Machine Translation with Real-Time Web Search*

Lei Cui[1], **Ming Zhou**[2], **Qiming Chen**[3], **Dongdong Zhang**[2], and **Mu Li**[2]

[1]Harbin Institute of Technology, Harbin, P.R. China
leicui@hit.edu.cn
[2]Microsoft Research, Beijing, P.R. China
{mingzhou,dozhang,muli}@microsoft.com
[3]Shanghai Jiao Tong University, Shanghai, P.R. China
simoncqm@gmail.com

Abstract

Contemporary machine translation systems usually rely on offline data retrieved from the web for individual model training, such as translation models and language models. In contrast to existing methods, we propose a novel approach that treats machine translation as a web search task and utilizes the web on the fly to acquire translation knowledge. This end-to-end approach takes advantage of fresh web search results that are capable of leveraging tremendous web knowledge to obtain phrase-level candidates on demand and then compose sentence-level translations. Experimental results show that our web-based machine translation method demonstrates very promising performance in leveraging fresh translation knowledge and making translation decisions. Furthermore, when combined with offline models, it significantly outperforms a state-of-the-art phrase-based statistical machine translation system.

Introduction

Web intelligence (Zhong et al. 2000; Yao et al. 2001) is recognized as a new direction for scientific research and development to explore the fundamental roles and practical impacts of artificial intelligence. Web-based methods work on a web platform that contains a heterogeneous collection of structured, unstructured, semi-structured, and inter-related web documents consisting of multi-lingual texts. In addition, a huge amount of multi-lingual user generated content from social networks and a variety of websites is produced every day. This content provides a great deal of web evidence that can be viewed as implicit human knowledge.

Machine Translation (MT) has been studied for the past decade. Among all the branches of MT research, the most popular one is Statistical Machine Translation (SMT), whose fundamental building blocks are bilingual corpora. From bilingual corpora, statistical translation models and reordering models can be effectively trained. Typically, a SMT problem is formulated as the following: given a sentence in source-side language, find its translation in target-side language that maximizes the posterior probability computed in the log-linear framework (Och and Ney 2002). A bilingual corpora collection is the key for MT development. Conventional SMT systems usually leverage web mining techniques to construct bilingual corpora in an offline way. A large number of parallel/comparable webpages and documents can be crawled from the web (Resnik 1998; Resnik and Smith 2003; Shi et al. 2006), which has helped push forward MT development. In addition, monolingual texts on the web are also crawled for language modeling. Although web mining has been successfully utilized to collect data for SMT, there is still fresh data emerging every day that cannot be exploited in real-time with conventional mining techniques. Therefore, real-time acquisition of translation knowledge is required for contemporary MT systems.

In this paper, we provide a new perspective for MT with real-time web search. MT is treated as a web search task and parallel data is acquired via search engines at runtime. The advantage of our method is that no offline bilingual corpora are required. Instead, fresh web search results are leveraged to train the models on the fly. These models play similar roles to offline SMT models. Commercial search engines are utilized for information acquisition since they have extensively indexed and processed an enormous amount of real-time data that contains a great deal of fractional cross-lingual knowledge. Hence, search engines provide complementary information to conventional web mining. This idea was inspired by web-based Question Answering (QA) (Brill, Dumais, and Banko 2002; Dumais et al. 2002). In web-based QA, questions are rewritten into a set of queries. The queries are sent to search engines, from which page summaries (snippets) are collected and analyzed. The answer is extracted from the n-grams of snippets with the assistance of the data redundancy.

There are some challenges in collecting data on the fly from the web to guarantee the translation quality, such as speed and coverage. To this end, bilingual corpora are obtained on demand according to the source text. Specifically, the source text is first rewritten into a set of queries with some hint words/phrases to trigger translation related web search. This may return search results that contain parallel data. We extract bilingual word pairs, phrase pairs, and sentence pairs from these search results and form a bilingual corpus. Using this corpus, a translation table is constructed on the fly that contains bilingual phrase pairs. To generate sentence-level translations, target phrases covered by source

*This work was done while the first and third authors were visiting Microsoft Research.

phrases are assembled into a Boolean "AND" query. The Boolean query is also sent to search engines to find a proximate permutation of the target words and phrases. Therefore, the monolingual data on the web indexed by search engines is regarded as a language model. Finally, we use the standard bottom-up decoding algorithm to generate the translation for the whole sentence based on the log-linear model for MT.

Compared to conventional SMT methods, experimental results demonstrate that our web-based MT method has several promising advantages:

1. Parallel data obtained via web search is much fresher. In contrast, traditional web mining of bilingual corpora is relatively cumbersome and time-consuming. Hence the data may be out-of-date from a practical perspective.
2. Web-based MT gets parallel data from web search through source-side information. The retrieved parallel data has high domain resemblance to the source texts. Therefore, our web-based MT approach is able to provide in-domain translation knowledge that is indispensable for making translation decisions.
3. Web-based MT is a good supplement of current SMT approaches. When SMT systems are combined with web-based MT, the translation accuracy is significantly improved compared to only using offline models.

Related Work

During the past decades, the web was mainly utilized as a huge repository where bilingual data and monolingual data could be mined to improve translation modeling and language modeling for MT. An early attempt at web mining parallel data was introduced in the STRAND system (Resnik 1998; Resnik and Smith 2003), aiming to automatically identify parallel texts on the web. After that, web mining techniques were widely adopted in parallel data construction (Zhao and Vogel 2002; Munteanu and Marcu 2005; Shi et al. 2006; Lin et al. 2008; Jiang et al. 2009). One common property of these approaches is that they used web data in an offline fashion. Webpages were crawled from the web, in which parallel data was extracted to train statistical models. Distinct from previous research, Huang, Zhang, and Vogel (2005) proposed using web search to mine key phrase translations from a commercial search engine that had indexed trillions of webpages. Therefore, the obtained phrase-level parallel data was fresher. Following the research in (Huang, Zhang, and Vogel 2005), we propose an end-to-end machine translation approach that not only acquires phrase-level translations but also composes sentence-level translations via search engines.

Web search based methods were also successfully applied in QA research (Brill, Dumais, and Banko 2002; Dumais et al. 2002). Questions were rewritten into queries, so that search engines could be used to obtain search results where potential answers might exist in the snippets. An abundance of QA related features were leveraged to learn a ranking model that was capable of finding the correct answer to a question. Compared to web-based QA, our web-based MT is a more challenging task since MT is a structured learning problem, where the search space is much larger and the correct answer is not unique. It is difficult to directly apply previous approaches in this scenario. Therefore, we need to devise a systematic approach and design MT specific features to achieve good translation performance.

Web-based MT Approach

In this section, we explain our web-based MT approach in detail. Figure 1 sketches the high-level overview that illustrates the web-based MT. The system contains three steps: query formulation, phrase-level translation generation, and sentence-level translation generation. The input is text in a source-side language and the output is the translation of the source text. A variety of features are assembled in a ranking model that helps generate N-best translation results.

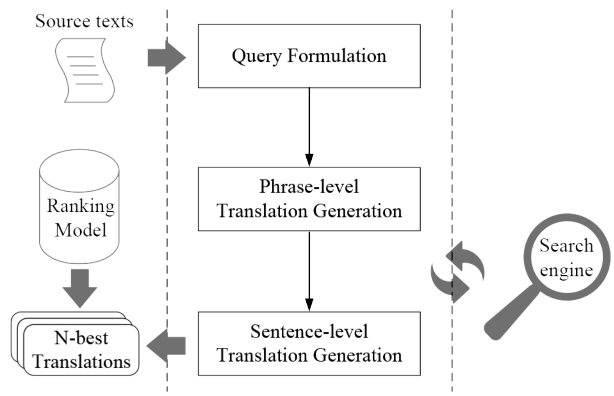

Figure 1: The overview of web-based MT approach

Query Formulation

In web-based MT, the aim of query formulation is to rewrite the source text into a group of n-gram queries for web search, which helps with finding relevant parallel data so that possible translations of the phrasal collocations can be extracted. To this end, the generated n-grams need to be searched using some hint words, which are able to trigger translation related searches. Currently, commercial search engines such as Google and Bing support question-style queries with patterns such as "what is X" or "how to X", where the question focus "X" can be answered via a proper search strategy. Similarly, in Chinese, the queries are "X 是什么" (what is X) and "怎样 X" (how to X). We select hint words/phrases that play two important roles: 1) Strong indicators to trigger translation related search; 2) Formulate a question-style query to obtain answers from search engines. Hence, we summarize several hint words/phrases in Table 1 to find potential translations of the queries from web search engines. For Chinese-to-English translation, we submit Chinese queries into web search engines. The motivation is that a large number of non-English speakers often discuss English translations on their native websites that are indexed in commercial search engines. Therefore, we are able to collect the relevant parallel data from the search results.

Figure 2: Search results of “失智症 英文翻译是什么”. The English phrases in the box are phrase-level translation candidates. The webpage on the right side is a search result from which example parallel sentences can be extracted.

Hint words/phrases templates
X 用英文怎样说 (how to say X in English)
X 英文翻译是什么(what is the English translation of X)
X 中英对照 (X Chinese-English parallel)
X 中英双语例句 (X Chinese-English example sentence)

Table 1: Hint words/phrases in Chinese where X is a source n-gram. The meaning of the query is described in English.

Phrase-level Translation Generation

Similar to phrase-based SMT methods, sentence-level translation is composed of multiple phrase-level translations. To find possible phrase-level translations from the web, the formulated queries are sent to popular search engines such as Google and Bing. Figure 2 shows a search example of the query “失智症 英文翻译是什么” (what is the English translation of 失智症 (dementia)) from Bing. In the search results, scattered translation knowledge can be found, such as “Alzheimer's disease” and “dementia”. These translations are mainly from wikis, weblogs, and other websites where the content is generated by active users. In addition, some search results come from online bilingual dictionaries, where examples of bilingual sentence pairs for English learners are obtained, as shown on the right side of Figure 2.

Therefore, parallel data is retrieved from two channels on the web: 1) Scattered and unstructured bilingual word/phrase pairs in the snippets; 2) Structured sentence pairs from online dictionaries. Sentence pairs from online dictionaries are easy to extract because example sentences satisfy some of the template constraints in webpages. For word/phrase pairs in the snippets, we use the frequency-distance model proposed in (Huang, Zhang, and Vogel 2005) to select potential translations of the queries. For each English phrase e in the snippets in Figure 2, a frequency-distance weight is computed as:

$$w(e) = \sum_{s_i} \sum_{f_{i,j}} \frac{1}{d(f_{i,j}, e)} \tag{1}$$

where s_i is a search snippet, the source phrase occurs in s_i is $f_{i,j}$ ($j \geq 1$ since f may occur several times in a snippet), and $d(f_{i,j}, e)$ is the distance between $f_{i,j}$ and e, i.e., the number of characters between the two phrases. We only retain the top-5 target phrases ranked by $w(e)$ scores.

After accumulating the extracted bilingual pairs (including word, phrase, and sentence pairs) to form a bilingual corpus, the HMM alignment method (Vogel, Ney, and Tillmann 1996) is performed. Although the corpus is very small, HMM still performs well since the parallel data is specially tailored for the source sentence and the vocabulary size is very small. With the word alignment, the phrase extraction method (Koehn, Och, and Marcu 2003) in SMT is used to extract phrase pairs. The procedure is fast because the corpus only contains hundreds of bilingual phrase/sentence pairs.

To train a web-based translation model $P(\bar{e}_1^I | \bar{f}_1^I, \mathbf{W})$ given the web $\mathbf{W}$, we use the maximum likelihood estimation method to compute the probability approximately as:

$$P(\bar{e}_i | \bar{f}_i, \mathbf{W}) \approx \frac{\text{Count}(\bar{f}_i, \bar{e}_i)}{\sum_{\bar{e'}_i} \text{Count}(\bar{f}_i, \bar{e'}_i)} \tag{2}$$

where $\{\bar{f}_i, \bar{e}_i\}$ is a phrase pair extracted from the parallel corpus collected from search engines, and $\text{Count}(\bar{f}_i, \bar{e}_i)$ is the count of $\{\bar{f}_i, \bar{e}_i\}$ in the corpus.

Sentence-level Translation Generation

With the phrase-level translations extracted from the retrieved search results, sentence-level translations are generated in which web search is utilized for language modeling. Conventional SMT approaches always collect a large amount of monolingual data to train an offline language model, which helps in making translation decisions and reordering. However, in the web-based MT approach, the language model is not pre-built offline. Instead, it is trained on-the-fly during the web search process since snippets from search engines can be leveraged. To use search engines for this purpose, we generate a set of Boolean queries during the MT decoding process. Suppose $f[i,j]$ denotes a source-side n-gram from i-th to j-th word and $e[i,j]$ is a translation of $f[i,j]$. When two source-side phrases $f[i,k]$ and $f[k+1,j]$ are combined into a larger phrase $f[i,j]$, the target-side phrases $e[i,k]$ and $e[k+1,j]$ are also combined in either monotonic order or reversed order to compose $e[i,j]$. At this time, a Boolean search query "$e[i,k]$ AND $e[k{+}1,j]$" is used to retrieve possible combinations of $e[i,k]$ and $e[k+1,j]$ from search snippets. We collect these snippets for all the phrases combined in a bottom-up decoding process based on Inversion Transduction Grammars (ITG) (Wu 1997) within a distortion limit Λ. The details are shown in Algorithm 1. Given an input sentence with n words (line 1), we first obtain phrase-level translations for all source-side phrases with a length no more than Δ (line 2-4). Then, for each span $[i,j]$ where $j-i \leq \Lambda$, we traverse all of its sub-span pairs to generate Boolean queries and combine two sub-spans in both monotonic and reversed orders (line 8-10). For each span $[i,j]$ where $j-i > \Lambda$, its sub-span pairs are combined only in the monotonic order (line 11-12). To speed up the decoding process, we use beam search algorithms that only retain the top-N translations for each source-side n-gram $f[i,j]$ (line 15). Finally, the N-best translations with the highest scores are obtained (line 18). In our experiments, Δ was set to 3 and Λ was set to 10.

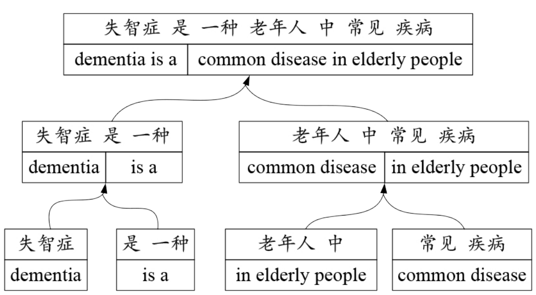

Figure 3: An example of the bottom-up decoding process, in which phrases are combined in a monotonic or reversed way.

Figure 3 illustrates an example that gives several phrase pairs and combines them from bottom to top. When "in elderly people" and "common disease" are combined, the query and search results are shown in Figure 4. The snippets show that the target-side phrases appear with contextual information that is appropriate to train a web-based language model on the fly. An m-gram language model is used to predict the next word based on the information from previous $m-1$ words. We use the stupid backoff approach (Brants et al. 2007) for web-based language model smoothing. The

Algorithm 1 Decoding in web-based MT

```
1:  Given an input sentence f[0, n−1]
2:  for all i, j s.t. 0 ≤ j − i ≤ Δ do
3:      Obtain phrase-level translation for f[i, j] as e[i, j]
4:  end for
5:  for length l ← 1 ... n − 1 do
6:      for all i, j s.t. j − i = l do
7:          for all k ∈ [i, j] do
8:              if l ≤ Λ then
9:                  ▷ generate a Boolean query "e[i, k] AND e[k +
                    1, j]" and retrieve search snippets.
10:                 ▷ combine e[i, k] and e[k, j] into e[i, j] in both
                    monotonic and reversed orders.
11:             else
12:                 ▷ combine e[i, k] and e[k + 1, j] into e[i, j] only
                    in the monotonic order.
13:             end if
14:         end for
15:         ▷ keep top-N translations for f[i, j] via web searched
            information.
16:     end for
17: end for
18: return e[0, n−1]
```

computation is:

$$p(w_i|w_{i-m+1}^{i-1}) = \begin{cases} \frac{c(w_{i-m+1}^{i})}{c(w_{i-m+1}^{i-1})} & \text{if } c(w_{i-m+1}^{i}) > 0 \\ \alpha \cdot p(w_i|w_{i-m+2}^{i-1}) & \text{otherwise} \end{cases} \tag{3}$$

where $c(\cdot)$ is the count of an m-gram from the snippets. α is backoff weight that is set to the same value ($\alpha = 0.4$) as in the original paper. Due to its fast and simple implementation, these m-gram counts can be accumulated during the translation procedure, so that the language models will be more accurate as more snippets are retrieved.

Model

Finally, we use a ranking model to generate N-best translations based on the proposed features. Given a sentence f in a source language, the web-based MT system searches for a target translation $\hat{e}$ that has the maximum posterior probability, given the web $\mathbf{W}$ as the knowledge source:

$$\begin{aligned} \hat{e} &= \arg\max_{e \in \mathbf{H}} P(e|f, \mathbf{W}) \\ &= \arg\max_{e \in \mathbf{H}} \{\sum_i w_i \cdot \log \phi_i(f,e)\} \end{aligned} \tag{4}$$

where $\mathbf{H}$ is the hypothesis space. The probability is given by the standard log-linear model in MT (Och and Ney 2002), where $\phi_i(f,e)$ denotes a feature function and w_i is the corresponding feature weight. Distinct from conventional machine translation methods, web-based MT utilizes web-based features, including a web-based translation model (two directions) and a web-based language model. In addition, we also use the standard features, including word penalty, phrase penalty, and NULL penalty.

Figure 4: Search results from Bing using the query "in elderly people" AND "common disease".

Experiments

We evaluated our web-based MT approach on two Chinese-to-English machine translation tasks: 1) phrase-level translation, which verifies the capability of web-based MT to collect fresh translation knowledge, and 2) sentence-level translation, which aims to further investigate whether web searched data provides in-domain translation knowledge according to the source text.

Setup

We evaluated phrase-level translation quality using a public dataset released by (Huang, Zhang, and Vogel 2005) and compared the accuracy to that approach. The dataset contains 310 Chinese phrases from 12 domains and their English translations. A hypothesized translation is considered correct when it matches one of the reference translations.

The performance of sentence-level translation was compared to a state-of-the-art SMT baseline, which is an in-house phrase-based SMT decoder based on ITG with a lexicalized reordering model (Xiong, Liu, and Lin 2006). The baseline system was trained on 30 million Chinese-English sentence pairs including training data from NIST and IWSLT, as well as a large amount of parallel data mined from the web (Lin et al. 2008; Jiang et al. 2009). An offline 4-gram language model was trained using web mined monolingual data, which contains 12 billion tokens in total. The development data for parameter tuning with MERT (Och 2003) is a human annotated dataset with 1,483 sentences for both the baseline and our web-based MT system. The testing data includes 4 datasets from different genres, which are shown in Table 2. Case-insensitive BLEU4 (Papineni et al. 2002) was used as the evaluation metric of the translation quality. A statistical significance test was performed using the bootstrap re-sampling method (Koehn 2004).

We conducted web search via a distributed system that contained 48 nodes, on which a large number of search queries were sent to Bing in order to obtain search results in parallel. We retrieved top-10 search results for each query in phrase-level translation and sentence-level translation.

Test Datasets	#Sentences	Genre
NIST 2008	1,357	newswire and weblog
IWSLT 2010	504	colloquial
Wikipedia	1,570	formal introduction
Novel	1,200	literature

Table 2: Testing data used in the experiments.

Phrase-level Translation Quality

We used Bing to search each phrase with the four hint phrases shown in Table 1. Table 3 shows the phrase-level translation accuracy of our method from the top-5 hypotheses, compared to the results reported in (Huang, Zhang, and Vogel 2005) as well as the baseline SMT system. Both our method and Huang's method outperformed the SMT method, which shows the effectiveness of fresh web search results. Although our method did not perform as well as Huang's method in top-1 accuracy, it achieved much higher accuracy (+6.1) within top-5 results only with four simple hint words. The advantage of our method is that we directly leverage some effective hint words/phrases so that our approach can be applied to common words/phrases translation. However, Huang's method has the limitation that it only focuses on key phrase translations such as entities. Therefore, our method is an effective combination of Huang's method and traditional SMT methods. In addition, our method only needed 40 search results to achieve this accuracy, compared to 165 search results in (Huang, Zhang, and Vogel 2005), which saves a lot of time. Hence, our approach is more capable of exploiting the potential of search engines. This also confirms that the Out-Of-Vocabulary (OOV) rate of web-based MT is much lower than that of the SMT methods.

Methods	Accuracy of the Top-N (%)				
	Top1	Top2	Top3	Top4	Top5
Baseline SMT	61.0	67.1	71.6	73.9	74.5
Huang's method	80.0	86.5	89.0	90.0	90.0
Ours	73.2	84.5	90.3	93.9	96.1

Table 3: Phrase-level translation accuracy.

Sentence-level Translation Quality

We evaluated the sentence-level translation accuracy on the datasets from different genres, with the results shown in Table 4. Although the web-based method does not utilize any offline training data, its translation performance is comparable to an SMT baseline trained with huge parallel data. This confirms that the translation knowledge retrieved from web search is extremely effective, especially on the Novel dataset. The main reason is that the Novel dataset contains many named entities, such as person names and place names, where our web-based method shows its superiority. Moreover, we also add an offline-trained reordering model, translation model, and language model into the log-linear model in order to investigate whether web-based methods can help conventional SMT systems. The results show that

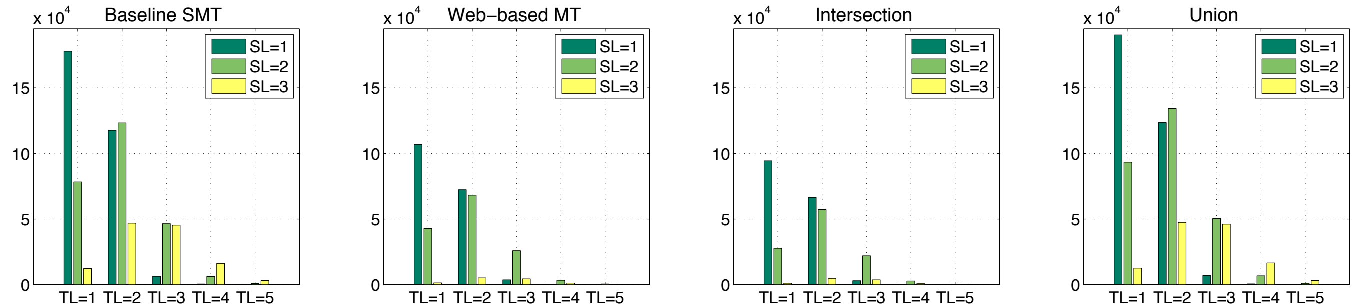

Figure 5: Phrase pair coverage of baseline SMT and web-based MT, as well as their intersection and union filtered by the Novel dataset. "SL" denotes the number of words in the source phrase and "TL" denotes the number of words in the target phrase. E.g., for a phrase pair ⟨失智症, alzheimer's disease⟩, it is a "SL=1,TL=2" pair.

web-based methods bring additional benefits to the translation systems. The best performance was achieved when we used the baseline features as well as the web-based features (web-based TM + web-based LM) in a domain adaptation way (Koehn and Schroeder 2007), where web-based TM and web-based LM were viewed as in-domain models while the baseline TM and LM were viewed as general models.

Settings	NIST	IWSLT	Wiki	Novel
SMT(RM+TM+LM)	30.48	39.95	27.66	26.26
WBMT(wTM+wLM)	25.23	35.2	23.21	24.27
WBMT+RM	26.08	35.80	23.93	25.11
WBMT+TM	27.79	38.51	26.32	26.03
WBMT+LM	27.03	37.93	25.18	25.84
SMT+wTM	30.81	40.85	28.11	28.21
SMT+wLM	30.51	40.41	27.78	27.54
SMT+WBMT	**30.88**	**41.03**	**28.18**	**28.68**

Table 4: End-to-end evaluation on sentence-level translation accuracy in BLEU% ($p < 0.05$). WBMT denotes web-based MT that utilizes web-based TM (wTM) and web-based LM (wLM). RM, TM, and LM denote the offline reordering model, translation model, and language model, respectively.

To further explore how the web helps MT, we looked into the details of web searched data. We compared the phrase pairs extracted from offline training data and those extracted from online retrieved data, as well as their intersection and union, with the results shown in Figure 5. We found that the phrase pair coverage of SMT was much larger than the web-based method, which is the main reason for the translation performance gap. However, we observed that the web-based method can provide new phrase pairs as the union shows, although the amount of offline training data was huge in the experiments. This confirms that using search engines can always deliver up-to-date translation knowledge. For example, the translation of a Chinese phrase "布莱克斯泰勃" in the Novel dataset does not exist in the traditional parallel corpus. However, using our web-based method, the translation can be easily found from a search snippet as:

> ...尽管菲利普属中产阶级，为跋足青年，自十岁成孤儿，在**布莱克斯泰勃**（Blackstable）与他伯父伯母住在一起，后来在寄宿学校及 ...

The search result is from a Chinese website *docin.com* that is full of ever-increasing user generated content, which is beyond the capacity of conventional web mining techniques such as (Lin et al. 2008). Furthermore, we analyzed the translation results in the testing datasets. Two examples from the Wikipedia dataset in Table 5 demonstrate that web searched parallel data is indeed more domain relevant to the translated texts. In the SMT system, although "重要性" and "阻力" are translated correctly in a literal manner, they are not the most accurate choice for the collocation "法律重要性" and "重力阻力". In contrast, the combined system provides more in-domain translation knowledge, hence it significantly outperforms the SMT baseline.

Src	功能性的不同有更多的法律重要性
Ref	functional differences have more legal significance
SMT	functional differences have more legal importance
SMT+WBMT	functional differences have more legal significance
Src	从而尽量减少重力阻力造成的损失
Ref	thereby minimizing losses due to gravity drag
SMT	to minimise damage caused by the gravity resistance
SMT+WBMT	so as to minimize the damage caused by gravity drag

Table 5: An example illustrating in-domain knowledge, where "Src" is the source sentence and "Ref" is the reference translation.

Conclusion and Future Work

In this paper, we propose a web-based MT approach using real-time web search, where parallel data is completely obtained via web search on the fly. The web-based MT method is able to provide fresh data and incorporate in-domain translation knowledge. Furthermore, it is language independent once the search engines are available. Experimental results demonstrated that our method leads to substantial improvements compared to a state-of-the-art baseline.

In the future, we will try to devise more intelligent search schemes that can alleviate the data coverage problem in web-based MT. Furthermore, we also plan to address the challenge that translation results may vary due to the ever-changing web search results.

Acknowledgments

We are grateful to the anonymous reviewers for their insightful comments. We thank Shujie Liu, Fei Huang (BBN), Nan Yang, Yajuan Duan, and Hong Sun for their helpful discussions. We also thank Ruyao Wu (HangZhou DianZi University) for picture design.

References

Brants, T.; Popat, A. C.; Xu, P.; Och, F. J.; and Dean, J. 2007. Large language models in machine translation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 858–867. Prague, Czech Republic: Association for Computational Linguistics.

Brill, E.; Dumais, S.; and Banko, M. 2002. An analysis of the askmsr question-answering system. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, 257–264. Association for Computational Linguistics.

Dumais, S.; Banko, M.; Brill, E.; Lin, J.; and Ng, A. 2002. Web question answering: Is more always better? In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '02, 291–298. New York, NY, USA: ACM.

Huang, F.; Zhang, Y.; and Vogel, S. 2005. Mining key phrase translations from web corpora. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 483–490. Vancouver, British Columbia, Canada: Association for Computational Linguistics.

Jiang, L.; Yang, S.; Zhou, M.; Liu, X.; and Zhu, Q. 2009. 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*, 870–878. Suntec, Singapore: Association for Computational Linguistics.

Koehn, P., and Schroeder, J. 2007. Experiments in domain adaptation for statistical machine translation. In *Proceedings of the Second Workshop on Statistical Machine Translation*, 224–227. Prague, Czech Republic: Association for Computational Linguistics.

Koehn, P.; Och, F. J.; and Marcu, D. 2003. Statistical phrase-based translation. In *Proceedings of HLT-NAACL 2003 Main Papers*, 48–54. Edmonton: Association for Computational Linguistics.

Koehn, P. 2004. Statistical significance tests for machine translation evaluation. In Lin, D., and Wu, D., eds., *Proceedings of EMNLP 2004*, 388–395. Barcelona, Spain: Association for Computational Linguistics.

Lin, D.; Zhao, S.; Van Durme, B.; and Paşca, M. 2008. Mining parenthetical translations from the web by word alignment. In *Proceedings of ACL-08: HLT*, 994–1002. Columbus, Ohio: Association for Computational Linguistics.

Munteanu, D. S., and Marcu, D. 2005. Improving machine translation performance by exploiting non-parallel corpora. *Comput. Linguist.* 31(4):477–504.

Och, F. J., and Ney, H. 2002. Discriminative training and maximum entropy models for statistical machine translation. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, 295–302. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics.

Och, F. J. 2003. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, 160–167. Sapporo, Japan: Association for Computational Linguistics.

Papineni, K.; Roukos, S.; Ward, T.; and Zhu, W.-J. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, 311–318. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics.

Resnik, P., and Smith, N. A. 2003. The web as a parallel corpus. *Computational Linguistics* 29(3):349–380.

Resnik, P. 1998. *Parallel strands: A preliminary investigation into mining the web for bilingual text*. Springer.

Shi, L.; Niu, C.; Zhou, M.; and Gao, J. 2006. A dom tree alignment model for mining parallel data from the web. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, 489–496. Sydney, Australia: Association for Computational Linguistics.

Vogel, S.; Ney, H.; and Tillmann, C. 1996. Hmm-based word alignment in statistical translation. In *Proceedings of the 16th Conference on Computational Linguistics - Volume 2*, COLING '96, 836–841. Stroudsburg, PA, USA: Association for Computational Linguistics.

Wu, D. 1997. Stochastic inversion transduction grammars and bilingual parsing of parallel corpora. *Comput. Linguist.* 23(3):377–403.

Xiong, D.; Liu, Q.; and Lin, S. 2006. Maximum entropy based phrase reordering model for statistical machine translation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, 521–528. Sydney, Australia: Association for Computational Linguistics.

Yao, Y.; Zhong, N.; Liu, J.; and Ohsuga, S. 2001. Web intelligence (wi) research challenges and trends in the new information age. In *Web Intelligence: research and development*. Springer. 1–17.

Zhao, B., and Vogel, S. 2002. Adaptive parallel sentences mining from web bilingual news collection. In *Proceedings of the 2002 IEEE International Conference on Data Mining*, ICDM '02, 745–. Washington, DC, USA: IEEE Computer Society.

Zhong, N.; Liu, J.; Yao, Y. Y.; and Ohsuga, S. 2000. Web intelligence (wi). In *Computer Software and Applications Conference, 2000. COMPSAC 2000. The 24th Annual International*, 469–470. IEEE.