Exploring the Further Integration of Machine Translation in Multilingual Information Access

Daqing He
School of Information Sciences
University of Pittsburgh, Pittsburgh PA 15260, USA
dah44@pitt.edu

Dan Wu
School of Information Management
Wuhan University, Wuhan 430072, China
woodan@whu.edu.cn

ABSTRACT

Machine Translation (MT) has been identified as a very important related technology for Multilingual Information Access (MLIA). Over the past decade, the usages of MT in MLIA are still largely concentrated on its capabilities for document translation, selection and examination. In this paper, by using a common evaluation framework, we explored the applications of MT in several unexplored or underexplored areas, which include query translation, relevance feedback, and out of vocabulary term resolution. Our experimental results demonstrate the unique contributions that MT can provide in those areas, and at the same time raise more interesting questions about how MT can be optimally integrated with MLIA.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – Search Process

General Terms
Design, Experimentation, Languages, Performance

Keywords
Multilingual Information Access, Machine Translation, Query Translation, Relevance Feedback, Out-Of-Vocabulary Term.

1. MACHINE TRANSLATION IN MLIA

With vast amount of multilingual information on the Web and the ability to read the results, it is natural for users to issue queries in one language, and access documents in other language(s). This so called Multilingual Information Access (MLIA) has been an active research area for more than a decade. Translation has played a very important role in MLIA. Although there are techniques for achieving MLIA without actually involving translation [1], most MLIA techniques often rely on translation methods to cross the language barriers between a query and the documents. Depends on whether it is the query, the documents, or both that are translated, we have document translation based MLIA (DT-MLIA), query translation based MLIA (QT-MLIA) and interlingual MLIA (IL-MLIA). Many resources have been exploited for the translation task, among which the most commonly used are machine-readable dictionary (MRD), parallel or comparable corpora, and machine translation (MT). MRD is probably the most commonly used in experiment setting, especially for translating queries [2]. However, considering that several companies such as Google and Yahoo are actively prompting their multilingual MT services on the Web, MT is probably the most easily accessible translation resource among the above three on the Web between commonly used language pairs. Therefore, it is important to examine the usages of MT in MLIA.

We believe that from the view point of MLIA, MT can be viewed as either a component of MLIA or as one of the major translation resources for MLIA. This motivated us to look at the usages of MT in MLIA in the steps of QT-MLIA such as query translation, relevance feedback, interactive MLIA, and out-of-vocabulary (OOV) term translation. The reason for concentrating on QT-MLIA is because it is query translation that makes MLIA different to monolingual information access. QT-MLIA reveals its translation process to the users so that the users can feel and build up more control of the search process. To the wide range of MLIA users, it is probably QT-MLIA that they will most interact with if they want to perform MLIA. Our goal is to obtain more insights about the wide range usages of MT in MLIA, and to help us and the community to identify promising future directions for both MT and MLIA.

The remainder of this paper is organized as follows. We discuss in detail our research topics of MT in MLIA and the corresponding experiment settings in Section 2. And in Section 3, we will present our insights to the usages of MT in MLIA, and conclude with brief highlights of our future work.

2. RESEARCH TOPICS AND EXPERIMENTS

2.1 General Experiment Settings

There are two research angles in our experiments and the discussions here. The first one examines the effectiveness of different query lengths on the studied techniques. The motivation is that Web search queries are often short, but queries in TREC like evaluation frameworks are often much longer. Previous studies show techniques often are more suitable for certain types of query lengths. Our experiments conducted under different query lengths, therefore, will provide more insights about the applicability of the techniques. The second angle is about technique integration. Over the years of active research, there have been many techniques and methods developed for MLIA. One insight obtained in the literature is that different techniques and methods often can be combined to obtain further

iConference'2010, February 3–6, 2010, Illinois, USA.

improvement. Therefore, when possible, we will examine the integration of complementary techniques in our studies.

All studies reported in this paper were performed on the same experiment environment. This not only helped to simplify the experiment design, but also made it possible to compare results across several studies.

Our experiments were performed between English queries and Chinese documents. The test collection contains documents from TDT4 and TDT5 Multilingual corpora. All the documents in the collection are news articles in the time period of 2000 to 2003 from several news agencies including Xinhua News Agency, Zaobao News Agency, China Broadcasting System, etc. The two TDT collections contain 83,627 Chinese documents and corresponding number of machine translation documents generated by ISI MT system. The collections also contain 306,498 English documents from several English news agencies at the same time period as Chinese documents.

We selected 44 TDT English topics and manually translated them into Chinese for monolingual Chinese search. These topics were also converted into TREC topic style with title, description and narrative fields for our study of the effect of different query lengths. Queries were automatically extracted from the topics with short queries containing titles only (T query), medium queries with title and description fields (TD query), and long queries with all the three fields (TDN query). The average length of the queries were: T query (4 terms), TD query (27 terms), and TDN query (127 terms).

The bilingual MRD used for our dictionary-based MLIA was an English-Chinese lexicon generated from a parallel bilingual corpus automatically [3]. The dictionary contains 126,320 English entries with translation probabilities for each Chinese translation alternative. During the translation of the queries with the MRD, to remove low probability translations which often are noises, a fixed threshold called Cumulative Probability Threshold (CPT) was selected. A threshold of 0 corresponds to the using the single most probable translation (a well-studied baseline), and a threshold of 1 corresponds to the use of all translation alternatives in the dictionary. In order to improve the coverage of the dictionary as much as possible, we adopted the back-off translation strategy [4] during the translation of the query terms.

In MRD based MLIA, we adopted a named entity translation component based on information extraction (IE) techniques. The NE component is designed to provide two functions in the MRD based query translation. The first one is to identify NEs in a given text, which could be queries, documents, or any parts of queries and documents. The function was provided by the NYU English and Chinese HMM-based name taggers trained on several years of ACE (Automatic Content Extraction) corpora. Both name taggers can identify names such as Person, Geo-Political Entity (GPE), Location, Organization, Facility, Weapon and Vehicle, and achieve about 87%-90% F-measure on newswire [5].

If translation enhancement (will be presented in section 2.3) was involved in the experiment, we selected TE-TWA method. The CPT threshold was 0.5, and λ was 0.5. Both of the two values were obtained via training [6]. If query expansion (QE) was involved in the experiments, we used the Indri's build-in PRF module which is based on Lavrenko's relevance model [7] . Depends on whether the QE is performed before and/or after query translation, we have pre-translation, post-translation and combined QE. The parameters in QE were set as top 20 terms from top 20 returned documents. The weight between original query and expanded terms is 0.5. This was based on our previous exploration of the parameters in Indri.

Unless mentioned specifically, the measure used was Mean Average Precision (MAP) over a ranked list. This measure is a commonly used evaluation measure in IR field. Statistical significance tests used in all our experiments were two tailed paired samples t-test, and we used $p\text{-value} < 0.05$ as the threshold for the statistically significance.

2.2 Topic 1: MT for Query Translation

The core step in QT-MLIA is the translation of queries, and MT can be integrated for translating queries. However, the effectiveness of using MT for translating queries comparing to MRD based methods is uncertain in previous studies [8]. Recently, both MT and MLIA have experienced rapid integration of statistical based language models and resources into their handling of translations. Statistical MT has become the state of the art for MT, and even some commercial MT systems such as Google Translate are statistical MT systems. Translation probabilities are widely used in MLIA for handling translation ambiguities or are even built as a part of the statistical language modeling for MLIA [2, 9, 10]. One important insight gained in MLIA from the usage of translation probabilities is that choosing multiple translations with their probabilities is a superior method than choosing only the top best translation. This insight actually to some degree argues against the current usage of MT output, which contains only one best translation for query terms or documents.

Therefore, the objective of this research topic is to examine again the effect of MT in query translation. We concentrate on using an out-of-box commercial MT system – Google Translate -- for the task. Our motivation is that if commercial MT systems have demonstrated their capabilities in translating queries, maybe effective MLIA capabilities can be easily constructed even by layman users. The users do not have to go through the steps of obtaining MRDs with high quality translation probabilities in order to perform MLIA. What they need is an online MT system.

Table 1: The MAP results of MT-based and MRD-based runs (* indicates that the improvement is statistically significant between MT Plain and MRD Base)

Run ID	T	TD	TDN
Mono Base	0.4739	0.5817	0.6215
MT Plain	0.4446*	0.5536*	0.6170*
MT QE-PreTrans	0.4922	0.5443	0.5580
MT QE-PostTrans	0.5284	0.6031	0.6292
MT QE-Combine	0.5604	0.5833	0.6001
MRD Base	0.3336	0.4251	0.4701
MRD QE-PreTrans	0.3714	0.4377	0.4477
MRD QE-PostTrans	0.4118	0.5080	0.5182
MRD QE-Combine	0.4415	0.5007	0.5170

The research questions associated with this study are: 1) when no other technique is integrated, can MT based query translation perform comparably to monolingual search or to MRD based query translation; 2) if performance enhancement technique such as query expansion (QE) based on pseudo relevance feedback (PRF) is used, would MT based query translation still work; and 3) is query length a factor that affects MT based query translation? When using the MT system for query translation in the experiments, we entered the whole query into the MT system at once.

The experiment revealed some interesting results. As shown in Table 1, when there is no RF performed, MT based run "MT-Plain" performed really well. Its performance values under three different query lengths were between 94% to 99% of monolingual run "Mono Base". This is comparable to the state of the art MRD based MLIA performance, which integrated many performance enhancement techniques. In addition, the MT based runs obtained significantly improvement over the plain MRD baseline "MRD Base" under all three different lengths of queries.

Query length affects MT based query translation methods. MT based query translation works the best with long queries. The performance of "MT Plain" can reach to 99% of that of "Mono Base". However, it is interesting to see that the superiority of MT method is shrinking along with the increasing of query length comparing to "MRD Base". Maybe this does not imply that MT method is not good for long queries, it probably just means that MRD method performs better with long queries too. The longer the queries are, the more information can be used in the MRD method to compensate the impact of translation ambiguities. Because the MT method worked well over short queries too, this helps to remove the worry that MT does not work when not much context is available for translation.

We know from previous studies in the literature, QE in general helps MRD-based methods. Our results in Table 2 show that QE was helpful to MT based method too. With QE, MT-based method was at least 90% of monolingual performance. Some QE methods even helped MT method to achieve over 100% of monolingual performance ("MT QE-Combine" at T queries achieved 120%). Therefore, with the help of a simple QE technique, most MT-based MLIA runs actually outperformed the corresponding monolingual runs. All these results confirm again that MT is useful for translating queries in MLIA.

Table 2: Comparison of QE methods for MT-based MLIA (* indicates that the improvement is statistically significant)

Run ID		Perc. Of Mono Base	Impr. over MT Plain
T	MT QE-PreTrans	103.86%	10.71%*
	MT QE-PostTrans	111.50%*	18.85%*
	MT QE-Combine	118.25%*	26.05%*
TD	MT QE-PreTrans	93.57%	-1.68%
	MT QE-PostTrans	103.68%	8.94%*
	MT QE-Combine	100.28%	5.36%
TDN	MT QE-PreTrans	89.78%*	-9.56%*
	MT QE-PostTrans	101.24%	1.98%
	MT QE-Combine	96.56%	-2.74%

Again, query length affects the performance of QE in MT based query translation. As shown in Table 2, similar to that in MRD-based MLIA, it is often the post-translation QE and/or the combined QE that are the best methods among the three MLIA QE methods. In fact, post-translation QE ("MT QE-PostTrans") seems to be helpful whatever the query length is, whereas pre-translation QE ("MT QE-PreTrans") only works for short queries and "MT QE-Combine" does not work for long queries. The reason why the longer queries don't work is that the translation resource is very good (the MT Plain condition is already with very high performance), thus there is less need for QE techniques to combat the translation failures.

Overall, not only MT is a reasonable query translation method in MLIA, but also it is a better tool for query translation than MRD, no matter what the query length is and whether QE is used.

2.3 Topic 2: MT for Relevance Feedback in MLIA

Revealed in the interactive track of Cross-Language Evaluation Forum experiments (such as in [11]), and demonstrated by Google's cross-language search engine inside Google Translate, a user who performs QT-MLIA tasks would not only need queries to be translated so that the retrieval can be performed, but also need the returned documents to be translated back so that the user can perform relevance judgment or document examination. This MLIA process with two translation steps imposes some interesting problems and opportunities about MT in MLIA, especially the MLIA relevance feedback (RF) techniques based on MT.

Translation Enhancement (TE) is one such relevance feedback technique in MLIA [6]. TE utilizes the MT outputs of relevant documents for improving query translation. It views that users' relevance judgments are performed on these MT outputs rather than on the original returned documents in another language. This understanding helps to treat the user confirmed relevant documents and their MT outputs as a small parallel corpus. By using a word alignment tool GIZA++ [12], individual words inside the relevant documents and their MT translations can be connected. Instances of translations of query terms and their probabilities can be extracted and integrated with the corresponding translation information in the original dictionary. This TE method is called Translation Extraction with Word Alignment (TE-TWA).

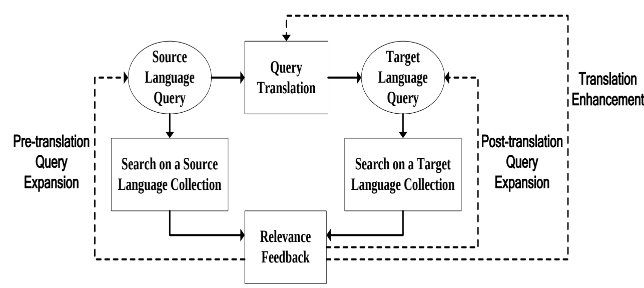

Figure 1: Translation Enhancement and Query Expansion (pre and post-translation) in MLIA

As shown in Figure 1, TE and various QE methods are applicable at different parts of the MLIA process, and they use different aspects of the relevant documents. The applications of their relevance feedback results are different too: QE results are integrated into the query, and TE results modify the query translation, respectively. Therefore, two important research questions to be answered are: 1) what is the effect of TE-TWA method; and 2) what is the effect of combining TE and QE in MLIA?

As shown in Table 3, our experiment results show that TE approach "TE-TWA" performed better than the plain MLIA baseline run "MRD Base". The differences were statistically significant with all three types of queries. This demonstrates that TE is a valid and effective RF technique for improving MLIA performance. Comparing to the higher MLIA baseline "MRD-QE"[1], the TE run "TE-TWA" outperformed "MRD-QE" when the queries were TD and TDN. However, only the improvement obtained with TDN queries was statistically significant. With TE alone, the MLIA performance cannot outperform the corresponding monolingual baselines "Mono-Base". However, "TE-TWA" matches 93.61% of the effectiveness of "Mono-Base" under the TDN queries, which is close to the state of the art MLIA performance.

Table 3: Comparison of TE-TWA with several baselines (* indicates that the improvement is statistically significant)

Run ID		MAP (Perc. of Mono-Base)	Impr. over MRD-Base
T	Mono Base	0.4739(100%)	+42.06%*
	MRD Base	0.3336(70.39%)	-
	MRD QE-Combine	0.4415(93.16%)	+32.34%*
	TE-TWA	0.3992(84.24%)	+19.66%*
TD	Mono Base	0.5817(100%)	+36.84%*
	MRD Base	0.4251(73.08%)	-
	MRD QE-PostTrans	0.5080(87.33%)	+19.50%*
	TE-TWA	0.5340(91.80%)	+25.62%*
TDN	Mono Base	0.6215(100%)	+32.21%*
	MRD Base	0.4701(75.64%)	-
	MRD QE-PostTrans	0.5182(83.38%)	+10.23%*
	TE-TWA	0.5818(93.61%)	+23.76%*

Query length affects the performance of RF techniques. As shown in the last column of Table 3, QE obtained the highest MAP improvement over the corresponding baselines "MRD-Base" with short T queries, and the improvement decreased with longer TD and TDN queries. However, the effect is different for TE runs. "TE-TWA" performed better with longer TD and TDN queries, but less so to that of the short T queries. Therefore, it seems that query length has different effects on TE and QE. This is another motivation to combine these two RF methods.

The combination of TE and QE ("TEQE") achieved comparable results to the monolingual baseline "Mono-Base" for all three types of queries (see Table 4). In the case of T and TD queries, "TEQE" even exceeded "Mono-Base". Of course, these runs still cannot outperform the higher monolingual baseline "Mono-QE". "TEQE" run also significantly outperformed TE only run under T and TD queries, and it significantly outperformed QE only run under TD and TDN queries.

Table 4: Comparison of combining TE and QE (TEQE) to several baselines and to TE or QE alone (* indicates that the improvement is statistically significant)

TEQE	MAP			
	MAP	Perc. of Mono-Base	Impr. over TE-TWA	Impr. over MRD-QE
T	0.4748	100.19%	+18.94%*	+7.54%
TD	0.5905	101.51%	+10.58%*	+16.24%*
TDN	0.5972	96.09%	+2.65%	+15.25%*

Another interesting point is that "TEQE" showed stable performance with all three types of queries. Different to TE that works better with long queries, and QE that works well with short queries but is losing performance with long queries, the combined run performed consistently comparable to the monolingual run with all three types of queries. It seems that the combination helped to use one's advantages to overcome the limitations of the other. Therefore, we can conclude that it is beneficial and effective to combine TE with QE.

2.4 Topic 3: MT for OOV Terms

Both MT and MLIA have to face out-of-vocabulary (OOV) terms. OOV terms refer to the words whose translations are not available in the translation resources such as MRDs [13]. In MRD based MLIA, it is beneficial to use dedicated data mining and information extraction methods for obtaining high quality translations for named entities [14], which are the most important and most common type of OOV [15]. We believe that a well designed MT system would be helpful in resolving OOV terms. This is not only true in query translation, but also true in TE method in relevance feedback for MLIA.

The research questions therefore are: 1) in MT for query translation, would MT which has its own handling of OOV terms generate comparable result to an MRD based query translation method that has a dedicated OOV module; and 2) can MT supported TE method helps in resolving OOV terms?

Table 5: The comparison between MT method and MRD method with dedicated OOV module (* indicates that the improvement is statistically significant over "MRD Base", † indicates that the improvement is statistically significant over "MRD NE Enhance")

Run ID	T	TD	TDN
Mono Base	0.4739	0.5817	0.6215
MT Plain	0.4446†	0.5536	0.6170†
MRD Base	0.3336	0.4251	0.4701
MRD NE Enhance	0.3934*	0.5034*	0.5563*

In the studying of MT for query translation, we observed from Table 5 that a dedicated NE module for MRD based query

[1] "MRD-QE" refers to corresponding "MRD QE-Combine" or "MRD QE-PostTrans" based on query length.

translation indeed helped the system performance. The run "MRD NE Enhance", which has a dedicated NE translation module, significantly outperformed "MRD Base" in all three types of queries. However, its performance was still inferior to MT based method "MT Plain", and in the case of T and TDN queries, the differences were statistically significant. This indirectly indicates that MT has its capability of handling OOV terms.

When examining individual topics, we noticed that there were many named entities in the topic statements, and Google Translate system handled them well. There were still a few NEs that Google Translate system cannot handle. But they were also OOV terms for the MRD even with the NE module. So they did not make obvious difference between the MT methods and the MRD methods.

Table 6: OOV terms and their translations found by TWA (# after the No indicates that the translation is wrong)

No.	Topic ID	OOV Term	Translations found by TE-TWA
1	55087	Bingol	宾格尔省
2	55087	diyarbakir	迪亚巴克尔
3#	40007	Garner	还
4	55087	kandilli	坎迪利
5#	55029	karolinska	推动/科技
6	55179/55127	Kumba	昆巴
7	41025	montesinos	蒙特西诺斯
8	40037	morariu	莫拉留
9	41012	ouattara	瓦塔拉
10	55181	Qurei	库赖
11	41025	vladimiro	弗拉迪米罗

In the experiments of TE, we found that during the process of extracting translation information from the small parallel corpus built from relevant documents and their MT translations, the TE method "TE-TWA" can identify the translations for some OOV terms with the help of word alignment information. Table 6 shows the 11 OOV terms and their translations found by "TE-TWA" method. Almost all these terms are NEs of people names, locations, etc. This is consistent with the finding by [13]. Therefore, it becomes an important advantage for "TE-TWA" (thus for MT) to be able to identify translations for some OOV NEs. Of course, as shown in In the studying of MT for query translation, we observed from Table 5 that a dedicated NE module for MRD based query translation indeed helped the system performance. The run "MRD NE Enhance", which has a dedicated NE translation module, significantly outperformed "MRD Base" in all three types of queries. However, its performance was still inferior to MT based method "MT Plain", and in the case of T and TDN queries, the differences were statistically significant. This indirectly indicates that MT has its capability of handling OOV terms.

When examining individual topics, we noticed that there were many named entities in the topic statements, and Google Translate system handled them well. There were still a few NEs that Google Translate system cannot handle. But they were also OOV terms for the MRD even with the NE module. So they did not make obvious difference between the MT methods and the MRD methods.

Table 6, some of the found translations are wrong (such as the translations for Nos. 3 and 5), which were the results of word alignment errors. However, the fact that majority found translations are correct indicates that it is reasonable reliable to use MT based TE method for resolving OOV terms.

In summary, based on the results obtained from our studies on MT for query translation, we can see that because MT systems often has its way of handling OOV terms, it often can overcome most OOV problems faced in MLIA process. Unlike a dedicated OOV term resolution module is needed in MRD based method, MT for query translation does not critically need an OOV module. Still its performance is either comparable or significantly better than MRD enhanced by an OOV module.

Based on the results from our studies on MT for RF in MLIA, it also makes sense to resolve OOV terms by performing translation enhancement. The extracted translation information contains possible solutions to some OOV terms. Because of the high quality of MT outputs, the translations of OOV terms obtained this way are in high quality too.

3. DISCUSSIONS

Overall, if the MT outputs of the whole collection are available, DT-MLIA is probably the simplest and cost-effective method. Literature has shown that the performance of DT-MLIA is among the best of the various MT usages in MLIA. Of course, relevance feedback techniques can still be applied to further enhance the results. However, QT-MLIA is better than DT-MLIA to give more transparency to MLIA searches. In this case, MT is a very effective method for translating queries. Considering that commercial MT systems are easily accessible online for many major language pairs, this is a very simple way of building MLIA capabilities. Query translation based on MT can by itself achieve comparable results to monolingual baseline, which is at the state of the art MLIA performance. If relevance feedback techniques like QE are applied, the final results would be close to performing QE on monolingual searches.

In addition, with MT's relatively full coverage of terms in general domain, QT-MLIA using MT for query translation does not have to have a dedicated NE module to handling OOV terms. This further simplified the design of MLIA system without sacrificing the retrieval performance. If MT output is readily obtainable, it is useful for resolving OOV terms too, and TE can achieve that.

In MLIA, besides QE as a basic relevance feedback method, we also demonstrate that TE, which is performed on the identified relevant documents and their MT outputs, is certainly a valid relevance feedback method too. Both QE and TE can achieve results close to but not exceed to the monolingual plain search. However, the most interesting property of them is that the combination of them can generate a much more robust MLIA relevance feedback mechanism that is capable of handling queries with different lengths.

It is better that MT should be applied before relevance feedback. The different effects obtained from different QE methods in MT based query translation raise an interesting insight about combining MT based query translation and QE. Because pre-translation QE adds many expansion terms before MT can translate the queries, adding too many words that are not part of a sentence could hurt the quality of MT even though some of the words are truly relevant. This is true especially when the original queries already have many words to work on (such as in TD and TDN). So if QE is integrated with MT in query translation, it is better that QE is performed after MT.

At the same time, our studies also demonstrate that further integration between MT and MLIA is needed. For example, our current TE study still needs to apply word alignment between MT outputs and their original documents. This is expensive and should not be necessary. Statistical MT systems should be able to provide information similar to word alignment, or even better phrase level alignment, as part of its translation output. Mining from there is a much better approach for TE. Another example is that MT for query translation still gives out only one best translation. When the translations of the queries are simple and straightforward, this is not problematic. However, when the hypotheses of the translations of the query terms are at low confidence or quality, it is actually better for MT system to give out n-best translations. But current MT systems do not provide such output even though they are capable of doing that.

Finally, we acknowledge the limitations of our studies. Only one language pair and translation direction was used. All our experiments were performed on TDT collections, which are just one of the several available MLIA test collections. Many findings and insights could need further testing in other collections, and the document collections were quite comparable, which makes it more likely that pre-translation QE will be effective.

4. CONCLUSION

Machine Translation (MT) has been identified as a very important related technology for Multilingual Information Access (MLIA). In this paper, the primary goal is to test how MT can be used in MLIA process and what effects MT will bring to certain aspects of MLIA process. From multiple aspects of MLIA process, our studies demonstrate that MT can be applied at the many places in the whole MLIA process. Although MT in DT-MLIA certainly is a simple and cost-effective way of integrating MT into MLIA, many other important aspects of MLIA can certainly benefit from MT too.

Our future work on integrating MT in MLIA is in the following areas. First, applying word alignment on the returned documents and their machine translations is actually not the optimal approach. Current statistical MT systems can generate outputs with word level alignment information. We will explore the usage of such information in TE experiments. Second, improving translation relationship at word level certain helps, but modern phrase MT systems can give us many phrase level translation information. We can integrate the phrase translation generated by MT system into MLIA.

ACKNOWLEDGEMENT

This work is partially supported by National Science Foundation of USA under the agreement NSF/IIS 0704628, National Post-doctor Foundation of China under the agreement 20090451078, and Social Science Foundation of Wuhan University under the agreement 09ZZKY097.

REFERENCES

[1] Oard, D.W. and A.R. Diekema, *Cross-Language Information Retrieval*, in *Annual Review of Information Science and Technology*, B. Cronin, Editor. 1998, American Society for Information Science. p. 223-256.

[2] Darwish, K. and D.W. Oard. *Probabilistic Structured Query Methods*. in *Proceedings of the 26th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2003: ACM Press.

[3] Wang, J. and D.W. Oard. *Combining Bidirectional Translation and Synonymy for Cross-language Information Retrieval*. in *Proceedings of the ACM SIGIR 2006*. 2006.

[4] Resnik, P., D. Oard, and G. Levow. *Improved Cross-Language Retrieval using Backoff Translation*. in *First International Conference on Human Language Technologies*. 2001.

[5] Grishman, R., D. Westbrook, and A. Meyers. *NYU's English ACE 2005 System Description*. in *ACE 2005 Evaluation Workshop*. 2005.

[6] He, D. and D. Wu. *Translation enhancement: a new relevance feedback method for cross-language information retrieval*. in *Proceeding of the 17th ACM conference on Information and knowledge management*. 2008. Napa Valley, California, USA: ACM.

[7] Lavrenko, V. and W.B. Croft. *Relevance-based Language Models*. in *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2001: ACM Press.

[8] Kraaij, W., *TNO at CLEF-2001: Comparing Translation Resources*, in *Evaluation of Cross-Language Information Retrieval Systems: Second Workshop of the Cross-Language Evaluation Forum, CLEF 2001 Darmstadt, Germany, September 3–4, 2001 Revised Papers*, C. Peters, et al., Editors. 2001, Springer Berlin: Heidelberg. p. 141-162.

[9] Lavrenko, V., M. Choquette, and W.B. Croft, *Corss-Lingual Relevance Models.* Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2002: p. 175-182.

[10] Xu, J. and R. Weischedel, *A Probabilistic Approach to Term Translation for Cross-Lingual Retrieval;*, in *Language Modeling for Information Retrieval*, W.B. Croft and J. Lafferty, Editors. 2003.

[11] Oard, D.W. and J. Gonzalo. *The CLEF2001 Interactive Track*. in *the Cross-Language Evaluation Forum (CLEF) 2001* 2001.

[12] Och, F.J. and H. Ney, *A Systematic Comparison of Various Statistical Alignment Models.* Computational Linguistics, 2003. **29**(1): p. 19-51.

[13] Demner-Fushman, D. and D.W. Oard. *The Effect of Bilingual Term List Size on Dictionary-Based Cross-Language Information Retrieval*. in *36th Annual Hawaii International Conference on System Sciences (HICSS'03) - Track 4*. 2003. Hawaii.

[14] Ji, H., et al. *NYU-Fair Isaac-RWTH Chinese to English Entity Translation 07 System*. in *NIST ET 2007 PI/Evaluation Workshop*. 2007.

[15] Thomas Mandl and C. Womser-Hacker. *The Effect of Named Entities on Effectiveness in Cross-Language Information Retrieval Evaluation*. in *ACM SAC'05*. 2005. Santa Fe, NM.