Chapter 6

Conclusions

This chapter presents a brief summary of the work in this thesis. Also, some conclusions derived from the experiment results and some limitations encountered in this work are addressed. In the last section, some suggestions for future work are listed.

6.1   Summary

The primary objective of this work is to automatically map English concepts to equivalent Indonesian concepts. Bilingual concept mapping task is introduced so as to achieve this objective. Specifically for bilingual concept mapping, this work attempts to map English concepts derived from Princeton WordNet, which serves as English WordNet, to Indonesian concepts derived from KBBI, which serves as Indonesian WordNet. Another task, namely bilingual term mapping, is introduced to map English terms to their Indonesian analogues. It was intended to verify the underlying process behind bilingual concept mapping and to find optimal variable configuration to carry out the concept mapping. A model employing LSA has been presented to compute these tasks.

Some experiments were carried out using several variable configurations. For bilingual term mapping, the variables used are collection size, rank approximation, removal of *stopword*, frequency weighting, source term selection, and target term mapping selection. For bilingual concept mapping, only the largest collection size of parallel corpus 1 was used. The removal of *stopwords*, rank approximation setup, and weighting schemes are identical to bilingual term mapping. Additionally, the source term selection and the target term mapping selection are replaced with source concept selection and target concept mapping selection.

90

Essentially, a parallel corpus is required to apply LSA for the purposes of this work. Two available resources used as parallel corpora are article pairs from ANTARA and verse pairs from an English Bible and an Indonesian Bible. For bilingual term mapping experiments, each parallel corpus was employed separately. For bilingual concept mapping, on the other hand, only the first parallel corpus was employed.

6.2 Conclusions

According to the research objectives listed in Section 1.3, this work attempts to provide Indonesian *synsets* automatically, which is a basic requirement for building an Indonesian WordNet. This objective can be accomplished by mapping English concepts listed in Princeton WordNet to Indonesian concepts listed in KBBI.

This work shows that LSA can be used to approximate the bilingual concept mappings between English and Indonesian concepts. Although LSA is not yet able to attain levels comparable to human judgements, it is generally better than random baselines (See Table 5.12). On the other hand, LSA performance in bilingual term mapping, that is to map English terms to their Indonesian analogues, is unsatisfactory. It is believed that the failure is due to the inappropriate metric used to evaluate the mapping results.

The following subsections express some detailed conclusions which can be drawn from bilingual term and concept mapping experiments. Moreover, some hypotheses for several issues in this work are listed.

6.2.1 Conclusions of Bilingual Term Mapping Experiments

Some conclusions derived from the bilingual term mapping experiment results are:

- Bilingual term mapping experiments for terms which appear frequently in a collection yield better results than terms which appear rarely. The frequent terms have better semantic context provided by the collection, thus enabling LSA to map the terms to the correct translations.

- Applying a weighting scheme to term-document matrices can improve the performance of both frequency baseline and LSA. TF-IDF weighting gives positive effects towards frequency baseline. Unfortunately, it gives the opposite

effects to LSA. On the other hand, Log Entropy weighting improves the mapping results for both frequency baseline and LSA.

- Collection size influences LSA performance greatly. As the size of collection increases, LSA performance also increases. According to the experiment results, LSA using P_{1000} of parallel corpus 1 and P_{3500} of parallel corpus 2 yields the highest precision and recall values compared to the smaller collections.

- In general, frequency baseline, which employs full rank term-document matrix, performs bilingual term mapping better than LSA, which employs k-rank approximation matrix. It is speculated that the rank approximation removes important details from the original matrix, thus information about the context of the terms becomes blurred.

- Although removal of *stopwords* is commonly used to improve Information Retrieval systems, the results of applying this technique to bilingual term mapping were likely to be worse. It is speculated that because of the small size of collections, in coincidence, *stopwords,* which consistently appear in a specific domain, may carry some semantic information about the domain. The precise explanation of this issue is unknown yet and is suggested for further research.

- Mapping selection can be of great importance to get the optimal precision and recall. Precision is the number of mapping results, which match bilingual dictionary pairs, divided by the number of Indonesian terms taken for an English term. Thus, as the number of translation pairs selected increases, the precision decreases. On the other hand, recall is the number of translation pairs, which match the bilingual dictionary pairs, divided by the number of pairs given by the bilingual dictionary. As the number of translation pairs selected increases, the possibility to find more pairs matching the pairs in bilingual dictionary increases. Thus, the recall value increases as well.

- Experiment results using bible verses as the fine-grained documents and the chapters as the coarser-grained documents show that LSA performance using the verses is indeed better than using the chapters. Thus, it confidently confirms the intuition that LSA performance can be improved by using finer-grained

segmentation of the parallel texts which enables it to capture more specific context of term-usage.

6.2.2 Conclusions of Bilingual Concept Mapping Experiments

Some conclusions derived from bilingual concept mapping experiment results are:

- The textual context sets for both English and Indonesian concepts are very important for LSA to yield correct mappings. Large textual context sets provide adequate contexts for LSA to understand the English concepts to be mapped and subsequently enable it to choose the correct Indonesian equivalent concepts. Conversely, limited textual context sets which are sparsely filled with terms that cannot convey the concepts are believed to be the reason why LSA is unable to identify correct mappings.

- Although the results are far from conclusive, it indicates that LSA is able to show some measure of semantic information provided by the textual context sets. Thus, LSA can perform better than random baseline, but is not yet able to attain level comparable to human judgements.

- In general, some failure of LSA is believed to be due to lack of context provided by the collection size. Thus, it suggests further work of experimenting with larger corpora. Indeed, similar LSA approaches are commonly trained on collections of texts numbering in the tens of thousands of articles, e.g. (Rehder, Littman, Dumais, & Landauer, 1997).

- Generally, LSA using Top 1 mapping selection yields higher level of agreement than other mapping selections, i.e. Average threshold, MinMax 10%, 25%, and 50% thresholds, and Top 3, and 5. Intuitively, since LSA using Top 1 mapping selection only selects one Indonesian concept, the level of agreement with human judges for not choosing the other concepts as the correct mapping should be high.

- LSA with TF-IDF weighting for bilingual concept mapping is likely to yield the highest average of level of agreement than LSA without weighting and LSA with Log Entropy weighting. The level of agreement average of LSA with Log

Entropy weighting is slightly lower than TF-IDF. But, it is still considerably higher than no weighting.

- For Average threshold and MinMax 10%, LSA performs optimally with the original term-document matrix without any weighting. On the other hand, for MinMax 25% and Top1, LSA performs better using term-document matrix with TF-IDF weighting. For MinMax 50%, Top 3, and Top 5, LSA performs better using term-document matrix with Log Entropy weighting.

6.3   Limitations

Some limitations encountered in this works are:

- **Limitation of hardware**

  SVD computation requires huge memory. As the term-document matrix size increases, the requirement of the memory also increases. In actual fact, the available memory size in this work is limited up to 1G. SVD computation for large corpora is usually run on supercomputers.

- **Limitation of resources**

  The availability of English-Indonesian parallel corpora is very limited. The ANTARA article pairs used in this work were paired semi automatically. Thus, the accuracy of the translation and alignment is rather doubtful. The corpora might be more suitably considered as comparable rather than parallel.

  English and Indonesian Bibles should serve as better resources with respect to finer alignment and translation. However, the content is of a very specific domain and homogenous, i.e. does not cover various domains.

  It is necessary to apply LSA on large corpora, but the number of available parallel documents is very limited. Given larger corpora, LSA should acquire much better analysis.

  Another limitation was found during the development of the English-Indonesian bilingual dictionary. Not many resources were available and their qualities are rather doubtful. Nevertheless, the bilingual dictionary was built by combining

these resources. Many entries of the dictionary use abbreviations, thus it should affect the precision and recall values in evaluating bilingual term mapping experiments.

As mentioned above, textual context set is very important for LSA to do concept mappings correctly. In reality, some entries of Princeton WordNet and KBBI even do not provide adequate ideal textual context. For example, many entries in KBBI do not have definition.

The list of common based concepts may contain concepts which are not truly common. For example, the English concept *"a successful stroke in an athletic contest (especially in baseball)"* conveyed by the term *"hit"* is not a common concept in Indonesian. Moreover, this concept hardly ever appears in Indonesian news articles.

- **Limitation of time**

As the period of time for doing this work is very short, the results are quite limited. A lot of questions remain to be answered in further research. Some suggestions are given in the following section.

6.4 Future Work

Several suggestions for future work are:

- **Using larger corpora with a wide variety of domains**

In this work, the collection size is very limited due to the limitation of hardware and the availability of parallel corpora, whereas the analysis of LSA greatly depends on the context provided by the collection size. Hence, it is recommended experimenting bilingual term and concept mapping with larger corpora. It is also recommended to employ corpora with a wide variety of domains to provide better distinction of the contexts.

- **Using corpora with finer granularity of alignment**

Bilingual term mapping experiments on varying text alignment granularity has shown that finer granularity of text alignment improves the accuracy of the

mappings. It is recommended to apply bilingual concept mapping using corpora with finer granularity of alignment.

- **Using larger bilingual dictionary with better quality**

    Specifically for bilingual term mapping, precision and recall value greatly depends on the bilingual dictionary. Thus, it is suggested to perform experiments using larger bilingual dictionary with better quality.

- **Using larger memory**

    As mentioned above, SVD computation requires very large memory size. The memory size determines the number of collection size which can be computed. In this work, a very limited collection size was experimented with due to this limitation of memory size. Thus, it is recommended using larger memory for future work.

- **Using another weighting scheme**

    In this work, only two weighting scheme were experimented with, i.e. TF-IDF and Log Entropy. Other weighting schemes are suggested to be explored in further work.

- **Observing the optimal rank approximation**

    Determining optimal rank approximation to acquire the best LSA performance is still an open research question. The bilingual concept mapping experiment results using top 3 suggest that the level of agreement between human judges and LSA with 10% rank approximation yields better results than 25% and 50%. However, it is not certain yet and is recommended for future work to observe the optimal rank approximation deeply.

    The optimal rank approximation can be determined by observing the singular values matrix of a bilingual term-document matrix. The rank approximation can be determined according to the number of singular values, which are nonnull. Since the singular values are likely to fall down after several values, the rank can be approximated according to the number of singular values before they collapse. Furthermore, the complexity of SVD computation can be reduced using these approaches.

- **Observing why excluding *stopwords* from collection decrease LSA performance**

  As described above, the results of applying this technique to bilingual term mapping are likely to be worse. The proper explanation of this issue is unknown yet and is suggested for further research.