

Mining Entity Translations from Comparable Corpora: A Holistic Graph Mapping Approach

Jinhan Kim[†] Long Jiang[‡] Seung-won Hwang[†] Young-In Song[‡] Ming Zhou[‡]

[†]Pohang University of Science and Technology, Pohang, Republic of Korea
{wlsqks08, swhwang}@postech.ac.kr

[‡]Microsoft Research Asia, Beijing, China
{longj, yosong, mingzhou}@microsoft.com

ABSTRACT

This paper addresses the problem of mining named entity translations from comparable corpora, specifically, mining English and Chinese named entity translation. We first observe that existing approaches use one or more of the following named entity similarity metrics: entity, entity context, and relationship. Inspired by this observation, in this paper, we propose a new holistic approach, by (1) combining all similarity types used and (2) additionally considering relationship context similarity between pairs of named entities, a missing quadrant in the taxonomy of similarity metrics. We abstract the named entity translation problem as the matching of two named entity graphs extracted from the comparable corpora. Specifically, named entity graphs are first constructed from comparable corpora to extract relationship between named entities. Entity similarity and entity context similarity are then calculated from every pair of bilingual named entities. A reinforcing method is utilized to reflect relationship similarity and relationship context similarity between named entities. According to our experimental results, our holistic graph-based approach significantly outperforms previous approaches.

Categories and Subject Descriptors

I.2.7 [Natural Language Processing]: Machine Translation, Text Analysis; H.2.8 [Database Application]: Data Mining

General Terms

Languages, Experimentation

Keywords

Named Entity Translation, Graph Mapping

1. INTRODUCTION

Named entities (NE) normally refer to a range of concepts like people names, place names, organization names, product names,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'11, October 24–28, 2011, Glasgow, Scotland, UK.
Copyright 2011 ACM 978-1-4503-0717-8/11/10 ...\$10.00.

and so on. Large quantities of new named entities appear every day in newspapers, web sites, TV programs, and so on. However, their translations usually cannot be found in general purpose bilingual dictionaries. Automatic NE translation extraction is thus important to machine translation systems and cross-language information retrieval applications.

For this NE translation extraction task, early researches focused on finding NE correspondence in parallel sentences (or parallel corpora) [12, 7, 3]. Though such NE translation extraction from parallel corpora can achieve very high accuracy, obtaining large-scale parallel corpora is not a trivial task, especially for uncommon language pairs.

To avoid the high costs of obtaining parallel corpora, many researchers have switched to comparable corpora for building large scale bilingual lexicons. Comparable corpora refer to those texts that are not translations of each other but talk about the same or related topics. For instance, news articles published in different languages within the same period can be regarded as comparable corpora. The amount of comparable corpora is much larger than that of parallel corpora, and they are easier to obtain, which makes them a desirable source for large-scale bilingual lexicon building.

Table 1: Taxonomy of cross-language similarity metrics.

	Entity	Relationship
Using entity names	E	R
Using textual context	EC	

Existing work can be categorized, based on the types of similarity metrics used (Table 1):

- **E**: Entity names in the two corpora can be compared based on their phonetic similarity, *e.g.*, high phonetic similarity between Obama and Chinese translation of Obama, 奥巴马 (pronounced 'Aobama'), is evidence that they are translation pairs.
- **EC**: Entity Context, *i.e.*, common text surrounding named entity translation pair in the two corpora, can be compared. For example, Obama and 奥巴马, surrounded by words with the same meanings, such as 'president' and '总统', support the inference that they are translation pairs.
- **R**: Relationships between a NE pair in one corpus and a NE pair in other corpus, quantified by named entity co-occurrence,

can be used as relational evidence that they are translation pair, *e.g.*, Barack and Michelle (巴拉克 和 米歇尔) frequently co-occur.

More specifically, early work applied E, using either rule-based phonetic similarity [19] or statistics-based phonetic similarity [11, 6]. However, pure E-based methods did not achieve satisfactory results, as some translations do not have similar pronunciation and there are also NEs with ambiguous pronunciation. In contrast, [15, 4] proposed to extract word translations from comparable corpora, to use EC. More EC-based approaches were proposed [2, 22, 8] adopting different context representation.

More recently, *holistic* approaches [16, 21] combining multiple similarity metrics have been proposed and outperformed preceding work. First, [16] combines E and EC using a language model and a probabilistic transliteration model respectively to find Chinese-English translation pair. In contrast, [21] combines E and R. As most NE translations do not appear in a general purpose bilingual word dictionary, [21] adopts an iterative approach to gradually refine R.

In this paper, we complete the picture, by (1) combining E, EC, and R, and (2) additionally considering RC, which is a missing quadrant in the taxonomy. Like EC, RC, Relationship Context, considers the textual context similarity between an English NE pair and a Chinese NE pair, *i.e.*, surrounding text explaining the relationship of the NE pair, as evidence for whether the two relationships are comparable. For example, Barack and Michelle described as couple in both corpora is a strong evidence.

Specifically, we first construct NE relationship graphs from comparable corpora and initialize pairwise translation similarity between bilingual NE pairs, using E and EC. We then iteratively reinforce the overall translation similarity until convergence using R and RC. Our experimental results suggest that our new holistic approach significantly outperforms previous state-of-art approaches.

To summarize, this paper has the following contributions:

1. We present a new named entity translation similarity metric, relationship context similarity (RC), to complete the taxonomy of cross-language named entity translation similarity metrics.
2. We propose a new holistic English-Chinese NE translation approach, by combining all translation similarity metrics (E, EC, R, and RC).
3. We perform various kinds of evaluations to present the effectiveness of our holistic approach by comparing with existing NE translation approaches.

The rest of this paper is organized as follows. Section 2 reviews existing work. Section 3 formally defines the problem and Section 4 then develops our framework. Section 5 reports experimental results and Section 6 concludes the paper.

2. RELATED WORK

As described earlier, existing work has focused on using entity context similarity (EC), entity similarity (E), and relationship similarity (R) for mining cross-language term translations from comparable corpora.

Early work applies entity similarity (E), using either rule-based phonetic similarity [19] or statistics-based phonetic similarity [11, 6, 17]. However, pure E-based methods did not achieve satisfactory results, as some translations do not have similar pronunciation and there are also NEs with ambiguous pronunciation. Meanwhile,

[15, 4, 2, 22, 8] extracted translation pairs with high entity context similarity (EC) using word co-occurrences. More recently, holistic approaches combining two similarity notions have been studied, using E and EC [16, 1], and E and R [21] for computing overall similarity respectively. In this paper, we implemented [16] and [21] as baseline approaches, which we denote as Shao and You respectively from this point on.

More specifically, Shao [16] combines E and EC. For computing E, Shao uses a probabilistic pronunciation model for machine transliteration using expectation maximization (EM) algorithm. For computing EC, Shao uses a language model used in a typical information retrieval (IR) problem which considers each Chinese word as a query and each English word as a document using their context words. Each of these two models ranks English translation candidate words for each Chinese word. After that, Shao examines the top M English translation candidate words in both ranked lists, finds English words that appear in top M positions in both ranked lists, and selects the English word with the highest average rank position as the translation of the Chinese word.

On the other hand, You [21] combines E and R. For computing E, You computes edit distance (ED) between Chinese Pinyins and English strings as they are pronounced similarly. For R, You uses monolingual entity co-occurrences, crawled from entity search web sites. You develops an entity matching algorithm which initializes the translation similarity between named entities by using E and iteratively reinforces the translation similarity by using R. After the convergence of the reinforcement step, You extracts one-to-one matches maximizing the overall similarity by using greedy matching algorithm.

Our work generalizes these efforts by combining three notions (E, EC, and R) and by also considering relationship context similarity (RC), with the following distinctions from each.

Compared to Shao, our approach has at least two advantages. First, NEs co-occur frequently, and co-occurrence information is an important factor to determine NE translation. However, NE translation pairs rarely appear in a dictionary and thus are ignored in Shao. By iterative reinforcement in our approach, the correctly translated NEs can help us find the translations of their neighbors in the next iteration, which enables us to make use of most NEs and provide more accurate NE translations. Second, in Shao, the candidate ranking of a given NE is independent of the rankings of other NEs. It is very likely that one target language NE is considered the translation of many source language NEs (*i.e.*, one-to-many mapping). However, in our scenario, we want to find the best one-to-one mapping. That is, if Chinese NE is mapped to an English NE, it will not be mapped to other English NEs.

Compared to You, our approach has the following advantages: First, You gets named entities and their relationship similarity (R) directly from entity search web sites and thus cannot be easily extended to other languages; Second, since the entity search web sites only provide the co-occurrence relationship between NEs, it does not contain common words but only the related NEs, and thus may not be sufficient for finding the right translation; Third, when mapping NEs in the source and target graphs, You only considers relationship similarity (R) and ignore the relationship context similarity (RC), which may result in incorrect mapping.

Meanwhile, some existing work focuses on Web corpus [10, 14, 9], of a larger scale with higher asymmetry than comparable corpora, as an alternative resource to extract NE translations with promising results. However, these approaches require either crawling the corpus on a scale that cannot be easily hosted, or frequently issuing external queries to web search engines, both of which are costly.

3. PROBLEM STATEMENT

This section formally defines the problem and introduces the notations.

We abstract the NE translation mining problem as the matching of two NE relationship graphs extracted from the comparable corpora. More formally, we construct two graphs $G_e = (V_e, E_e)$ and $G_c = (V_c, E_c)$ from each English and Chinese corpus where V is a set of NEs (nodes) and E is a set of relationships (edges). Our goal is to populate a hard decision (or similarity) matrix R^* , a $|V_e|$ -by- $|V_c|$ matrix, where each R_{ij}^* denotes whether the two NEs $e_i \in V_e$ and $c_j \in V_c$ are translations of each other, *i.e.*, 1 if they are and 0 otherwise. As an intermediate goal, we populate a soft decision matrix R where each R_{ij} indicates the likelihood to be a translation.

Towards this goal of populating R , we holistically combine all the similarity notions used in the past— E, EC (Shao [16]), and R (You [21])— together with the new notion of RC.

R_0 using E and EC: A soft decision matrix R can be initialized by using pairwise entity similarity (E), as done in You. However, it cannot represent translation pairs with little phonetic similarity, *e.g.*, Jackie Chan and 成龙 (pronounced Cheng Long). To compensate for this weakness, we also consider entity context similarity (EC) around the NEs, as done in Shao. For example, if words like ‘actor’ co-occur frequently with Jackie Chan in the English corpus, the Chinese word for ‘actor’ is highly likely to co-occur with 成龙 in the Chinese corpus.

In order to consider E and EC, we first construct two $|V_e|$ -by- $|V_c|$ translation similarity matrices, S^E and S^{EC} , for E and EC respectively. We then combine S^E and S^{EC} to build the initial soft decision matrix R_0 .

R_∞ reflecting R and RC: You uses relationship similarity (R) between monolingual NE pairs in each corpus, which considers co-occurrences between NEs, to iteratively reinforce the soft decision matrix R . For instance, the Chinese translations of some NEs co-occurring frequently with ‘Jackie Chan’ in the English corpus, *e.g.*, ‘Chris Tucker’, are also likely to co-occur frequently with 成龙 in the Chinese corpus. As a result, if many English named entities co-occur with ‘Jackie Chan’ and their Chinese translations with 成龙, we can give a high similarity score for the NE translation pair (‘Jackie Chan’, 成龙).

We take one step further to take advantage of the relationship context similarity (RC) between two monolingual NE pairs in the two corpora. In other words, the relationship between ‘Jackie Chan’ and ‘Chris Tucker’ can be represented by context words, co-occurring with these named entity pairs, such as ‘Rush Hour’. Similarly, in G_c , context words for the translations of ‘Jackie Chan’ and ‘Chris Tucker’ can be collected, *e.g.*, the Chinese translation for ‘Rush Hour’. The similarity of context words between two monolingual NE pairs in the two NE graphs, or RC, would give additional evidence that the right translation pairs have been found.

However, in order to consider R and RC, we must already know the translations of NEs, which makes this problem a chicken-and-egg dilemma. The nature of this problem therefore suggests an iterative process. That is, as we identify highly likely NE translation pairs from R_0 , we can update these translated pairs to the named entity dictionary, such that, the remaining NEs can be updated to R_1 from this dictionary. From this update, we may identify more translation pairs, which we continue until pairwise similarity converges to R_∞ .

This holistic approach complements the current state-of-the-art approaches, as summarized in Table 2.

Table 2: The similarity metrics used by current state-of-art approaches and our new holistic approach.

	E	EC	R	RC
Shao [16]	✓	✓		
You [21]	✓		✓	
Our holistic approach	✓	✓	✓	✓

4. OUR FRAMEWORK

As described in the previous section, our goal is to populate a hard decision matrix R^* , using both entity-based (E and EC) and relationship-based (R and RC) similarities. Our framework consists of the following steps:

- Building NE graphs and initializing R_0 :** building two NE relationship graphs, G_e and G_c , from comparable corpora and initializing a soft decision matrix R_0 by combining two entity-based similarity metrics, E and EC.
- Reinforcing into R_∞ :** reinforcing the soft decision matrix (or translation similarity matrix) R by exploiting R and RC as additional evidence.
- Populating R^* from R :** generating a binary hard decision matrix R^* by using greedy matching algorithm.

We discuss each of these steps in the following sections respectively.

4.1 Initialization Step

This section discusses how we build two NE relationship graphs, G_e and G_c , from each English and Chinese corpus, and how we generate the initial soft decision matrix R_0 by combining two entity-based similarity metrics, E and EC.

4.1.1 Constructing NE Relationship Graphs

To extract the NE relationship information from the English and Chinese corpora, we first build two NE relationship graphs, $G_e = (V_e, E_e)$ and $G_c = (V_c, E_c)$, from each English and Chinese corpus where V is a set of NEs (nodes) and E is a set of relationships (edges) between NEs.

For mining V_e from the English corpus, we use a CRF-based English NE tagger. We then build an English relationship graph G_e , regarding an NE that appears more than δ times in the English corpus as a node and connecting two nodes (NEs) if the corresponding NE pair co-occurs more than δ times in the English documents in the corpus. We also build a Chinese graph G_c using the same approach after mining V_c from the Chinese corpus. We segment the Chinese corpus and identify NEs using a Chinese word breaker based on [5]. In this paper, we set $\delta = 20$ based on empirical observations.

4.1.2 Initializing R_0

We build the initial soft decision matrix R_0 by using two entity-based similarity metrics, E and EC. We first compute the entity similarity matrix S^E and the entity context similarity matrix S^{EC} . We then combine S^E and S^{EC} to generate R_0 .

For computing the entity similarity matrix S^E between English NE e in V_e and Chinese NE c in V_c , we adopt the Edit-Distance (ED) between e and c used in You [21].

More specifically, we first convert c into its Pinyin form PY_{c_j} . Pinyin is the romanization representation of pronunciation of Chinese character. We then compute the string similarity between English string e and Chinese Pinyin PY_{c_j} , as English strings and Chinese Pinyins are pronounced similarly. To compute the string similarity, we use the Edit-Distance defined in You. Using the Edit-Distance score, entity similarity between English NE e_i and Chinese NE c_j is computed as:

$$S_{ij}^E = 1 - \frac{ED(e_i, PY_{c_j})}{Len(e_i) + Len(PY_{c_j})}$$

where $e_i \in V_e$, $c_j \in V_c$, and $Len(e_i)$ and $Len(PY_{c_j})$ are the length of e_i and PY_{c_j} respectively.

For computing the entity context similarity matrix S^{EC} , we first define the entity context as the surrounding words w of the NE. Formally, let $d = \{w_1, w_2, \dots, w_n\}$ be a given document which contains an English NE e in a corpus and l be the length of the entity context (or window size). When e is the i^{th} word w_i in the document d , entity context word list is:

$$\{w_{i-\frac{l}{2}}, w_{i-\frac{l}{2}+1}, \dots, w_i(=e), \dots, w_{i+\frac{l}{2}-1}, w_{i+\frac{l}{2}}\}$$

In this paper, we set the length of the window size $l = 24$, based on empirical observations.

For the given English NE e , we aggregate all the context windows of e by summing up the occurrence of each context word w in the English corpus, and build N_e -by- N_w co-occurrence matrix where N_e is the number of English NEs and N_w is the number of context words. We then define a context association vector CA_e of NE e , using the Log-Odds (LO) association measure, which was the winner among all the context association metrics reported in [13], including chi-square and log-likelihood. The LO association measure between NE e and a single context word w is defined as:

$$LO(e, w) = \log \frac{(O_{11} + \frac{1}{2})(O_{22} + \frac{1}{2})}{(O_{12} + \frac{1}{2})(O_{21} + \frac{1}{2})}$$

where O_{ij} are the cells of the 2×2 contingency matrix of e and w . That is, O_{12} and O_{21} stand for the sum of co-occurrence matrix elements containing either e or w exclusively (i.e., e but not w , or w but not e respectively), O_{11} for both and O_{22} for neither. CA_e can be similarly defined for Chinese NEs.

Now, we define the vector similarity between CA_e and CA_c , which shows the entity context similarity (EC) of English NE e and Chinese NE c , as CA_e and CA_c contain the relation score between NE and its context words. For such a computation, we need to ‘align’ two vectors, such that scores in the same vector position correspond to the words of the same meaning, such as ‘president’ and ‘总统’. For this purpose, we use an English-Chinese word dictionary to align words with the same meaning.

Specifically, among entity context words of e and c , we stem the words and keep $w_e \in CA_e$ and $w_c \in CA_c$, only if the context word pair (w_e, w_c) appears as a translation in the dictionary. In other words, we prune out entity context words (1) if the English word does not appear in the dictionary or (2) its dictionary translation does not appear in the Chinese context. We also prune out stopwords and numbers as they do not add much contextual information.

However, as the context word translation pair may not map one-to-one to each other, there can be a word with multiple translations. In this case, we divide the association score among all possible translations: if w_e has two translations $w_c^{(1)}$ and $w_c^{(2)}$ with their occurrences in the Chinese corpus as n and m respectively, we divide w_e into $w_e^{(1)}$ and $w_e^{(2)}$ indicating the two different senses of

the word, corresponding to $w_c^{(1)}$ and $w_c^{(2)}$ respectively. We then add new context association vector entries $CA_e(w_e^{(1)})$ and $CA_e(w_e^{(2)})$ as $CA_e(w_e) \times \frac{n}{n+m}$ and $CA_e(w_e) \times \frac{m}{n+m}$ respectively, then delete $CA_e(w_e)$. For example, if $CA_e = \{(w_e^a, 0.5), (w_e^b, 0.7)\}$ and w_e^a has two Chinese translations (w_c^{a1} and w_c^{a2}) with their occurrences in the Chinese corpus as 4 and 6 respectively, we remove $(w_e^a, 0.5)$ and add two new entries $(w_e^{a1}, 0.5 \times \frac{4}{4+6})$ and $(w_e^{a2}, 0.5 \times \frac{6}{4+6})$. As a result, new context association vector of NE e is $CA_e = \{(w_e^{a1}, 0.2), (w_e^{a2}, 0.3), (w_e^b, 0.7)\}$. We similarly refine context association vectors for Chinese words.

Once all context association vectors have been refined, we compute the cosine similarity between context association vector of English NE e_i and context association vector of Chinese NE c_j as:

$$S_{ij}^{EC} = \frac{CA_{e_i} \cdot CA_{c_j}}{\|CA_{e_i}\| \cdot \|CA_{c_j}\|}$$

where $e_i \in V_e$ and $c_j \in V_c$.

Finally, we discuss how we combine S^E and S^{EC} . We first use min-max normalization for the values in S^E and S^{EC} into the same range of [0:1]. After that, we merge two similarity matrices using the following equation:

$$R_{ij} = S_{ij}^E \times S_{ij}^{EC}$$

where R_{ij} is the initial similarity score (R_{ij}^0) between English NE e_i and Chinese NE c_j .

4.2 Reinforcement Step

From the initial similarity matrix R_0 , we iteratively reinforce R_{ij} of the nodes $i \in V_e$ and $j \in V_c$ from the matching similarities of their neighbor nodes u and v using relationship-based similarity metrics, **R** and **RC**.

The basic intuition is built on the assumption that two NEs with a strong relationship co-occur frequently in both corpora (**R**) with the similar context words (**RC**). In order to express this intuition, we formally define an iterative reinforcement model as follows. Let R_{ij}^t denote the similarity of nodes i and j at t -th iteration:

$$R_{ij}^{t+1} = \lambda \sum_{(u,v) \in B^t(i,j,\theta)} \frac{R_{uv}^t \times (S_{iu,jv}^{RC})^\gamma}{2^k} + (1 - \lambda)R_{ij}^0$$

where $B^t(i, j, \theta)$ is an ordered set which contains the best neighbor matching pairs between node i 's neighbor u and node j 's neighbor v , and $S_{iu,jv}^{RC}$ is the **RC** similarity matrix element between edge $(i, u) \in E_e$ and edge $(j, v) \in E_c$ for the pair (u, v) listed in $B^t(i, j, \theta)$. We will discuss how we populate $B^t(i, j, \theta)$ later. γ is the parameter to control the weight of the **RC** similarity.

Our reinforce model is expressed as a linear combination of (a) relationship similarity and relationship context similarity $\sum (R_{uv}^t \times (S_{iu,jv}^{RC})^\gamma) / 2^k$ and (b) initial similarity R_{ij}^0 . λ is the coefficient for interpolating the two terms.

Similar to entity context similarity S^{EC} , relationship context similarity S^{RC} is computed as the cosine similarity between two context association vectors, $CA_{(e,e')}$ and $CA_{(c,c')}$, defined for all English NE pairs and all Chinese NE pairs respectively. The differences are that we generate context association vector $CA_{(e,e')}$ based on the co-occurring context words from the documents where two NEs e and e' co-occur, and set the size of window l as the size of the given document. Similarly, we compute $CA_{(c,c')}$ and quantify the cosine similarity.

Algorithm 1 Find Common Neighbors ($R^t, S^{\text{RC}}, i, j, \theta, \gamma$)

```
1:  $B^t(i, j, \theta) \leftarrow \emptyset$ 
2: for all  $(i, u) \in E_e$  and  $(j, v) \in E_c$  do
3:    $SIM_{iu,jv} = R_{uv}^t \times (S_{iu,jv}^{\text{RC}})^\gamma$ 
4: end for
5:  $S \leftarrow \text{Sort}(SIM)$   $\{S$  is the ordered set of elements in  $SIM\}$ 
6: for  $SIM_{iu,jv} \in S$  do
7:   if  $SIM_{iu,jv} < \theta$  then
8:     return  $B^t(i, j, \theta)$ 
9:   end if
10:  if both  $u$  and  $v$  are not matched yet then
11:     $B^t(i, j, \theta) \leftarrow B^t(i, j, \theta) \cup \{(u, v)\}$ 
12:  end if
13: end for
14: return  $B^t(i, j, \theta)$ 
```

In our reinforcement model, as evidence that English NE i is a translation of Chinese NE j , we consider the relationship similarity R between i and j . In R , we check the extent to which i and j share common neighbors using $B^t(i, j, \theta)$, which is an ordered set of the best matching translation pairs between neighbors of i and j , as similarly used in You [21] (Algorithm 1). More specifically, we consider the neighbor pair (u, v) with $R_{uv}^t \times (S_{iu,jv}^{\text{RC}})^\gamma \geq \theta$, as a matching pair, where u is a neighbor of i and v is a neighbor of j (Line 6-13). But, instead of keeping all such pairs in $B^t(i, j, \theta)$, we make sure to eliminate many-to-many matches, by not adding (u', v') , if u' or v' appears in the pair with a higher matching score (Line 10-12). When $(u, v)_k$ is the pair in $B^t(i, j, \theta)$ with the k -th highest similarity score, we apply decaying coefficient $1/2^k$, to guarantee the convergence, as $\sum_{k=1}^{\infty} 1/2^k = 1$.

4.3 Matching Step

Once R converges to R_∞ , we use R_∞ as a soft decision matrix, representing the likelihood of each bilingual NE pair being a translation, to generate a binary hard decision matrix R^* , showing the one-to-one correspondence that is most likely. Such a problem can be stated as the maximum weighted bipartite matching problem and has been studied in [20]. However, computing the optimal matching solution incurs high computational cost. We thus approximate this problem using a greedy matching algorithm (Algorithm 2).

Algorithm 2 Greedy Matching Algorithm (R, θ)

```
1:  $S \leftarrow \text{Sort}(R)$   $\{S$  is the ordered set of element  $R_{ij}\}$ 
2: for  $R_{ij} \in S$  do
3:   if both  $i$  and  $j$  are not matched yet then
4:     if  $R_{ij} \geq \theta$  then
5:        $R^*_{ij} = 1$   $\{i$  and  $j$  are matched $\}$ 
6:     else
7:        $R^*_{ij} = 0$   $\{i$  and  $j$  are not matched $\}$ 
8:     end if
9:   else
10:     $R^*_{ij} = 0$   $\{i$  or  $j$  are already matched $\}$ 
11:   end if
12: end for
13: return  $R^*$ 
```

Algorithm 2 works as follows: (1) Elements in the soft decision matrix R are sorted by the decreasing order of the similarity score R_{ij} (Line 1); (2) We check each element in S and set $R^*_{ij} = 1$,

if neither $i \in V_e$ nor $j \in V_c$ was selected and the similarity score R_{ij} is larger than the given threshold θ (Line 2-12), otherwise we set $R^*_{ij} = 0$.

5. EXPERIMENTS

This section reports our experimental results to evaluate our proposed approaches. First, we report our experimental settings in Section 5.1. Second, we validate the effectiveness of our approaches over a real-life dataset in Section 5.2.

5.1 Experimental Settings

This section reports our experimental settings to evaluate our proposed approaches.

We used the same dataset (but with different time period) as Shao [16] did. For the English corpus, we used the English Gigaword Corpus (LDC2009T13) published by the Xinhua News Agency which contains news articles from January 2008 to December 2008. For the Chinese corpus, we used the Chinese Gigaword Corpus (LDC2009T27) also published by Xinhua News Agency during the same period. The English corpus contains 100,746 news documents and the Chinese corpus contains 88,029 news documents. As two corpora handle the news documents during the same period, they are regarded as comparable corpora in our experiments. From the English and Chinese corpora, we extracted person names by using NE taggers and built English and Chinese person name graphs independently as discussed in Section 4.

For the purpose of comparison, we implemented the following four existing approaches with ours:

- **EC**: Entity context-based approach which ranks NE candidates using word context similarity S_{ij}^{EC} .
- **E**: Entity name-based approach which ranks NE candidates using NE pronunciation similarity S_{ij}^{E} .
- **Shao**: Holistic approach which uses the average rank of NE candidates from two ranked lists computed by using EC and E [16].
- **You**: Holistic approach which initializes the translation similarity using E and iteratively reinforces the translation similarity using R from the two NE graphs [21].
- **E+EC+R**: Our approach without considering RC by setting $\gamma = 0$.
- **E+EC+R+RC**: Our approach considering RC setting non-zero γ .

We re-implemented You to use the same NE relationship graphs we extracted from the comparable corpora, and used the entity similarity (E) module of You for re-implementing Shao as well. In implementing EC, Shao, and our methods, we used an in-house English-Chinese general word dictionary containing more than 80,000 English-Chinese translation word pairs from Web resources such as online dictionaries to compute contextual similarities (EC and RC). As a result, this dictionary does not contain NE translation pairs and mainly consists of the pairs of general words.

For the main evaluation metrics, we used standard precision, recall, and F1 scores:

- **Precision**: the ratio of correctly returned translation pairs over returned translation pairs.
- **Recall**: the ratio of correctly returned translation pairs over the total number of correct translation pairs.

English Name	EC	E	E+EC+R	E+EC+R+RC
Fidel	菲德尔	费德勒	菲德尔	菲德尔
Trickett	孙晔	特里克特	特里克特	特里克特
Gbagbo	杜拉希	巴博	巴博	巴博
Amorim	拉夫罗夫	阿莫林	阿莫林	阿莫林
Federer	杰维季耶娃	菲德尔	费德勒	费德勒
Kikwete	纳希德	斯科特	基奎特	基奎特
Mottaki	雅伯	摩加迪	穆塔基	穆塔基
Calderon	西梅翁	麦德林	卡尔德龙	卡尔德隆

Figure 1: The translation examples of person NEs. The shaded cells indicate the correctly translated pairs.

- F1-score: the harmonic mean of precision and recall.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

We designed our experiments to follow the experiment setting used in most previous work (Shao and You) on name translation for fair comparison with them. Their settings aim to estimate a precision of name translations rather than recall, for which we need to ensure that the gold standard set has (a) sufficient amount of named entities in Chinese and English corpora and (b) such entities are not biased. To build ground truth on above conditions, we firstly selected 500 English person names randomly from all the person names extracted from the English corpus. Then, we hired an annotator to translate them into their Chinese translation names. These 500 English and Chinese named entity translation pairs were used for evaluation as ground truth. Among the 500 Chinese person name translations in the pairs, only 203 were matched to the person names extracted from our Chinese corpus. We used all 500 pairs to find the best parameters and computed the evaluation measures only using 203 matched pairs, as it is more suitable for real-life translation scenarios. This also follows Shao’s [16] practice to make the results easy to compare.

In the evaluation, we optimized parameters as following: We used a fixed value of $\lambda = 0.15$ which is one of the robust parameter settings in [21]. We performed a 5-fold cross-validation by dividing 500 English-Chinese NE translation pairs into 5 groups (each group contains 100 NE translation pairs). We used four groups to optimize necessary threshold parameters of each method (for instance, θ in EC, E, You, and our two approaches, γ in our approach, and M value in Shao), and used the remaining group as a test dataset using optimized parameters. The optimal parameters were determined by the ones maximizing F1-scores.

5.2 Experimental Results

This section reports our experimental results using the evaluation datasets. In particular, we evaluated (1) the effectiveness of the overall framework using precision, recall, and F1-score by comparing our approaches with four existing approaches including two state-of-the-art approaches [16, 21], (2) the effectiveness of our framework in terms of frequency of NEs, (3) the contribution of our reinforcement model, greedy matching algorithm, and parameters in improving NE translation quality, (4) other types of NEs using precision, recall, and F1-score, and (5) the effect of fuzzy matching. (In all experimental results, **Bold numbers** indicate the best performance for each metric).

5.2.1 Performance Comparison

We first evaluated the effectiveness of the overall framework using precision, recall, and F1-score by comparing our approaches with four existing approaches.

Table 3: The average precision, recall, and F1-score of each approach computed using 5-fold cross-validation. E+EC+R shows the best recall, and E+EC+R+RC shows the best precision and F1-score.

	Precision	Recall	F1 Score
EC	0.0819	0.2053	0.1160
E	0.3082	0.3255	0.3159
Shao	0.3000	0.3943	0.3379
You	0.3409	0.3965	0.3630
E+EC+R	0.4480	0.6288	0.5225
E+EC+R+RC	0.5105	0.5713	0.5388

Table 3 presents our experimental results. Our proposed approaches, E+EC+R and E+EC+R+RC, clearly outperform other approaches including both You and Shao. Specifically, E+EC+R shows the best recall, and E+EC+R+RC shows the best precision and F1-score. These performance improvements of ours over both You and Shao were statistically significant according to the Student’s t-test at $P < 0.01$ level.

More specifically, E+EC+R performs significantly better than You and Shao. This result indicates that considering entity context similarity (EC) complements [21] by supplying more evidence for name translation, while considering relationship similarity (R) complements [16], and the reinforcement framework can effectively integrate them for name translation. Lastly, considering relationship context similarity (RC) was effective in terms of precision, but shows relatively lower recall compared to ours without RC (*i.e.*, E+EC+R). This can be explained by the sensitivity of RC to data sparsity—When data are sparse, *e.g.*, for rare name relationship pairs, the number of documents where such pairs co-occur is small and using RC may not be sufficiently precise to eliminate rare but relevant pairs (*i.e.*, low recall). However, when data are sufficiently abundant, RC can be a very effective device for precise name translation identification.

In addition, we present representative translation examples in Figure 1, where the shaded cells indicate the correctly translated pairs. We can observe that each notion of E, EC, R, and RC contributes to translation accuracy.

5.2.2 Impact of NE Frequency

In this section, we evaluated the performances of all approaches for high frequency and low frequency NEs to understand how the frequency of an NE affects the performances. To build high frequency and low frequency NE sets, we divided our 500 person NE evaluation dataset into two equally sized subsets according to the frequencies of their English corpus parts (top 250 NEs and bottom 250 NEs). The average numbers of frequency of high frequent NE set and low frequent NE set in our experiment are 213.62 and 27 respectively.

Table 4 presents our experimental results. *HF* shows the results of high frequent NEs and *LF* shows the results of low frequent ones. *Ratio* shows the “decreasing ratio” on the two subsets, *HF* and *LF*. To compute this, we divide the decrease from the score from the high frequency NEs to low frequency ones by the score from the high frequency NEs. That is, the lower decreasing ratio is, the more robust the scheme is over varying NE frequencies. As the

Table 4: Precision, Recall, F1 score of each approach using high frequent (HF) and low frequent (LF) NEs. In both high frequent and low frequent NEs, both of our approaches outperform all other approaches.

	Precision			Recall			F1 Score		
	HF	LF	Ratio	HF	LF	Ratio	HF	LF	Ratio
EC	0.1360	0.0480	0.6465	0.3063	0.1087	0.6451	0.1884	0.0667	0.6461
E	0.3364	0.2913	0.1341	0.3333	0.3261	0.0217	0.3348	0.3077	0.0811
Shao	0.3765	0.1535	0.5922	0.5766	0.4022	0.3025	0.4555	0.2222	0.5122
You	0.4845	0.2778	0.4267	0.4234	0.3804	0.1015	0.4519	0.3211	0.2895
E+EC+R	0.5033	0.3939	0.2173	0.6847	0.5652	0.1745	0.5802	0.4643	0.1997
E+EC+R+RC	0.5274	0.4189	0.2059	0.6937	0.5326	0.2322	0.5992	0.4689	0.2175

E approach does not use contextual similarity, its performance will not be affected by NE frequency. Considering this point, we also show the second lowest scores in **bold**.

As Table 4 shows, our approaches, E+EC+R and E+EC+R+RC, clearly outperform other approaches in both high frequent and low frequent NEs. The *Ratio* results also show that our approaches are relatively more robust over most frequencies.

5.2.3 Effectiveness of Reinforcement Model

We evaluated the contribution of our reinforcement model using MRR [18], the average of the reciprocal ranks of the query results as the following formula:

$$\text{MRR} = \frac{1}{|Q|} \sum_{q \in Q} \frac{1}{\text{rank}_q}$$

where q is a ground truth translation pair, Q is the set of ground truth translation pair, and rank_q is the rank of the right translation answer in the list of translation candidates returned by our approach.

Specifically, we compared the MRR score of our framework using initial (or baseline) similarity matrix (R_0) with the score using reinforced similarity matrix (R_∞).

Table 5: MRR score of baseline and reinforced matrices using E+EC+R+RC. Our reinforcement model effectively improves MRR scores in all cases.

Set	MRR	
	Baseline R_0	Reinforced R_∞
$\gamma = 0.1$	0.7113	0.7459
$\gamma = 0.2$	0.7113	0.7473
$\gamma = 0.3$	0.7113	0.7408
$\gamma = 0.4$	0.7113	0.7456
$\gamma = 0.5$	0.7113	0.7519
$\gamma = 0.6$	0.7113	0.7473
$\gamma = 0.7$	0.7113	0.7438
$\gamma = 0.8$	0.7113	0.7340
$\gamma = 0.9$	0.7113	0.7300
$\gamma = 1.0$	0.7113	0.7246
Average	0.7113	0.7411

Table 5 summarizes our experimental results. We observed an increase in MRR scores in all cases after applying our reinforcement model, which indirectly shows the effectiveness of our reinforcement model. Note, as the initial matrix R^0 is not affected by γ , the MRR scores of all baselines are identical.

5.2.4 Effectiveness of Matching Approach

We also evaluated the effectiveness of matching a soft decision matrix R into a hard decision matrix R^* , using precision, recall, and F1-score. As matching translations using R can occur one-to-many assignment problem, jointly finding translations for all source NEs can avoid many errors that may occur in Shao.

Table 6 shows the evaluation results of E+EC+R+RC. Our holistic approach using R^* shows consistently higher precision than R . The average number of returned translation pairs is 54.00 with R and 45.4 with R^* . Among these pairs, the number of correctly returned pairs is 24.00 and 23.60 with R and R^* respectively. These results indicate that R^* can effectively filter some wrong translation pairs— on average, a 15.56% reduction of incorrect translation pairs.

In overall, E+EC+R+RC using R^* shows consistently higher F1-score which indicates our matching algorithm improves the translation quality.

5.2.5 Influences of Parameters

In this section, we study our performance over varying γ to see the effect of relationship context similarity (RC). We also varied θ for all γ values and found $\theta = 0.25$ to be optimal for most γ values, suggesting the robustness of our framework.

Table 7: Influence of the parameter γ of our approach E+EC+R+RC.

	Precision	Recall	F1-Score
$\gamma = 0.1$	0.4569	0.6222	0.5263
$\gamma = 0.2$	0.4741	0.6155	0.5349
$\gamma = 0.3$	0.4755	0.5990	0.5296
$\gamma = 0.4$	0.4892	0.5951	0.5364
$\gamma = 0.5$	0.4953	0.5800	0.5336
$\gamma = 0.6$	0.5017	0.5713	0.5338
$\gamma = 0.7$	0.5105	0.5713	0.5388
$\gamma = 0.8$	0.5137	0.5668	0.5384
$\gamma = 0.9$	0.5193	0.5583	0.5373
$\gamma = 1.0$	0.5141	0.5524	0.5321

Table 7 presents our results. A higher γ tends to lead to higher precision but lower recall, while a lower γ leads to high recall and low precision.

In overall, when $\gamma = 0.7$, E+EC+R+RC shows the best F1-score.

5.2.6 Other Type of NE

Table 6: Precision, Recall, F1-score of E+EC+R+RC using a soft decision matrix R and a hard decision matrix R^* . R^* shows higher precision and F1-score, and R shows higher recall.

	E+EC+R+RC					
	Precision		Recall		F1 score	
	Using R	Using R^*	Using R	Using R^*	Using R	Using R^*
$\gamma = 0.1$	0.3784	0.4569	0.6300	0.6222	0.4712	0.5263
$\gamma = 0.2$	0.3871	0.4741	0.6233	0.6155	0.4767	0.5349
$\gamma = 0.3$	0.3881	0.4755	0.6069	0.5990	0.4725	0.5296
$\gamma = 0.4$	0.4065	0.4893	0.6029	0.5951	0.4843	0.5364
$\gamma = 0.5$	0.4183	0.4953	0.5917	0.5800	0.4889	0.5336
$\gamma = 0.6$	0.4247	0.5017	0.5792	0.5713	0.4888	0.5338
$\gamma = 0.7$	0.4378	0.5105	0.5792	0.5713	0.4973	0.5388
$\gamma = 0.8$	0.4389	0.5137	0.5753	0.5668	0.4969	0.5384
$\gamma = 0.9$	0.4402	0.5193	0.5629	0.5583	0.4927	0.5373
$\gamma = 1.0$	0.4391	0.5141	0.5550	0.5524	0.4890	0.5321

To check the extension possibility of our approaches, we compared the performances of all approaches using the location type of named entities in the same corpora. We first extracted location type of NEs by using the same NE taggers. We then randomly selected 100 English location NEs from the English corpus and found their Chinese translations manually. As a result, 86 of their Chinese translations were found from the Chinese corpus. We used same parameters used in the person NE translation.

Table 8: The measure scores of all approaches computed using 100 location type of NEs. E+EC+R+RC outperforms all other approaches.

	Precision	Recall	F1 Score
EC	0.4684	0.4302	0.4485
E	0.6667	0.2558	0.3697
Shao	0.6557	0.4651	0.5442
You	0.4328	0.3372	0.3791
E+EC+R	0.6623	0.5930	0.6258
E+EC+R+RC	0.6709	0.6153	0.6424

Table 8 presents our experimental results. Compared to the Table 3, EC, E, and Shao show high precision (more than 50% precision improvement), while You shows the worst precision improvement. These results indicates that each translation similarity measure has a different contribution for different types of NEs— EC and E have big contribution for location type of NEs, and R has big contribution for person type of NEs.

As also shown in the Table 8, our proposed approaches, E+EC+R and E+EC+R+RC, clearly outperform other approaches. Specifically, E+EC+R+RC shows the best performance over all evaluation metrics. This result indicates that our approach, combining and optimizing many similarity metrics, is robust and effective in cross-type NE translation.

Figure 2 shows the difference of F1-score of each approaches. Our two approaches clearly outperform other approaches in cross-type NE translation.

5.2.7 Fuzzy Matching

In this section, we study our performances using fuzzy matching not exact matching. Fuzzy matching is the matching considering

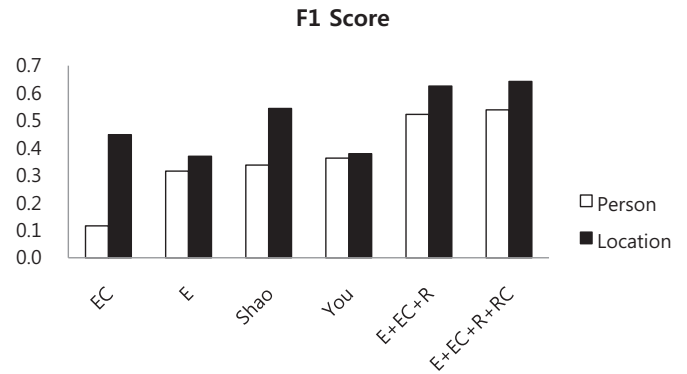


Figure 2: The difference of F1-Score between person type of NE and location type of NE.

partial name match as the right translation. For example, if our approach returns a translation pair (Barack Obama, 奥巴马), the exact matching algorithm considers it as a wrong translation, as 奥巴马 is Obama not Barack Obama, while fuzzy matching algorithm considers it as a right translation.

Table 9 shows the evaluation results. Compared to the Table 7, fuzzy matching results show higher precision, but lower recall. It is because both of the number of pairs that have right translation and the number of correctly returned pairs are increased, but the former is larger than the latter.

6. CONCLUSION AND FUTURE WORK

This paper studies the problem of mining named entity translations from comparable corpora. Current state-of-the-art approaches are holistic, as they combine entity, entity context, and relationship similarity. In this paper, we propose a new holistic approach, by combining all similarity types used and additionally considering relationship context similarity between named entities. Specifically, named entity graphs are first constructed from comparable corpora. Entity and entity context similarities are then calculated from every pair of bilingual named entities. A reinforcing method is utilized to reflect relationship and relationship context similarities between named entities. According to our experimental results, our new

Table 9: The performance of E+EC+R+RC using fuzzy matching. Fuzzy matching results show higher precision but lower recall.

	Precision	Recall	F1-Score
$\gamma = 0.1$	0.5292	0.4245	0.4694
$\gamma = 0.2$	0.5506	0.4214	0.4748
$\gamma = 0.3$	0.5542	0.4127	0.4708
$\gamma = 0.4$	0.5709	0.4099	0.4748
$\gamma = 0.5$	0.5819	0.4012	0.4726
$\gamma = 0.6$	0.5859	0.3951	0.4691
$\gamma = 0.7$	0.5920	0.3922	0.4690
$\gamma = 0.8$	0.5963	0.3892	0.4681
$\gamma = 0.9$	0.6037	0.3835	0.4665
$\gamma = 1.0$	0.5991	0.3806	0.4627

holistic graph-based approach significantly outperforms previous approaches and is extensible for various types of NEs.

We leave the followings as future work:

- **Comparable Corpora:** Currently, we used English and Chinese Gigaword Corpora published by the Xinhua News Agency which contains news articles from January 2008 to December 2008. Though they can be regarded as comparable corpora, they do not always talk about the same or related topics. As a result, there can be some noises that disturb the NE translation quality. Using larger dataset can reduce the effect of noises and yield higher-quality NE translation. Building our approach using larger dataset remains as future work.
- **Measures:** To construct the context association similarity matrix, we used Log-Odd association measure and cosine similarity measure, as they were the winners among various kinds of measures reported in [13]. The quality of our approaches may improve by using other association measure and similarity measure. Testing the performances of our approaches using other measures remains as future work.
- **NEs with No Translation:** Some English names may not have translations in Chinese corpus and deciding whether the translation exists or not is a clear challenge. However, existing work focuses on finding the most likely translation and does not discuss or evaluate how to determine the case with no translation. Our model consider this "asymmetry" case, by discarding English names with similarity ranked under threshold θ , but did not compare with existing work as they are not designed for this problem. Evaluating the effectiveness of asymmetry pruning remains as future work.

Acknowledgments

This research was supported by Microsoft Research and the MKE (The Ministry of Knowledge Economy), Korea, under IT/SW Creative research program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2010-C1810-1002-0003).

7. REFERENCES

- [1] A. Baron and M. Freedman. Who is Who and What is What: Experiments in Cross-Document Co-Reference. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, EMNLP '08. Association for Computational Linguistics, 2008.
- [2] M. Diab and S. Finch. A statistical word level translation model for comparable corpora. In *Proc. of the Conference on Content-based Multimedia Information Access RIAO*, RIAO '00, 2000.
- [3] D. Feng, Y. Lü, and M. Zhou. A New Approach for English-Chinese Named Entity Alignment. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, EMNLP '04, pages 372–379. Association for Computational Linguistics, 2004.
- [4] P. Fung and L. Y. Yee. An IR approach for translating new words from nonparallel, comparable texts. In *Proceedings of the 17th international conference on Computational linguistics - Volume 1*, COLING '98, pages 414–420. Association for Computational Linguistics, 1998.
- [5] J. Gao, M. Li, and C.-N. Huang. Improved source-channel models for Chinese word segmentation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, pages 272–279. Association for Computational Linguistics, 2003.
- [6] L. Haizhou, Z. Min, and S. Jian. A joint source-channel model for machine transliteration. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, ACL '04. Association for Computational Linguistics, 2004.
- [7] F. Huang, S. Vogel, and A. Waibel. Automatic extraction of named entity translingual equivalence based on multi-feature cost minimization. In *Proceedings of the ACL 2003 workshop on Multilingual and mixed-language named entity recognition - Volume 15*, MultiNER '03, pages 9–16. Association for Computational Linguistics, 2003.
- [8] H. Ji. Mining name translations from comparable corpora by creating bilingual information networks. In *Proceedings of the 2nd Workshop on Building and Using Comparable Corpora: from Parallel to Non-parallel Corpora*, BUCC '09. Association for Computational Linguistics, 2009.
- [9] L. Jiang, S. Yang, M. Zhou, X. Liu, and Q. Zhu. Mining bilingual data from the web with adaptively learnt patterns. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2*, ACL '09, pages 870–878. Association for Computational Linguistics, 2009.
- [10] L. Jiang, M. Zhou, L.-F. Chien, and C. Niu. Named entity translation with web mining and transliteration. In *Proceedings of the 20th international joint conference on Artificial intelligence*, pages 1629–1634. Morgan Kaufmann Publishers Inc., 2007.
- [11] K. Knight and J. Graehl. Machine transliteration. *Comput. Linguist.*, 24:599–612, December 1998.
- [12] J. Kupiec. An algorithm for finding noun phrase correspondences in bilingual corpora. In *Proceedings of the 31st annual meeting on Association for Computational Linguistics*, ACL '93. Association for Computational Linguistics, 1993.
- [13] A. Laroche and P. Langlais. Revisiting context-based projection methods for term-translation spotting in comparable corpora. In *Proceedings of the 23rd International Conference on Computational Linguistics*, COLING '10. Association for Computational Linguistics, 2010.
- [14] D. Lin, S. Zhao, B. V. Durme, and M. Pasca. Mining Parenthetical Translations from the Web by Word Alignment. In *Proceedings of the 46nd Annual Meeting on Association*

for *Computational Linguistics*, ACL '08, pages 994–1002. Association for Computational Linguistics, 2008.

- [15] R. Rapp. Automatic identification of word translations from unrelated English and German corpora. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, ACL '99, pages 519–526. Association for Computational Linguistics, 1999.
- [16] L. Shao and H. T. Ng. Mining new word translations from comparable corpora. In *Proceedings of the 20th international conference on Computational Linguistics*, COLING '04. Association for Computational Linguistics, 2004.
- [17] P. Virga and S. Khudanpur. Transliteration of proper names in cross-language applications. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. ACM, 2003.
- [18] E. M. Voorhees. The TREC question answering track. *Nat. Lang. Eng.*, 7:361–378, December 2001.
- [19] S. Wan and C. M. Verspoor. Automatic English-Chinese name transliteration for development of multilingual resources. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 2*, ACL '98, pages 1352–1356. Association for Computational Linguistics, 1998.
- [20] D. B. West. *Introduction to Graph Theory*. Prentice Hall, second edition, 2000.
- [21] G.-w. You, S.-w. Hwang, Y.-I. Song, L. Jiang, and Z. Nie. Mining name translations from entity graph mapping. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10. Association for Computational Linguistics, 2010.
- [22] K. Yu and J. Tsujii. Bilingual dictionary extraction from Wikipedia. In *Machine Translation Summit XII*, 2009.