Multilingual Plagiarism Detection

Menna Mostafa
University of Waterloo
200 University Avenue West
menna.mostafa@uwaterloo.ca

Lalit Agarwal
University of Waterloo
200 University Avenue West
lalit.agarwal@uwaterloo.ca

ABSTRACT

Cross lingual plagiarism detection has recently caught attention due to copy-right violations occurring in many fields such as education, journalism, scientific research, literature, screenplays, etc, where an author would translate an article in language L_1 into language L_2 and then either publish/submit it or change some of the sentences to suit his/her motivations. Therefore, the need for a robust method for cross lingual plagiarism detection arises. Most of the existing work on cross lingual plagiarism detection uses machine translation to translate the suspect document in L_1 into L_2 and then search for similar documents in L_2. However, we argue that this approach suffers from the following limitations: (1) machine translation does not capture different writing styles that differ from field to another, (2) online machine translation that allows anonymous users to suggest better translations which suffers from tampering and incorrect suggestions, (3) the limited ability to identify different types of plagiarisms, for example two articles describing an accident might be labeled as plagiarized although they originated from different sources. Therefore, we propose an approach that will attempt to remedy the above three limitations by using machine learning and crowd sourcing techniques.

Author Keywords

Plagiarism detection, cross lingual, machine learning, crowd-sourcing

INTRODUCTION

Plagiarism is the act of copying someone else's work without their consent and publishing it to the world as your own. The explosion in the world wide web helped plagiarism to go undetected for decades now, however with the recent advances in machine learning algorithms, more effort is put into automated tools for plagiarism detection. Plagiarism in monolingual setting is considered to be an easier task compared to multilingual setting, in monolingual setting models for document similarity such as TFiDF, cosine similarity, N-gram can be used on document or paragraph levels. However in multilingual setting, the problem gets complicated as different languages differ in alphabet, sentence structure, and the translation is hardly a word to word mapping. What further complicates the problem is that the definition for plagiarism depends on the context of the document, for example in journalism if the document is describing an event or accident, then you will probably find many similar sentences, however if the document is describing the point of view of the author about some political issue, then having many similar sentences is considered plagiarism. Moreover, in a field such as scientific research having any similar sentences is considered plagiarism.

While a word to word inspection is ideal for detecting plagiarism in monolingual documents, it is hardly applicable in multilingual documents due to the differences in the sentences structure across languages. Also authors might switch sentences order or use unusual synonyms to hide plagiarism. We argue that techniques that inspect a document as whole, would perform better since these techniques can capture the latent topics between training documents from different languages.

Therefore, in this work, we use collection based algorithms that train models using manually translated documents (parallel documents) covering various topics. The model infers the latent topics in all the documents, such that when two documents (did not appear in the training data) are projected to the model, a similarity measure between the two documents can be determined based on their similarity to the inferred latent topics.

We therefore, formalize the problem of multilingual plagiarism detection to be:

A document d in language L_1 can be modeled as a vector of weighted topics, each weighted topic $\langle t, wt\_t \rangle$ is a topic t and the weight wt_t of topic t in the document. Let T be the set of all topics inferred, a topic $t \in T$ is a vector of weighted words $\langle w, wt\_w \rangle$ where wt_w is the contributing weight of the word w in the topic t, the words can be in multiple languages. Now, given two documents d_1 in any arbitrary language L_1 and d_2 in another arbitrary language L_2, the problem is reduced to finding the weighted topics vectors v_1 and v_2 for d_1 and d_2 respectively, and then calculating the similarity measure $S_{\{d_1,d_2\}}$ between the two vectors (ex: cosine similarity), and furthermore based on this similarity measure, the technique should be able to determine whether the two documents are plagiarized or not.

In this work we chose *Latent Semantic Indexing* commonly referred to as *LSI* as our collection based model, our choice for LSI is based on its simplicity and its simple support for

multilingual indexing and the fact that its implementation is publicly available. We give a brief description for LSI in the framework section.

In order to use collection based algorithms to determine plagiarism, one needs to train a collection based model using parallel documents from the domain under inspection, and then find the similarity threshold that differentiates similar and plagiarized documents, however having a fixed threshold can results in erroneous verdicts due to context sensitivity, where the context of the documents plays an important role in judging if it is plagiarized or not. To explain this idea more, we show the definition of Plagiarism from Wikipedia[1]: "Plagiarism is presenting someone else's work including their language and ideas as your own, whether intentionally or inadvertently. "Plagiarism is not a concern where the content lacks creativity."

Hence, while two news articles in different languages describing the same incident (merely describing events that happened), might have a relatively high similarity measure due to the fact that they have many topics in common, they should not be considered plagiarized unless the similarity measure is significantly high. We therefore turn to looping humans in this cycle, by training the model to learn two things from the human judgments:

- Topics Sensitivity: which topics are more sensitive to plagiarism.
- Missing Topics: If the test documents include topics that were not available in the training data then we request that the participants report these topics so the model can be retrained with parallel data covering these topics.

Hence, we summarize the contributions in this paper:

- Formalizing a method for identifying plagiarism threshold according to document context.
- Benefiting from human judgment in evaluating the training data.

In the next section we give a brief summary about the state of the art work in multilingual plagiarism detection describing different methods and tools, in the framework section we describe our proposed framework and in the experiments section we describe the set of experiments we intend to do and describe the UI for the users studies.

RELATED WORK

In this section we give a brief overview about the state of the art approaches in multilingual plagiarism detection. In[6], the authors present a translation based approach where a document is split into paragraphs and each paragraph is translated into English and a similarity index is used to retrieve the highly similar documents in English. To decide if two documents are plagiarized, they created an artificial corpus based on EuroParl dataset and trained a classifier to classify documents they intensionally plagiarized. The disadvantages in this approach are obviously in the usage of machine translation is expensive and is not fully context aware as it may fail to select the appropriate synonyms for different contexts. In[1], the authors present a method to decide if two synonyms are plagiarized based on their relative location in the two documents, however the complexity of such approach is expensive as it has to study all the synonyms in the documents. In[4], the authors use google translate on the preprocessed suspicious documents to translate into English and then use a semantic net to identify the concepts in the suspicious documents and then a similarity measure is applied to identify plagiarized documents. Their method suffers from (1) high dependence on Google Translate that has the same Machine Translation limitations mentioned before, and (2)the semantic net might not cover all topics introduces in the documents. In[3], the authors present an approach similar to [4] where they label each sentence in the document in its original language with a concept, and then form a graph using all the concepts in the documents, and since they use BabelNet which can identify similar concepts in different languages using Wikipedia lingual links, similar concepts in different languages are unified and then a graph similarity technique is applied. However the authors do not provide performance evaluation for the technique. I In [8], the authors present KCCA algorithm for cross lingual information retrieval, the algorithm attempts to find latent topics across different languages by learning from parallel corpora. The algorithm performs well in terms of retrieval accuracy, however it suffers from (1) difficulty in adding new documents after training phase (2) cubic complexity. In [5], the authors use LSI in cross lingual retrieval of documents on the web, although LSI is designed for monolingual documents, the users overcome this challenge by concatenating the parallel documents and feeding the concatenated corpora to the algorithm.

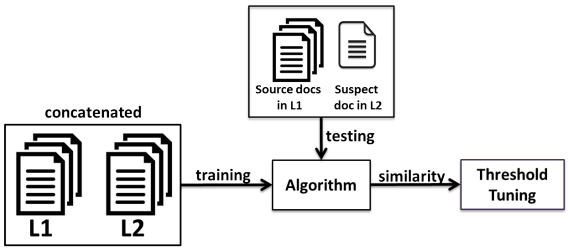

Figure 1. Overview of the framework.

FRAMEWORK

In this section we describe our framework, tools and data used. Figure 1 shows an overview of the our framework, we will explain each block in the next subsection.

Corpora

We chose to implement our model in English-Chinese languages for the following reasons:

- Chinese is widely spoken language, and it would be easy to find participants for the user study.
- Chinese and English use different alphabet which shows the powerfulness of the technique.

[1] http://en.wikipedia.org/wiki/Wikipedia:Plagiarism

- Chinese content on the web is powerful and it would be easy to find parallel data.

There exist two types of corpora, (1)parallel corpora and (2)comparable corpora. Parallel corpora is where both documents are manually translated by humans who are domain experts and aware of the context. Comparable corpora is where both documents discuss the same topic but are not exact translations and are probably not written by the same author, Wikipedia is a very good example for comparable corpora.

In order to have good coverage of many topics, we use a mix of parallel and comparable corpora. We used the following copora in training:

- UN multilingual corpora provided by the *United Nations* which includes manual translations for the sessions since the year 2000 through 2009. In total there were around $67,000$ documents.

- Wikipedia Comparable documents can be acquired from Wikipedia by processing the publicly available dumps and the multilingual links in the Chinese documents to the English documents, in total we were able to acquire around $150,000$ documents. However we did not apply any sampling on the categories, which would definitely enhance the performance.

For testing we used manually collected Chinese-English documents from bilingual Chinese news sites such as:

- Wikipedia Chinese featured documents [2].

- Xinhuanet bilingual zone. [3]

- Yeeyan manually translated articles. [4]

- peopledaily [5]

- chinadaily [6]

- NyTimes [7]

Latent Semantic Indexing

LSI in a technique devised in the late 1980's by Deerwester [2], LSI uses Singular Value Decomposition (SVD) on the document by term (word) matrix to decompose the matrix into the most influential orthogonal factors that form the matrix, these factors represent the latent topics in the training data. We used the publicly available implementation of LSI in gensim [8] library [7].

[2]http://en.wikipedia.org/wiki/Wikipedia:WikiProject_China/Featured_and_good_content

[3]http://www.xinhuanet.com/english/bilingual/news.htm

[4]http://www.yeeyan.org/

[5]http://en.people.cn/

[6]http://www.chinadaily.com.cn/

[7]http://cn.nytimes.com/ (Some articles in the Chinese version is a translation from the English version)

[8]http://radimrehurek.com/gensim/index.html

Preprocessing

The parallel training data in L_1 and L_2 is prepared for LSI by concatenating each parallel pair to form a new document containing all the terms that appeared in the two documents. The all the common and stop words [9] from both languages are removed from the concatenated document. We note that further customization and preprocessing should be applied according to the language, for example segmentation for Chinese [10], stemming for English [11] and Arabic.

Then the term-document frequency matrix is constructed from the processed parallel data. It is usually recommended to use a global weighting technique such as $tf - idf$ short for term frequencyinverse document frequency, which is used to give weights to each term in each document based on its importance in the all the training documents. The term-document frequency matrix A is constructed from n training documents D and total m terms, where the rows of the matrix are the documents and the columns are the terms, and each cell $a_{i,j} = tf(d_i, w_j) \times idf(w_j, D)$ where $tf(d_i, w_j)$ is the simple frequency of the appearance of the term w_j in document d_i and $idf(w_j, D) = \log(\frac{n}{|\{d \in D : w_j \in d\}|})$.

In order to further reduce the number of terms in the training data which we refer to as *dictionary*, we trim the words that appeared in less than 5 documents and the words that appeared in more than half the documents.

Training

Matrix A undergoes *Singular Value Decomposition* for a specific number of topics r to generate, the decomposition results in three matrices such that $A = T \times S \times D^T$ such that T is an $m \times r$ term-topic matrix, S is an $r \times r$ singular values matrix, and finally D is an $n \times r$ documents-topic matrix. The singular matrix represents the ranking for the r topics. Figure 2 show an example for a topic inferred by LSI.

Figure 2. Example of one of the topics inferred by LSI.

Projection

In order to compare a test document d_{L1} in language L_1 against set of documents D_{L2} in language L_2, we project the documents to the dictionary generated in the preprocessing stage, note that if a document has many words that are not present in the dictionary (bad coverage), then the algorithm will not have enough information to generate appropriate topics. Afterwards the documents should be represented in $tf-idf$ format by recalculating the weights for the training

[9]https://code.google.com/p/stop-words/

[10]http://code.google.com/p/smallseg/

[11]http://www.nltk.org/_modules/nltk/stem/porter.html

data with the test data, then the test document d_{L1} is multiplied by $T \times S^{-1}$ to result in a $1 \times r$ matrix V_{L1} representing the weights for each topic in r. The same is applied for each document in D_{L2}. Then V_{L1} is compared to each vector in D_{L2} topics using *Cosine-Similarity* which is measure of similarity between vectors, based on not magnitude and direction, since the ranking for the topics can take negative values, vectors can have opposite directions, hence cosine similarity can capture it.

Thresholding

The method described above works very good in cross lingual search, however in this work we aim to tune the threshold for topics according to the sensitivity of the context of the document to plagiarism. To do that we ran an experiment with participants, where we asked participants to mark document pairs in English and Chinese as either plagiarized or not. The complete description for the experiment is discussed in the experiments section. The documents presented to the users had different topic rankings and were a mix of plagiarized and non plagiarized documents, the idea is to correlate the cosine similarity measure generated by LSI to the judgment of users and either raise or lower the threshold for that topic.

Missing Topics

We noticed that the algorithm performed significantly bad for some topics that did not make enough appearance in the training data such as news about tennis, since the training data were from the UN (which is mainly politics, economics and human rights) and Wikipedia random sample, we could not ensure that sports and specifically tennis was included. Therefore we ran another experiment where we asked users to let us know about the topics that were present in one document and not present in the other. Using this information we would iteratively attempt to fetch documents that cover the topics reported by the users and add them to the training data and retrain the model.

CHALLENGES

We faced several challenges building this system, we list them briefly with the solutions implemented (if exists):

- Corpora Size: The size of the UN and Wikipedia corpora combined was too big to fit in the memory, therefore we had to sample documents from both. We were able to fit maximum of $35,000$ documents in the memory using a dual core machine (Intel i5-3320M 2.6Ghz) with 7.5 GB memory. In the future we plan on using a more powerful machine that can fit bigger corpora in memory.

- Number of Participant: unfortunately, due to the limited time and the restrictions for applying the experiments on random individuals, we were not able to get enough participants to collect enough data to implement the thresholding technique, and we did not want to exhaust our participants by asking them to evaluate large number of documents. However we plan in the future to use crowd-sourcing websites like Amazon Turk to overcome this problem.

- Due to time constraints we performed the missing topics experiment via email correspondence with the participants. Apparently the participants did not understand what was meant by missing topics and hence they provided very detailed topics relative to the document pair and we were unable to use the information provided by them to augment our training data, however we also intend to clarify the question and redo the experiment.

EXPERIMENTS

In the following experiments, we used a corpora consisting of a sample of $5,000$ documents from UN corpus, and $30,000$ documents from Wikipedia, we tested our manually collected dataset against other sized and figured that this was the best size. Figure 3 shows a graph for system evaluation against different sizes. We conducted two user studies, the first was for tuning the threshold for different topics, while the second was for identifying the missing topics. In the next subsections we describe each user study inputs and outputs.

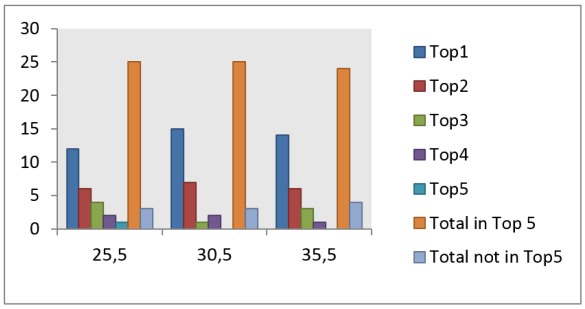

Figure 3. A chart showing testing results for different corpora sizes. On the X-axis the size of the Wikipedia corpora is the first number and the size of the UN corpora is the second number, on the Y-axis is the number of documents retrieved in a manual dataset of 28 documents.

User Study 1

The interface which we implemented for the first study was a simple interface which provided users features such as paragraph highlighting, display of top 10 words in both the languages. However, after analyzing the results from our first user study, we realized that users did not use some of the features which we thought would be useful to classify the documents as being similar to each other or not. Also a general feedback was to reduce the text length for both the documents to make the task easier and interesting. We modified the interface by incorporating some of these changes and conducted a second user study using our new interface.

Participant Sample

We were able to get 4 participants for both the user studies. Since most of the users who participated in our study were Chinese (Simplified) and English bilingual users, we purposely collected articles in Chinese and English for the study.

Dataset Collection

We specifically selected news articles for the study from various resources primarily because of two reasons: (1)we wanted to make the text documents interesting, so we manually picked news articles on recent topics such as Ebola, ISIS activities etc and (2) since we could not read Chinese, finding news articles in Chinese and English on the web which were

Users	doc 1	doc 2	doc 3	doc 4	doc 5
User1	Partially	yes	Partially	No	yes
User2	Partially	yes	No	No	No
User3	Partially	yes	Partially	No	Partially
User4	yes	yes	No	Partially	No

Table 1. Results from User Study 1, doc 1,2,5 are plagiarized, doc 3,4 are not. Yes indicates the users thinks the pair are plagiarized, No means the user does not think the pair is plagiarized while partially indicated that the user thinks some parts of the document might be plagiarized.

similar was relatively an easier task compared to other areas such as blog posts, assignment submissions etc. The test news articles were collected from popular Chinese news websites including peopledaily, chinadaily, xinhuanet and nytimes as mentioned before. Most of these websites published articles in both Chinese and English languages which made it easy to collect data for the study. We manually selected few news articles and while selecting the articles, the length of the article and the topic of the article was taken into account.

We selected similar articles on recent topics and made sure that the length of the document did not exceed 2-3 paragraphs in order to keep the user engaged during the entire task. We made extensive use of Google translate to understand the content of the news articles written in Chinese to confirm that it has similar content compared to the corresponding English article. Thus we were able to collect 7-8 document pairs in English and Chinese which were used for the user study.

The initial user interface which we used displayed a document pair in both English and Chinese along with the list of top 10 words in both the languages as shown in Figure 4. The interface also had a feature which allowed the users to highlight sentences/paragraphs which they found to be plagiarized. Out of the 7-8 document pairs which we had manually picked from the web, we selected only 5 pairs of documents which had a high similarity index according to our machine learning algorithm. We presented different types of document pairs to the user to analyze their responses to them. The following document types were selected for the study:

- 2 pairs on the same topics (Ebola, ISIS) from the same publishers (plagiarized).
- 2 pairs on the same topics (Online Privacy, IPhone Security) from different publishers (not plagiarized).
- 1 pair manually translated from the UN corpus (plagiarized).

Before the start of the study, each participant was given a basic idea about plagiarism and was told how to differentiate between plagiarized and similar pair of documents. After that, they were given a brief tutorial on how to use the interface. Each user was given 5 document pairs which we selected manually beforehand. We did not give them any time-limit to complete the task and they were allowed to complete it at their own pace. For each pair, the users were only required to read the two documents and report whether they found the document to be plagiarized, partially plagiarized or not plagiarized. Various user activities such as time taken to complete each task, plagiarism level selection of the user were logged in a local database at the backend.

Once the user was done with classifying all the document pairs, we asked them individually to give their feedback on the system using an online form. We asked users to rate their overall experience in using the system, their engagement level throughout the entire task and if they found the paragraph highlighting feature useful. Apart from this, we also asked users the number of document pairs they were willing to classify as plagiarized and not plagiarized in one-sitting. The main idea behind the on-line form was to measure the performance of the system based on some quantitative feedback from the users.

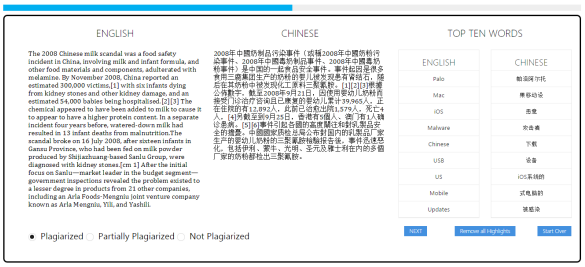

Figure 4. The interface for study 1.

Results

The users mostly gave a positive feedback to the interface with a few suggestions to improve the interface. Though one of the users suggested to reduce the length of the documents, other users generally had no problem with the document length. The users took around 2-3 minutes on an average to read through each document pair and mark whether they found it to be plagiarized or not. The table 1 shows the responses of the 4 users for each document pair.

It is evident that from the table that asking users if the document is plagiarized or partially plagiarized is a tough question to ask. For the first document pair which consisted of articles from the same publisher and hence had almost the same content, 75 % users marked it as partially plagiarized as opposed to plagiarized. Similarly for document 5 pair which consisted of exactly translated text, only one user classified it as being plagiarized. As suggested by one of the users, we realized from these results that displaying the similarity index of the document pair given by our algorithm alongside the documents would definitely assist the user in the classification task. However, we were not able to use the results in tuning the threshold since four participants is not enough to gain confidence in the results specially with the obvious differences in the responses.

The results we got from the online forms also gave us useful feedback about the interface. Users gave us an average rating of 3.75 on a 5-point scale for their overall engagement during the entire task. Users generally found the interface to be easy to use. We also asked users if they were willing to classify more than 5 document pairs in the future using the

on-line form. The user response was divided on this- while two users preferred to classify less than 5 document pairs, the other two users were willing to classify more documents if given an opportunity.

The two users who did not want to classify more than 5 document pairs wanted the document length to be a bit shorter. One of them wanted the topics to be more interesting and require very less domain/technical knowledge. The users did not extensively used the paragraph highlighting feature during the entire task but they found it to be very useful Apart from this, users also highlighted some issues and gave few suggestions to improve the interface. One of the issues highlighted by couple of users was that the Chinese document text needed reformatting and sentence restructuring, indicating that the document source might have used machine translation. Users also reported that it was hard to detect similar words in the text and wanted us to include a feature to highlight similar words when they hover the mouse on one of the words.

User Study 2

The purposed of this used study is two fold: (1) find topics that should be added to the training data to enhance the performance, (2) evaluate the system performance.

Interface

Based on the feedback we got from the users during the first study, we incorporated some changes to our interface. Instead of showing two documents in English and Chinese like in the previous study, we showed the users the three English documents which had the highest similarity index with the corresponding Chinese document. The main idea behind this was to ask users to rate the performance of our algorithm by looking at the top three English documents and their similarity indexes with the Chinese document. We displayed the similarity index alongside the documents as calculated by our algorithm to assist the user in classifying the document as being plagiarized or not. The similarity index ranged from -1 to 1 with 1 representing maximum similarity between the documents.

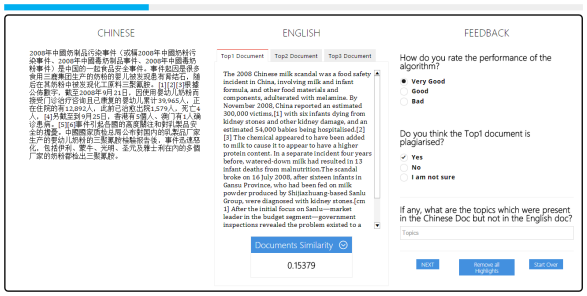

Figure 5. The interface for study 2.

Once the user was done reading all the documents, we asked them to rate the performance of our algorithm as very good, good, bad based on the content of these documents. This was done to get an idea if the similarity index given by our algorithm was correct compared to the content of documents.

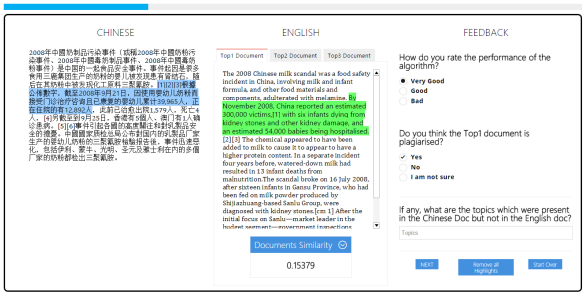

Figure 6. Paragraph highlighting in study 2.

We also specifically asked users if they found the top1 document i.e. the English document with the highest similarity index to be plagiarized or not. The users were asked to identify some of the topics which they thought were present in the Chinese document but not in the English document and vice versa. We wanted to identify certain topics to augment to the training and improve the performance of our algorithm.

We also wanted to include the feature to highlight similar words on mouse hover but due to the limited time could not implement it. Since we didn't have time to conduct the second user study face to face with the users and had to depend on email correspondence, we could not get feedback on our new interface from all the users. Only 1 of the four users tried our interface where we showed them two document sets and asked them to answer some of the questions mentioned above.

For the rest of three users, we sent out an email along with two document sets each consisting of a Chinese document and the top three matching English documents along with their similarity index (ranging between -1 and 1) and asked them to respond back with their responses to the above questions by email.

Results

Each user rated the performance of the algorithm for 2 document sets. We got a mixed response from the users towards the performance of our algorithm. While the performance was rated good four times and it was rated bad equal number of times. In most of the cases where the algorithm did not perform well according to the users, the similarity indexes of the documents were relatively lower which we think that the users didn't notice and down-voted the algorithm performance.

Also when the users were asked if they found the top1 document to be plagiarized, 4 out of 7 times, the user marked it to be plagiarized even when the top1 document in English and the Chinese document were from different sources but had similar content. We believe that differentiating between similar and plagiarized documents is still a difficult task for the users and more training should be given to the users before the start of the study to assist them with this task.

The topics provided by the users were very detailed and not useful in inferring what documents to add to the training data, we believe this problem is due to the lack of communication

with the users, we intend to resolve this issue by giving the users examples before they start using the system.

CONCLUSION

We finally conclude that while LSI works good in the field of multilingual retrieval, it needs some customization to be used in the field of multilingual plagiarism. We proposed looping humans in to benefit from the human judgment in differentiating between similar and plagiarized topics. We plan on taking this work a step forward and using crowd-sourcing and larger machines to solve the problems discussed in the challenges section. We find our solution to be original and useful in building a multilingual plagiarism detection that can be robust in detecting plagiarism and not subject to tampering and weak translations as most of the other suggested systems.

REFERENCES

1. Ceska, Z., Toman, M., and Jezek, K. Multilingual plagiarism detection. *Artificial Intelligence: Methodology, Systems, ...* (2008), 83–92.
2. Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. Indexing by latent semantic analysis. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE 41*, 6 (1990), 391–407.
3. Franco-Salvador, M., Gupta, P., and Rosso, P. Cross-language plagiarism detection using a multilingual semantic network. *Advances in Information Retrieval*, 269180 (2013), 710–713.
4. Kent, C. K., and Salim, N. Web Based Cross Language Plagiarism Detection. *2010 Second International Conference on Computational Intelligence, Modelling and Simulation* (Sept. 2010), 199–204.
5. Lee, C.-H., Yang, H.-C., and Ma, S.-M. A novel multilingual text categorization system using latent semantic indexing. In *Innovative Computing, Information and Control, 2006. ICICIC'06. First International Conference on*, vol. 2, IEEE (2006), 503–506.
6. Pereira, R. C., Moreira, V. P., and Galante, R. A New Approach for Cross-Language Plagiarism Analysis. 15–26.
7. Řehůřek, R., and Sojka, P. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, ELRA (Valletta, Malta, May 2010), 45–50. `http://is.muni.cz/publication/884893/en`.
8. Vinokourov, A. Inferring a semantic representation of text via cross-language correlation analysis. *Advances in neural ...* (2002).