any y_n in $\mathcal{P}(x_i)$ that is a duplication to y_b. We define that two words are duplicates of each other if they are the same after stemming. This process is repeated for the remaining words in $\mathcal{P}(x_i)$, until $\mathcal{P}(x_i)$ is empty or the size of $\mathcal{P}_{NMS}(x_i)$ reaches a maximum threshold T_w. Algorithm 1 details the word-level NMS method.

Algorithm 1: Word-level NMS

```
Input: P(x_i) = {y_1, ··· , y_k}; S(x_i) = {s_1, ··· , s_k}; T_w
    P(x_i) is the set of candidate translations
    S(x_i) is the set of corresponding scores for word in P(x_i)
    T_w denotes the maximum size threshold
Initialize P_NMS(x_i) = {}
while P(x_i) ≠ {} and len(P_NMS(x_i)) ≤ T_w do
    s_a = arg max S
    P_NMS(x_i) = P_NMS(x_i) ∪ {y_a}
    P(x_i) = P(x_i) − {y_a}
    S(x_i) = S(x_i) − {s_a}
    for y_b ∈ P(x_i) do
        if stem(y_a)==stem(y_b) then
            P(x_i) = P(x_i) − {y_b}; S(x_i) = S(x_i) − {s_b}
        end
    end
end
Return P_NMS(x_i)
```

Next, we use the softmax function to normalize the word scores in $\mathcal{P}_{NMS}(x_i) = \{y_1, \cdots, y_{T_w}\}$. For mention $m = \{x_1, x_2, \cdots, x_k\}$ with k words, we generate all T_w^k combinations[3]. We denote these combinations as *plausible English mentions* because they may be out of word order. For each plausible English mention we denote its relevance score to the original Spanish mention m as the averaged score of words in it, and T_m plausible English mentions with the highest scores are selected in the final intermediary collection $\mathcal{P}(m)$.

Equipped with the selective mechanism, semantic retrieval is capable of generating diverse English words which are related to the original Spanish word, and avoids the vocabulary mismatch problem from which bilingual lexicon-based methods suffer.

4.3 Filling the Mention-Entity Gap

Given the final intermediary collection of plausible English mentions, we search the candidate entities from the knowledge base using each element of the collection.

We first construct a search space with all the entities in the knowledge base. Each entity is represented by splitting its surface string into words and converted to lowercase. For example, `Manhattan_Bridge` is converted to *manhattan bridge*, `ChessPlayer` is converted to *chess player*. The lexical retrieval model uses word overlap information to score query-entity pairs. We use BM25 [26] to generate the query-entity score based on query statistics and entity statistics. The lexical matching score of a plausible English mention q and an entity e is defined as,

$$lex\_score(q, e) = Sim(q, m) \cdot \text{BM25}(q, e), \tag{6}$$

where $Sim(q, m)$ is the relevance score of plausible English mention q to its original Spanish mention m. The top N entities are selected as the candidate entities according to their lexical score.

In the process of bridging mention-entity gap, our method is flexible compared with hard matching methods using anchor-text links. It also runs quickly to search the whole entity space because statistics-based lexical retrieval is more efficient than the high dimensional vector retrieval used in semantic-based methods.

5 EXPERIMENTS

5.1 Datasets

We evaluate our method on the following two cross-lingual entity linking datasets, spanning 11 languages.

- **QALD**: We collect cross-lingual entity linking data from the multilingual QALD dataset[4], which is a benchmark for the task of cross-lingual question answering over knowledge base (KBQA). The first step of KBQA is XEL, which links *mentions* in other languages to their corresponding entities in the English KB. Each item in this dataset contains a *question*, *mentions* in this question, and the *SPARQL* to answer this question. We extract gold entities of mentions from the *SPARQL* query. The used knowledge base is DBpedia[5], with 6 million entities. Specifically, we merge all multilingual QALD data, from QALD-4 to QALD-9, and filter out questions whose SPARQL cannot be executed in this knowledge base. For the remaining data, we collect all mentions and their corresponding gold entities to perform the candidate retrieval task. These mentions are from eight languages, namely German, French, Russian, Spanish, Italian, Dutch, Romanian, and Portuguese. We released the used QALD data in our experiment on Github[6].
- **WIKI-LRL**: This is a cross-lingual entity linking dataset[7] for low-resource languages (LRL) collected by Zhou et al. [39]. The knowledge is Wikipedia. The candidate retrieval is conducted on 2 million entities of proper nouns in Wikipedia. The mentions are in three low-resource languages, namely Marathi (Indo-Aryan language spoken in Western India, written in Devanagari script), Lao (a Kra-Dai language written in Lao script), and Telugu (a Dravidian language spoken in southeastern India written in Telugu script).

In our experiments, we compare our methods with other candidate retrieval methods on these two challenging datasets. Previous works [20] show that most of the candidate retrieval methods perform well on the Wikipedia-based dataset but fail to generalize beyond Wikipedia, to news and social media text. For a more convincing evaluation, we collect the QALD dataset where mentions are extracted from the user's short search question. Moreover, the

[3]Note we set $T_w = 10$ and usually $k <= 2$, so there are only about 100 combinations. So the time cost for this step is very small.

[4]The dataset is available on https://github.com/ag-sc/QALD.
[5]We use the DBpedia 16-10 version: https://wiki.dbpedia.org/downloads-2016-10
[6]https://github.com/qianliu0708/PivotsCR/tree/main/QALD_data
[7]This dataset is available in https://github.com/shuyanzhou/pbel_plus.

Table 2: Top-1000 recall (R@1000) of different methods on the QALD dataset. #Mentions denotes the number of mentions for each language in QALD.

Languages (#mentions)	German (672)	French (672)	Russian (309)	Spanish (621)	Italian (672)	Dutch (621)	Romanian (615)	Portuguese (309)	Average
TRANS-Match	0.525	0.365	0.375	0.422	0.451	0.514	0.514	0.434	0.450
TRANS-Search	0.609	0.588	0.458	0.562	0.570	0.607	0.486	0.553	0.554
SemSearch	0.579	0.484	0.518	0.507	0.540	0.452	0.512	0.489	0.510
Spotlight	0.430	0.342	0.346	0.396	0.374	0.443	-	0.469	0.400
TagMe	0.338	-	-	0.316	-	-	-	-	0.327
OurMethod	0.824	0.801	0.722	0.815	0.799	0.828	0.828	0.812	0.804

existing low-resource XEL performance still lags far behind its high-resource counterparts [39]. We use the low-resource WIKI-LRL dataset to evaluate the robustness of our method to low-resource scenarios.

5.2 Baselines

We compare our method with the following five candidate retrieval methods, including lexicon-based methods and semantic-based methods.

- **TRANS** [23]: This is the most widely used lexicon-based candidate retrieval method for state-of-the-art XEL systems such as XELMS [32]. It translates the source-language mention into English in order to predict the entity link. Following Rijhwani et al. [25], we generate a bilingual lexicon with word alignments on parallel Wikipedia titles[8] using `fast_align` [8], which is a fast and unsupervised word aligner. Each word in the source-language mention is translated into English words using the lexicon. Then we experiment with two varieties to generate candidate entities. **Match** employs the English mention-entity lookup table[9] to generate candidate entities. **Search** utilizes the translated mention as a query and generates candidate entities by a lexical search of the entity space.
- **SemMatch** [15]: This is a semantic-based candidate retrieval method, leveraging cross-lingual word embeddings [4]. Following Pan et al. [22], we convert source-language mentions and target-language entities as fixed-length vectors in an aligned embedding space. We use the approximate nearest neighbors search tool to generate candidate entities. We use MUSE[10] to learn the aligned multilingual word embeddings. Each mention and entity are represented as averaged vector of words it contains. It is notable that some aggregation methods (such as BiLSTM and Transformer) are more powerful, however they are too complex for large-scale entity representation and retrieval to be feasible.
- **Spotlight** [5]: This is a publicly available tool[11] to automatically annotate mentions of DBpedia resources in text, providing a solution for linking unstructured information sources to the structural DBpedia. In our experiment, we use the pyspotlight[12], which is a thin python wrapper around the DBperdia Spotlight and supports ten languages including German, Dutch, French, Italian, and Spanish.
- **TagMe** [9]: This is a fast tool[13] to efficiently and judiciously augment plain text with the corresponding entities in Wikipedia. It is available in English, German and in Italian. We use the tagme-python version[14] in our experiment.
- **PBEL** [25]: This is a pivot-based entity linking for low-resource language (LRL) tasks. It performs cross-lingual string matching based on an entity gazetteer between a related high-resource language and English. This method removes reliance on the resource of LRL, and achieves state-of-the-art for candidate retrieval in low-resource XEL. In our experiment, we compare our method with PBEL on the WIKI-LRL dataset.

5.3 Main Results

5.3.1 Comparison on QALD. We first conduct the evaluation of different candidate retrieval methods on the QALD dataset. Table 2 shows the overall performance of our method as well as the baseline methods on the QALD dataset. The gold entity recall of top-1000 (R@1000) candidate entities is reported. We observe that,

- our method performs the best compared with the baseline methods mainly because it leverages both semantic matching and lexical matching information.
- our method and TRANS-Search both use lexical retrieval to generate candidates from the entity space. Our method significantly outperforms TRANS-Search, which implies that the plausible English mentions generated in our method perform much better than the lexicon generated from parallel Wikipedia titles. This indicates that semantic matching information is helpful in candidate retrieval for XEL. TRANS-Search performs slightly better than TRANS-Match, indicating lexical retrieval is more effective than a lookup table.
- the SemSearch method also employs semantic retrieval to fill the cross-lingual gap. It performs worse than our method mainly because a low-dimensional vector is not so accurate enough to represent a mention or an entity, resulting in an

[8]The parallel Wikipedia titles are available in https://linguatools.org/tools/corpora/wikipedia-parallel-titles-corpora/.
[9]https://github.com/dbpedia/lookup
[10]https://github.com/facebookresearch/MUSE
[11]https://www.dbpedia-spotlight.org/
[12]https://github.com/ubergrape/pyspotlight
[13]The official TagMe API: https://tagme.d4science.org/tagme/.
[14] https://github.com/marcocor/tagme-python

Table 3: Comparison of different methods in terms of average recall on QALD dataset. CR denotes the candidate retrieval in XEL. ED denotes entity disambiguation on the top-1000 candidate entities.

	Avg.	TRANS-Search	SemSearch	Ours
CR	**R@50**	0.381	0.408	0.544
	R@200	0.436	0.434	0.719
	R@500	0.513	0.467	0.765
	R@1000	0.554	0.510	0.804
ED	**R@1**	0.399	0.356	0.573
	R@5	0.486	0.451	0.739
	R@10	0.502	0.468	0.763

inaccurate mention-entity similarity measure. Our method employs plausible English mentions as pivots, and leverage lexical matching information to improve the accuracy.

- our method achieves better performance than Spotlight and TagMe. This indicates that our method is more flexible and feasible for mentions extracts from a user's actual questions.

For a more comprehensive comparison, we vary the size of the candidate entities in range of $\{50, 200, 500, 1000\}$, and report the average recall of TRANS-Search, SemSearch, and our method in Table 3. Moreover, we take the top-1000 candidate entities as input, and perform downstream entity disambiguation using the state-of-the-art method, i.e., multilingual-BERT [6]. For each mention-entity pair, we concatenate the *question* where the mention extracted from and the short *abstract* of the entity as a string, and perform entity disambiguation as the text classification task. The training data is collected from LC-QuAD [7], which is an English KBQA task. Similar to QALD, we extract questions and their corresponding mentions and entities to train the classifier. Table 3 reports the average recall at the top-1, top-5, and top-10 of different methods in entity disambiguation. We observe that,

- in candidate retrieval (CR), our method is consistently superior to other methods with different sized candidate entities, indicating the robustness of our method, and
- in entity disambiguation, pre-trained language model (i.e., multilingual-BERT) is powerful to learn the relevance between the source-language text and the target-language entity. Compared with the other method, our method achieves better performance, mainly due to the high recall in upstream candidate retrieval.

5.3.2 Comparison on WIKI-LRL. Then, we compare our method with the other baselines on the WIKI-LRL dataset in Table 4. Following [3], we report top-30 gold candidate recall. In the WIKI-LRL dataset, the source-language mentions are Wikipedia titles and the TRANS methods that rely directly on the Wikipedia titles as lexicons are excluded from the comparison. We observe that our method achieves the best performance across all three languages. PBEL is the state-of-the-art candidate retrieval method for low-resource language, and it is effective to leverage related high-resource languages as pivots to reduce the disconnect between mentions and entities. Our method leverages plausible English mentions as an intermediate without additional high-resource language information and achieves better results. Compared with SemSearch, our method performs better mainly because it combines the semantic similarity and lexical similarity between the mention and entity using plausible English mentions as the intermediary, instead of directly computing their similarity in the aligned latent space.

Table 4: Top-30 recall (R@30) of different methods on the WIKI-LRL dataset. PBEL_Char and PBEL_BiLSTM denote the PBEL method which encodes entities into vectors using BiLSTM and character-based CNN, respectively.

Languages (#mentions)	Marathi (2449)	Lao (799)	Telugu (1742)	Average
SemSearch	0.596	0.195	0.418	0.403
PBEL_BiLSTM	0.535	0.210	0.407	0.407
PBEL_CharCNN	0.477	0.180	0.246	0.348
OurMethod	0.702	0.307	0.532	0.514

5.4 In-depth Analysis

The intermediary collection $\mathcal{P}$ plays an important role in our method. To analyze the performance of different modules and investigate their impact on the final results, we evaluate the effect of character information, word-level NMS, and the size of the intermediary collection. Then, we analyse the bilingual-resource reliance and time-efficiency of our method.

5.4.1 Effect of Character Information. When filling the cross-lingual gap, if a source-language word x_i is out of vocabulary of the embedding space $\mathcal{X}$, we cannot find its semantically related English words. Inspired by the previous method [25, 34], we leverage character-level semantic matching to alleviate this problem.

To be specific, we randomly initialize all the characters in the source and target languages as fixed-length embeddings. Then, we design two character-level BiLSTM to encode words in the source and target languages in the latent vector space. Consider a source-language word x_i and its parallel target-language word y_i. Each word is a sequence of characters. The character embeddings are used as input to the BiLSTM and the final character embedding of each word is the concatenation of the last states from the forward and backward LSTMs. We train the model with a max-margin loss to maximize the cosine similarity between words which have same meaning in different languages, and minimize the similarity between negatively sampled word pairs:

$$\mathcal{L} = \max(0, sim(x, y^-) - sim(x, y) + \lambda), \tag{7}$$

where x and y is a word-pair in the seed dictionary which have the same meaning, y^- is a negative word in target language, and λ is the margin.

In our experiment, for the out-of-vocabulary source-language words, we search its most similar target-language words according to their character cosine similarities. We evaluate the performance of character information in the QALD dataset in Table 5. We observe that 4% mentions are out-of-vocabulary in word-level embedding space. Character-level information helps to improve our method, with an average performance gain of 1.2%.

Table 5: R@1000 on the QALD dataset to investigate the influence of character information. OOV denotes the percentage of our-of-vocabulary mentions. Δ denotes the performance improvement.

Languages	OOV(%)	w/ Char	w/o Char	Δ
German	4.17%	0.821	0.824	0.003
French	4.03%	0.796	0.801	0.004
Russian	4.17%	0.718	0.722	0.003
Spanish	4.17%	0.805	0.815	0.010
Italian	4.03%	0.786	0.799	0.013
Dutch	3.74%	0.821	0.828	0.006
Romanian	3.88%	0.811	0.828	0.016
Portuguese	3.88%	0.780	0.812	0.032
Average	4.01%	0.792	0.804	0.012

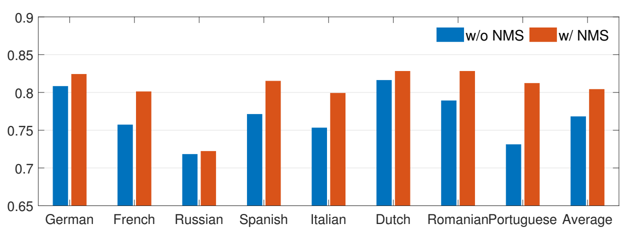

Figure 3: R@1000 on the QALD dataset to investigate the effectiveness of the NMS component.

5.4.2 Effect of Word-level NMS. We assess whether the word-level NMS is effective for generating diverse English mentions in Figure 3. We observe that our method achieves a significant performance gain using the word-level NMS method, with an averaged performance gain of 3.6%. This improvement is mainly comes from duplication reduction of the NMS component, which enhances the diversity of the intermediary collection and better covers the salient information in the mention. For example, the top-5 retrieved English words for the Spanish word *milpiés* (*millipede*) are {*springtails*, *centipedes*, *mantis*, *mantises*, *centipede*}. With the word-level NMS mechanism, the English word {*millipedes*} is included, which is salient in searching the gold entity, i.e., `Millipede` in downstream lexical retrieval.

5.4.3 Size of the Intermediary Collection. For each source-language mention m, we generate an intermediary collection with T_m plausible English mentions. To investigate the influence of T_m on candidate retrieval, we vary T_m between 1 and 10. The detailed results of R@1000 for different languages are plotted in Figure 4. The green bars represent the averaged recall of different languages. We observe that it performs worst when P_m only contains one plausible English mention (i.e., $T_m = 1$). This is mainly because that a word or phrase usually has multiple expressions, and one plausible English mention may be inaccurate and incomplete to capture the original source-language mention. Our method achieves best performance when T_m is set to 7. It is notable that adding plausible English mentions will result in a linear increase in time complexity of the lexical retrieval process. In practice, the recommend T_m is in range of [3,7].

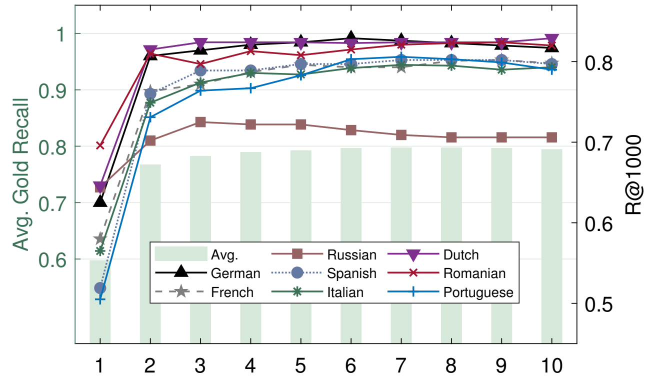

Figure 4: Influence of the size of intermediary collection on the QALD dataset. The x-axis shows the size of the intermediary collection, the left y-axis corresponds to the average R@1000 across eight languages, and the right y-axis denotes R@1000 of each language.

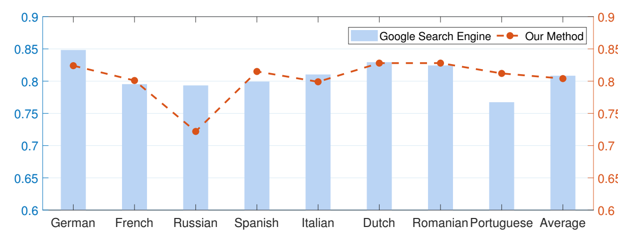

Figure 5: Comparison between our method and Google translator on the QALD dataset. The y-axis denotes R@1000 score of candidate retrieval.

5.4.4 Bilingual Resource Reliance. Our method only needs a bilingual word dictionary to align the source and target embedding space, which is a low-resource reliance method. We compare our method with the Google translator, which translates source-language mentions to target-language mentions and then generates candidate entities with lexical retrieval using BM25. It is important to note that the Google translator is trained on massively bilingual resources and is not available in many practical and industry scenarios. Figure 5 compares the performance of our method with the Google translator on the QALD dataset. The blue bar denotes the performance of Google translator in different language. The red line denotes the performance of our method. We observe that our method achieved better performance for Portuguese, Spanish, and French, but a bit worse for Russian. Considering the average R@1000 of eight languages, the Google translator (i.e., 0.808) only achieves a slight improvement of 0.4% over our method (i.e., 0.804). This demonstrates the effectiveness of intermediate collection, and the effectiveness of semantic retrieval and selection mechanisms in filling the cross-language gap.

Source Mention	Gold Entity	Plausible English Mentions
Norte Mar (Portuguese)	North_Sea	norte sea, south sea, **north sea**, south sea, southern mar
francés quinto República (Spansih)	French_Fifth_Republic	**french fifth republic**, france five republic, france fourth republic, french fifth republic, french fifth republican
burro di noccioline (Italian)	Peanut_butter	**butter di peanuts**, lard di peanuts, burro di peanuts, butter di custard, burro di syru
Финляндия (Russian)	Finland	**finland**, finnish, sweden, estonia, norway

Figure 6: Examples in the QALD dataset. The red plausible mentions are salient mentions to recall gold entity, marked by human evaluation.

5.5 Case Study

In this section, we present several examples from the QALD dataset in Figure 6 to give an intuitive impression of our method. We present the source mentions, their corresponding gold entities, and plausible English mentions generated by our method. We observe the plausible English words are effective to fill the cross-lingual gap between source and target language. For example, semantic retrieval is accurate to connect *Finland* in Russian and English scripts. The plausible English mentions that are important to recall the gold entity in downstream lexical retrieval are marked in red. For example, *butter di peanuts* is an effective query to search the entity `Peanut_butter`.

6 CONCLUSION

In this paper, we proposed a pivot-based candidate retrieval method for cross-lingual entity linking. It takes an intermediary set of plausible target-language mentions as pivots to bridge two types of gaps: cross-lingual gap and mention-entity gap. The learned plausible target-language mentions are capable of capturing the semantics of source-language mentions, and are effective to recall gold entity in the lexical retrieval. In the experiments, we evaluate our method on two challenging XEL datasets and the results demonstrate the competitiveness of our method. In the future, we plan to improve the quality of the intermediary set by automatically detecting the key-phrase of the source-language mention and alleviating the out-of-vocabulary problem.

ACKNOWLEDGMENTS

We would like to thank Tao Shen from University of Technology Sydney for his insightful comments and suggestions.

REFERENCES

[1] Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics, ACL*. 789–798.

[2] Yixin Cao, Lei Hou, Juanzi Li, and Zhiyuan Liu. 2018. Neural Collective Entity Linking. In *Proceedings of the International Conference on Computational Linguistics, COLING*. 675–686.

[3] Shuang Chen, Jinpeng Wang, Feng Jiang, and Chin-Yew Lin. 2020. Improving Entity Linking by Modeling Latent Entity Type Information. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*. 7529–7537.

[4] Xilun Chen and Claire Cardie. 2018. Unsupervised Multilingual Word Embeddings. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP*. 261–270.

[5] Joachim Daiber, Max Jakob, Chris Hokamp, and Pablo N. Mendes. 2013. Improving efficiency and accuracy in multilingual entity extraction. In *Proceedings of International Conference on Semantic Systems, ISEM*. 121–124.

[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT*. 4171–4186.

[7] Mohnish Dubey, Debayan Banerjee, Abdelrahman Abdelkawi, and Jens Lehmann. 2019. LC-QuAD 2.0: A Large Dataset for Complex Question Answering over Wikidata and DBpedia. In *Proceedings of International Semantic Web Conference, ISWC*. 69–78.

[8] Chris Dyer, Victor Chahuneau, and Noah A. Smith. 2013. A Simple, Fast, and Effective Reparameterization of IBM Model 2. In *Proceedings of The Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT*. 644–648.

[9] Paolo Ferragina and Ugo Scaiella. 2010. TAGME: on-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM*. 1625–1628.

[10] Octavian-Eugen Ganea and Thomas Hofmann. 2017. Deep Joint Entity Disambiguation with Local Neural Attention. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP*. 2619–2629.

[11] Luyu Gao, Zhuyun Dai, Zhen Fan, and Jamie Callan. 2020. Complementing Lexical Retrieval with Semantic Residual Embedding. *CoRR* abs/2004.13969 (2020). arXiv:2004.13969 https://arxiv.org/abs/2004.13969

[12] Ben Hachey, Will Radford, Joel Nothman, Matthew Honnibal, and James R. Curran. 2013. Evaluating Entity Linking with Wikipedia. *Artificial Intelligence* 194 (2013), 130–150.

[13] Minghao Hu, Yuxing Peng, Zhen Huang, and Dongsheng Li. 2019. Retrieve, Read, Rerank: Towards End-to-End Multi-Document Reading Comprehension. In *Proceedings of Association for Computational Linguistics, ACL*. 2285–2295.

[14] Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2017. Billion-scale similarity search with GPUs. *CoRR* abs/1702.08734 (2017).

[15] Guillaume Lample, Alexis Conneau, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2018. Word translation without parallel data. In *International Conference on Learning Representations, ICLR*.

[16] Phong Le and Ivan Titov. 2018. Improving Entity Linking by Modeling Latent Relations between Mentions. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics, ACL*. 1595–1604.

[17] Qian Liu, Heyan Huang, Yang Gao, Xiaochi Wei, Yuxin Tian, and Luyang Liu. 2018. Task-oriented Word Embedding for Text Classification. In *Proceedings of the International Conference on Computational Linguistics, COLING*. 2023–2032.

[18] Qian Liu, Heyan Huang, Junyu Xuan, Guangquan Zhang, and Jie Lu. 2020. A Fuzzy Word Similarity Measure for Selecting Top-k Similar Words in Query Expansion. *IEEE Transactions on Fuzzy Systems* (2020). https://doi.org/10.1109/TFUZZ.2020.2993702

[19] Qian Liu, Heyan Huang, Guangquan Zhang, Yang Gao, Junyu Xuan, and Jie Lu. 2018. Semantic Structure-Based Word Embedding by Incorporating Concept Convergence and Word Divergence. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*. 5261–5268.

[20] Yuval Merhav and Stephen Ash. 2018. Design Challenges in Named Entity Transliteration. (2018), 630–640. arXiv:1808.02563 http://arxiv.org/abs/1808.02563

[21] Tomás Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings of Conference on Neural Information Processing Systems (NeurIPS)*. 3111–3119.

[22] Xiaoman Pan, Thamme Gowda, Heng Ji, Jonathan May, and Scott Miller. 2019. Cross-lingual Joint Entity and Word Embedding to Improve Entity Linking and Parallel Sentence Mining. In *Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP, DeepLo@EMNLP-IJCNLP 2019*. 56–66.

[23] Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017. Cross-lingual Name Tagging and Linking for 282 Languages. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics, ACL*. 1946–1958.

[24] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 6 (2017), 1137–1149.

[25] Shruti Rijhwani, Jiateng Xie, Graham Neubig, and Jaime G. Carbonell. 2019. Zero-Shot Neural Transfer for Cross-Lingual Entity Linking. In *Proceedings of Conference on Artificial Intelligence, AAAI*. 6924–6931.

[26] Stephen E. Robertson and Hugo Zaragoza. 2009. The Probabilistic Relevance Framework: BM25 and Beyond. *Foundations and Trends in Information Retrieval* 3, 4 (2009), 333–389.

[27] Azriel Rosenfeld and Mark Thurston. 1971. Edge and Curve Detection for Visual Scene Analysis. *IEEE Trans. Comput.* 20, 5 (1971), 562–569.

[28] Avirup Sil and Radu Florian. 2016. One for All: Towards Language Independent Named Entity Linking. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL.*

[29] Avirup Sil, Gourab Kundu, Radu Florian, and Wael Hamza. 2018. Neural Cross-Lingual Entity Linking. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*. 5464–5472.

[30] Chen-Tse Tsai and Dan Roth. 2016. Cross-lingual Wikification Using Multilingual Embeddings. In *Proceedings of The Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT*. 589–598.

[31] Chen-Tse Tsai and Dan Roth. 2018. Learning Better Name Translation for Cross-Lingual Wikification. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*. 5528–5536.

[32] Shyam Upadhyay, Nitish Gupta, and Dan Roth. 2018. Joint Multilingual Supervision for Cross-lingual Entity Linking. In *Proceedings of the Empirical Methods in Natural Language Processing, EMNLP*. 2486–2495.

[33] Shyam Upadhyay, Jordan Kodner, and Dan Roth. 2018. Bootstrapping Transliteration with Constrained Discovery for Low-Resource Languages. (2018), 501–511.

[34] John Wieting, Mohit Bansal, Kevin Gimpel, and Karen Livescu. 2016. Charagram: Embedding Words and Sentences via Character n-grams. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP*. 1504–1515.

[35] Chao Xing, Dong Wang, Chao Liu, and Yiye Lin. 2015. Normalized Word Embedding and Orthogonal Transform for Bilingual Word Translation. In *Proceedings of North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT*. 1006–1011.

[36] Mengge Xue, Weiming Cai, Jinsong Su, Linfeng Song, Yubin Ge, Yubao Liu, and Bin Wang. 2019. Neural Collective Entity Linking Based on Recurrent Random Walk Network Learning. In *Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI*. 5327–5333.

[37] Wen-tau Yih, Ming-Wei Chang, Xiaodong He, and Jianfeng Gao. 2015. Semantic Parsing via Staged Query Graph Generation: Question Answering with Knowledge Base. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics, ACL*. 1321–1331.

[38] Boliang Zhang, Ying Lin, Xiaoman Pan, Di Lu, Jonathan May, Kevin Knight, and Heng Ji. 2018. ELISA-EDL: A Cross-lingual Entity Extraction, Linking and Localization System. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics, NAACL-HLT*. 41–45.

[39] Shuyan Zhou, Shruti Rijhwani, John Wieting, Jaime G. Carbonell, and Graham Neubig. 2020. Improving Candidate Generation for Low-resource Cross-lingual Entity Linking. *Transactions of the Association for Computational Linguistics* 8 (2020), 109–124.