sltu

by Budi Irmawati

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 81 (2016) 214 – 220

Procedia Computer Science

5th Workshop on Spoken Language Technology for Under-resourced Languages, SLTU 2016, 9-12 May 2016, Yogyakarta, Indonesia

Exploiting Syntactic Similarities for Preposition Error Corrections on Indonesian Sentences Written by Second Language Learner

Budi Irmawati[a,b,*], Hiroyuki Shindo[a], Yuji Matsumoto[a]

[a]Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan
[b]University of Mataram, Jalan Majapahit 62, Mataram 83125, Indonesia

Abstract

We propose a method to artificially generate training data to correct preposition errors in Indonesian sentences written by second language learners. Basically, we injected large size of native sentences with preposition errors learned from learners' sentences. Our method copies a preposition error from a learner sentence to a native sentence by firstly calculating a syntactic similarity score between the native sentence and the learners' sentence. Then, it chooses the preposition error from the learner sentence that has the highest syntactic similarity score to the native sentence, to replace the original preposition in the native sentence.

Experimental results show that the preposition error correction model trained on the artificial data resulted from our method outperforms the correction model trained on the similar size of native data.

Peer-review under responsibility of the Organizing Committee of SLTU 2016

Keywords: Artificial training data; Indonesian language; preposition error correction; syntactic similarities; under-resource language

1. Introduction

Grammatical error corrections on sentences written by second language (L2) learners for rich-resource languages such as English get many attentions recently[1]. However, research on this area for Indonesian, such as spelling checker[2,3], are very few. Researchers who work on this language are still struggling to develop language resources like name entity recognizers (NER), annotated corpora, morphological analyzers, or parsers[4,5,6,7,8] to explore more sophisticated methods that have been applied to other languages. The limitations are not only on the existence of mature language tools but unfortunately, also on the availability of the annotated language data. L2 data, for instance, need to be annotated with error tags as well as the linguistic features such as part-of-speech (PoS) and/or morphological information.

We initiated our work of grammatical error correction on prepositions, a closed class words, because the number of preposition choices is limited, considering that our learner corpus is small. Therefore, we utilized the corpus as the test data, which means we need to find a way to generate training data. For this purpose, we worked on native

* Corresponding author. Tel.: +62-370-636126 ; fax: +62-370-636523.
E-mail address: budi-i@is.naist.jp, yzakodek@yahoo.com

Peer-review under responsibility of the Organizing Committee of SLTU 2016
doi:10.1016/j.procs.2016.04.052

sentences as training data, but such a model cannot learn to correct errors made by learners. The model only learns corrections from the local context obtained from the native sentences.

To deal with those limitations, we propose a method to generate artificial training data for Indonesian, an under-resource language. We followed other works in generating artificial data by replacing a correct preposition with an incorrect preposition learned from learners' sentences. Our method copies errors made by learners to a native sentence based on the similarity score between the native sentence and the learners' sentence. We explain the difference of our method compared to other works in Section 2. Our method employs dependency-based word embeddings proposed by Levy and Goldberg[9] to calculate the similarity score between a native sentence and a learners' sentence as explained in Section 3. Then we show the performance of our proposed method among some baselines in Section 6.

2. Related works

Many error correction tasks employ supervised learning[10,11,12] that needs a large scale training data. Using native sentences as training data is an alternative to get huge training data for error corrections[10,13], but Rozovskaya and Roth[12] showed that introducing errors sophisticatedly on native sentences gives better corrections. They used information about error distribution and learners' first language (L1) to select appropriate sentences to be injected with an error. Unfortunately, our data do not have those information used in their work.

On the other hand, even though the results were not satisfied, Sidorov et al.[14] used syntactic information available in the training data provided by the organizer in the CoNLL Shared Task 2013 to build a rule-based system to correct five error types including prepositions. Then, Hernandes and Calvo[15] proposed an error correction model using syntactic n-gram in the CoNLL Shared Task 2014.

Levy and Goldberg[9] proposed word embeddings using dependency-based contexts as features. Their qualitative evaluation showed that the dependency-based embeddings bring about more functional similarity than the original skip-gram embeddings[16], which is domain similarity.

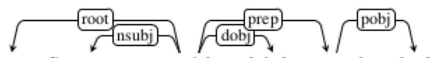
Inspired by the methods proposed by Sidorov et al.[14] and Hernandes and Calvo[15], we exploited syntactic information from a large native corpus. However, different from their work, we used the syntactic information to calculate a syntactic similarity score between a native sentence and the learners' sentences, so we could copy a preposition error in the appropriate learners' sentence to the native sentence to generate an artificial sentence.

3. Artificial training data generation

Our method includes two components: Training word embeddings using a large number of native sentences and copying preposition errors from learners' sentences to native sentences.

3.1. Training Dependency-based Word Embeddings

To generate artificial training data, we need large-scale native sentences that were automatically annotated with dependency relations. We extracted contexts based on the syntactic relations between a target word w and its head h and its modifiers $m_1, \ldots, m_k$: $(m_1, lbl_1), \ldots, (m_k, lbl_k), (h, lbl_h^{-1})$ where lbl is the type of the dependency relation between w and a modifier while -1 is used to mark the inverse-relation of w and h[9]. For the sentence given in Example (1), the features for ***provide*** are $(Someone, nsubj)$, $(drinks, dobj)$, $(desk, prep\_on)$, $(\text{ROOT}, root^{-1})$. Relations that include a preposition are 'collapsed dependency'[17]. We trained word embeddings on those features using the dependency-based embeddings proposed by Levy and Goldberg[9].

(1) 

ROOT Someone provides drinks *on* the desk

3.2. Injecting Artificial Data

As our goal is to correct preposition errors, we only need information about the head and the object of the target preposition. Given a native sentence as a *source* sentence, our proposed method finds a learners' sentence that has the highest similarity score to the native sentence following Equation (1)

$$\hat{h}_l, \hat{o}_l = \underset{h_l, o_l}{\arg\max}\ cos(\boldsymbol{v}_{h_n}, \boldsymbol{v}_{h_l}) \times cos(\boldsymbol{v}_{o_n}, \boldsymbol{v}_{o_l}) \tag{1}$$

where p is the preposition, h is its head, and o is its object. $\boldsymbol{v}_h$ and $\boldsymbol{v}_o$ are the vector representations of h and o. The subscripts n and l refer to the native and the learner sentences.

The objective function is to find a learners' sentence that contains h_l and o_l where the cosine similarity between the head h_n in the native sentence and the head h_l in the learner sentence and the cosine similarity between the object o_n in the native sentence and the object o_l in the learner sentence are maximum. For each native sentence, we only took the first 100 words that were similar to h_n and the first 100 words that were similar to o_n. So, to cover a native sentence where its similar h_n match to h_l in some learners' sentences, but its similar o_n do not match to o_l in any of those learners' sentences, we defined an *exception rule*:

"If the method cannot find a learner sentence whose both h_l and o_l are highly similar to o_l and o_n respectively, check if there is a learner sentence whose h_l is highly similar to h_n while its o_l has the same PoS as o_n, and assign $cos(\boldsymbol{v}_{o_n}, \boldsymbol{v}_{o_l})$ to 0.1. Otherwise, ignore the native sentence."

For easier understanding, we give an illustration by an English sentence. In Example (2), we have a native sentence (2a) as the *source* sentence. We assume that a learners' sentence (2b) has the highest similarity to the sentence (2a) among other learners' sentences since its $cos(\boldsymbol{v}_{provides}, \boldsymbol{v}_{put}) \times cos(\boldsymbol{v}_{desk}, \boldsymbol{v}_{table})$ is maximized.

(2) a. Someone **provides** drinks *on* the **desk** (prep: provides → on; pobj: on → desk)
h_n ... p_n ... o_n

b. We **put** a book **in*/*on* the **table** (prep: put → in/on; pobj: in/on → table)
h_l ... $\tilde{p}_l/p_l$ ... o_l

As preposition $\tilde{p}_l$ in the learners' sentence (2b) is corrected to p_l, our method replaces the correct preposition p_n in the *source* sentence (2a) to $\tilde{p}_l$. The native sentence is then converted to a new artificial sentence:

"*Someone provides drinks *in the desk*"

4. Error Correction Model

To correct the preposition errors, we trained a probabilistic classifier, *Naïve Bayes* on the artificial error sentences obtained by our proposed method. The classifier used context-word features in ± 2 window, bi-gram, tri-gram, and PoS n-gram. It also employed the head and the object of the targeted preposition, and the PoS of the head and the object of the targeted prepositions.

We employed a *one-vs-all* approach[18] to perform a multi-class classification. For M target prepositions, we assigned feature vectors for a classifier p as positive examples and feature vectors of $M - 1$ classifiers as negative examples. Then, we ranked the candidate corrections given by each classifier based on the confidence score obtained from the classifiers.

5. Experiment

5.1. Language Resources

We use following language resources in our experiment:

1. **Learner data**. They are taken from the Indonesian part of Lang-8 data[1] [19] crawled from lang-8 web service[2]. It contains 6,488 pairs of learners' sentences and 77,201 tokens. We automatically aligned a learners' sentence with

[1] http://cl.naist.jp/nldata/lang-8
[2] http://lang-8.com

each corrected sentence to tag the error types and error positions[20]. Then we took the learners' sentences that contain preposition errors and asked two native speakers who hold a master degree in social science to check and manually re-annotate the incorrect alignments. It consists of 382 sentences containing at least one preposition errors.

2. **Native data**. They are newspaper data taken from the Indonesian part of Leipzig corpora[21], that contains 1M sentences.
3. **Confusion Set**. Confusion set is the set of pairs between incorrect preposition $\tilde{p}$ with a correct preposition p extracted from the learner data.
4. **Morphind**[3]. A morphological analysis system for Indonesian[6] that covers affixations and clitics.
5. **Dependency parser**. We built a dependency parser by training the Minimum Spanning Tree (MST) parser[22] on 1,032 sentences annotated manually with labeled dependency relations.

The learner data and the native data were PoS tagged using Morphind and parsed using the dependency parser.

5.2. Evaluation Metrics

To evaluate our model, we followed the evaluation metrics proposed by Dahlmeier and Ng[23], defined as:

$$precision(P) = \frac{\#correct\_prepositions}{\#correction\_given\_by\_the\_model}$$

$$recall(R) = \frac{\#correct\_prepositions}{\#preposition\_errors}$$

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall}$$

5.3. Experiment Setting

We used all sentences from the native data set as the training data to calculate the word embeddings, but we only used 490K native sentences from the same set as the *source* sentences for artificial error data. We employed all learner data in ten-fold cross-validation. Each time, one fold was used as the test data for the error correction model and the nine folds were used to generate the artificial error data. With the limitations of our learners' sentences, we only worked on 13 preposition errors whose error frequency is more than five in the learners' sentences.

In these experiments, we wanted to investigate whether the word similarity calculated from syntactic features works on the preposition error correction (for replacement errors) task. The preposition errors may be classified to semantic errors, politeness, or functional errors. However, we could not differentiate those error types due to the size of learner data. Then to find out which syntactic features that work well, we discriminated the experiments into: (1) using similarity scores of both the preposition head and the preposition object if they exist (*depWE-HeadObj*) and applying the *exception rule* when the preposition object does not have similar words in the learners' sentences; and (2) only using similarity score of the preposition head (*depWE-Head*).

As baseline models, we used two training data for the error correction model: (1) 150K native sentences taken from original 1M sentences (*Native*) and (2) native sentences (the same data as *Native*) that were injected by preposition errors randomly (*CSRnd*), but we restricted the injected prepositions based on the confusion set explained in Subsection 5.1.

6. Results and Discussions

Figure 1 presents the F_1 score comparison of the error correction results on different size of training data. To have a proper evaluation, we iteratively sampled N instances (at least 100 times) without replacement and calculated the

[3] `http://septinalarasati.com/work/morphind/`

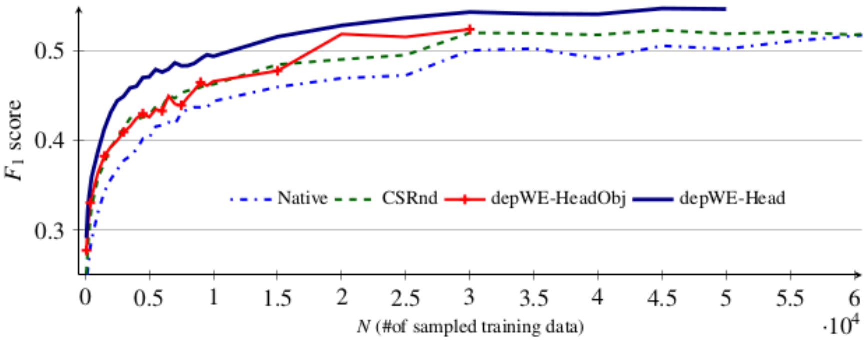


Fig. 1: **The comparison results of preposition error correction model trained on different size of sampled training data.** We iteratively sampled N instances without replacement from each training data. The native baseline is drawn in *blue dash-dotted line* and the *CSRnd* baseline is drawn in a *green dashed line*. Our proposed method is competitive compared to baselines, even though it produces smaller training data. It also shows that the dependency-based error injection using only head information *depWE-Head* outperforms other methods in similar size of training data.

average F_1 score of each training data. We present 60K samples since the performances are saturated on that number of samples. The figure shows that the models trained on the three artificial data outperform the models trained on the native data where their size is similar. It indicates that error correction task does get benefit from the training data that contain errors as explained in Rozovskaya and Roth[12] where the number of samples is less than 60K. On a larger scale training data, *Native* works well and catches *CSRnd* on 60K training data. Therefore, enlarging the learner data is important as larger training data perform better.

Figure 1 also shows that the artificial training data that employs both the head and the object (*depWE-HeadObj*) works slightly better than the random injection one (*CSRnd*) when the number of training data is small. The reason is that *CSRnd* needs larger training data to cover more error patterns. They may also contain noises because some artificial prepositions have never been made by learners.

We also notice that similarity score calculated only from the preposition head (*depWE-Head*) works the best in this case. *depWE-Head* generated more artificial data than *depWE-HeadObj*. The reason is that combining two dependency features to calculate the similarity score is difficult because the object in the learners' sentences that is similar to the object in the native sentence is not always available. On the other hand, the *source* (native) sentences have wider word variations and contain more complex inflections than the inflections used in the learners' sentences. These variations and complexities make difficult to find similar words in the learners' sentences.

The upper part of **Table 1** shows that our proposed methods (written in italic) obtained a better score than the baselines when the training data size is the same. It means that we obtained a better quality of artificial data than *CSRnd*. The bottom part of the table lists the training data in larger sizes than the upper part. We list certain size of training data whose F_1 score is similar to *depWE-Head*. It indicates that larger size of training data improve the precision, but *depWE-Head** is promising because it reaches the highest recall with less training data. To obtain a similar F_1 score to *depWE-Head**, *CSRnd* needs 50K more training samples while *Native* needs 100K more training samples. This part also shows that *CSRnd* is better than *Native* when training data is small, but they reach the same performance when *Native* becomes large as shown in Figure 1.

We outline an error analysis of our proposed method that may lead us to potential improvements below. **Single gold answer:** Our data only have one answer for each preposition error. Acceptable answers by the context are considered as incorrect if they are different from the gold answer. **Limited context:** Some prepositions can be decided based only on the head (verb or noun) such as 'focus on' while some prepositions rely on both the head and the object and they are chosen based on the semantic. Our model could not predict well these kinds of errors such as 'go to' or 'go with'. Our model also could not decide the proper preposition where the choice is decided based on the previous context such as 'come from', 'come to', or 'come in' even though they have the same object. **Out-of-domain:** Figure 1 shows that *depWE-HeadObj* generated less data than *depWE-Head*. It indicates that the number of similar objects between the native data and the learner data is limited. Moreover, the vocabularies chosen in native sentences are

Table 1: **The comparison results of the error correction model on a certain number of training data.** The upper part of the table shows that *depWE-Head* outperforms other artificial error injection *CSRnd* and *depWE-HeadObj* on the same size of training data. The bottom part shows that *depWE-Head* is still competitive or even better even though it trained on smaller training data.

Training data	**Size**	**Precision**	**Recall**	F_1 **score**
Native	25K	0.6592	0.3684	0.4727
CSRnd	25K	0.6783	0.3902	0.4954
depWE-HeadObj	25K	0.6722	0.4181	0.5155
depWE-Head	25K	**0.7059**	**0.4331**	**0.5368**
Native	125K	**0.7607**	0.4147	0.5367
Native	75K	0.7322	0.3979	0.5157
CSRnd	75K	0.7267	0.4181	0.5308
*depWE-Head**	**25K**	0.7059	**0.4331**	**0.5368**

The proposed methods are written in italic.
The bottom part lists the training data in larger size.

much more formal compared to the learners' choices, which are a bit simple. **Parsing accuracy:** The parser accuracy is only 81.2%. It might incorrectly assign the head or the object of the preposition, so our method chose inappropriate preposition errors. The parsing mistake might also generate inappropriate features for the error correction model. This analysis directs us to work further on improving the word similarity, especially on the preposition object, which is more challenging because the preposition object in the native sentences may not be available in the learners' sentences. We plan to adopt name entity recognizer (NER) and/or domain adaptation.

7. Conclusion

We proposed a method to generate artificial learner data for an under-resource language where the original learner data are insufficient to be used as training data. Our method includes two components: training word embeddings using a large size native sentences and copying preposition errors from learners' sentences to native sentences. Then, to correct the preposition errors in learners' sentences, we trained a probabilistic classifier on the artificial error data and applied to the learners' sentences.

We only produced a few artificial training data from a bunch of native sentences because the word variations in the learners' sentences are much lesser than those in the native sentences. However, as an initial work, even though our results are lower than the state of the art of the preposition error correction of English, this method can be applied to other languages when the language resources are limited or are not publicly available.

For future direction, we will utilize name entity recognition (NER) features to improve the sentence similarity when the preposition objects are employed. We will also apply this method to re-rank the preposition candidates given by an error correction model.

Acknowledgements

We acknowledge the anonymous reviewers for their constructive comments. This study is supported in part by the DGHE, Republic of Indonesia under BPPLN Scholarship Batch 7 fiscal year 2012-2015.

References

1. Leacock, C., Chodorow, M., Gamon, M., Tetreault, J.. *Automated Grammatical Error Detection for Language Learners*. Morgan and Claypool Publishers; 2014.
2. Aqsath, R., Kamayani, M., Reinanda, R., Simbolon, S., Soleh, M., Purwarianti, A.. Application of document spelling checker for bahasa indonesia. In: *Advanced Computer Science and Information System (ICACSIS)*. 2011, p. 249–252.
3. Soleh, M., Purwarianti, A.. A non word error spell checker for indonesian using morphologically analyzer and hmm. In: *Electrical Engineering and Informatics (ICEEI), 2011 International Conference on*. 2011, p. 1–6.

4. Budi, I., Bressan, S., Wahyudi, G., Hasibuan, Z.A., Nazief, B.A.A.. Named entity recognition for the indonesian language: Combining contextual, morphological and part-of-speech features into a knowledge engineering approach. In: *Proceedings of the 8th International Conference on Discovery Science*. Berlin, Heidelberg: Springer-Verlag. 2005, p. 57–69.
5. Larasati, S.D.. IDENTIC Corpus: Morphologically Enriched Indonesian-English Parallel Corpus. In: *LREC*. European Language Resources Association (ELRA). p. 902–906.
6. Larasati, S.D., Kuboň, V., Zeman, D.. Indonesian Morphology Tool (MorphInd): Towards an Indonesian Corpus. In: *Proceedings of the Systems and Frameworks for Computational Morphology - 2nd International Workshop*. Zurich, Switzerland; 2011, p. 119–129.
7. Kamayani, M., Purwarianti, A.. Dependency Parsing for Indonesian. In: *International Conference on Electrical Engineering and Informatics, ICEEI*. IEEE; 2011, p. 1–5.
8. Green, N., Larasati, S.D., Žabokrtský, Z.. Indonesian Dependency Treebank: Annotation and Parsing. In: *Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation*. Bali, Indonesia: Faculty of Computer Science, Universitas Indonesia; 2012, p. 137–145.
9. Levy, O., Goldberg, Y.. Dependency-Based Word Embeddings. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, Volume 2: Short Papers*. Baltimore, MD, USA: The Association for Computer Linguistics. p. 302–308.
10. De Felice, R., Pulman, S.. Automatic Detection of Preposition Errors in Learner Writing. *CALICO Journal* 2009;**26**(3):512–528.
11. Wagner, J., Foster, J., van Genabith, J.. Judging Grammaticality: Experiments in Sentence Classification. *CALICO Journal* 2009; **26**(3):474–490.
12. Rozovskaya, A., Roth, D.. Training Paradigms for Correcting Errors in Grammar and Usage. In: *Proceedings of the Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics*. Los Angeles, California, USA. 2010, p. 154–162.
13. Gamon, M.. Using mostly native data to correct errors in learners' writing: A meta-classifier approach. In: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*; HLT '10. Stroudsburg, PA, USA: Association for Computational Linguistics. p. 163–171.
14. Sidorov, G., Gupta, A., Tozer, M., Catala, D., Catena, A., Fuentes, S.. Rule-based system for automatic grammar correction using syntactic n-grams for english language learning (l2). In: *Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task*. Sofia, Bulgaria: Association for Computational Linguistics; 2013, p. 96–101.
15. Hernandez., S.D., Calvo, H.. CoNLL 2014 Shared Task: Grammatical Error Correction with a Syntactic N-gram Language Model from a Big Corpora. In: *Proceedings of the 18th Conference on Computational Natural Language Learning: Shared Task*. Baltimore, Maryland: ACL; 2014, p. 53–59.
16. Mikolov, T., Yih, W., Zweig, G.. Linguistic Regularities in Continuous Space Word Representations. In: *Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta, Georgia: Association for Computational Linguistics; 2013, p. 746–751.
17. de Marneffe, M., Manning, C.D.. *Stanford Typed Dependencies Manual*. 2013.
18. Rozovskaya, A., Roth, D.. Generating confusion sets for context-sensitive error correction. In: *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA: Association for Computational Linguistics; 2010, p. 961–970.
19. Mizumoto, T., Komachi, M., Nagata, M., Matsumoto, Y.. Mining Revision Log of Language Learning SNS for Automated Japanese Error Correction of Second Language Learners. In: *Proceedings of 5th International Joint Conference on Natural Language Processing*. Chiang Mai, Thailand: Asian Federation of Natural Language Processing; 2011, p. 147–155.
20. Irmawati, B., Komachi, M., Matsumoto, Y.. Towards Construction of an Error-Corrected Corpus of Indonesian Second Language Learners. In: Almeida, F.A., Barrera, I.O., Toledo, E.Q., Cuervo, M.S., editors. *Input a Word, Analyse the World: Selected Approaches to Corpus Linguistics*. Newcastle upon Tyne: Cambridge Scholars Publishing; 2016, p. 425–443.
21. Quasthoff, U., Richter, M., Biemann, C.. Corpus Portal for Search in Monolingual Corpora. In: *Proceedings of the 5th International Conference on Language Resources and Evaluation*. Genoa; 2006, p. 1799–1802.
22. McDonald, R., Lerman, K., Pereira, F.. Multilingual Dependency Analysis with a Two-stage Discriminative Parser. In: *Proceedings of the 10th Conference on Computational Natural Language Learning*. Stroudsburg, PA, USA: ACL; 2006, p. 216–220.
23. Dahlmeier, D., Ng, H.T.. Grammatical error correction with alternating structure optimization. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*. Stroudsburg, PA, USA: Association for Computational Linguistics. p. 915–923.

sltu

ORIGINALITY REPORT

12%	9%	10%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	natural-sciences.nwu.ac.za Internet Source	2%
2	Heck, Michael, Sakriani Sakti, and Satoshi Nakamura. "Unsupervised Linear Discriminant Analysis for Supporting DPGMM Clustering in the Zero Resource Scenario", Procedia Computer Science, 2016. Publication	1%
3	Lecture Notes in Computer Science, 2011. Publication	1%
4	Submitted to United States Coast Guard Academy Student Paper	1%
5	researchr.org Internet Source	1%
6	repositorio.ufpe.br Internet Source	<1%
7	"Natural Language Processing and Chinese Computing", Springer Science and Business Media LLC, 2018 Publication	<1%

8	Muhammad M. Rafique, Zhichun Zhu. "CAMPS", Proceedings of the 47th International Conference on Parallel Processing, 2018 Publication	<1%
9	Submitted to University of Sheffield Student Paper	<1%
10	web.tohoku.ac.jp Internet Source	<1%
11	Hariyadi, , and Hiroki Tamai. "Enhancing the Performance of Porous Concrete by Utilizing the Pumice Aggregate", Procedia Engineering, 2015. Publication	<1%
12	journal.equinoxpub.com Internet Source	<1%
13	cyberleninka.org Internet Source	<1%
14	journals.plos.org Internet Source	<1%
15	Lecture Notes in Computer Science, 2015. Publication	<1%
16	www.aclweb.org Internet Source	<1%
17	Lecture Notes in Computer Science, 2013. Publication	<1%
	Studies in Computational Intelligence, 2015.	

18	Publication	<1%
19	cl.naist.jp Internet Source	<1%
20	ijphysics.fi.itb.ac.id Internet Source	<1%
21	www.uleth.ca Internet Source	<1%
22	Jianbin Zhu, Xiaojun Shi, Shuanghua Zhang. "Machine Learning-Based Grammar Error Detection Method in English Composition", Scientific Programming, 2021 Publication	<1%
23	Lecture Notes in Computer Science, 2014. Publication	<1%
24	M. Iqbal D. Putra, Budi Irmawati, Wirarama Wedashwara, Dita Pramesti, Siti Oryza Khairunnisa. "Age Group Based Document Classification in Bahasa Indonesia", 2020 International Conference on Advancement in Data Science, E-learning and Information Systems (ICADEIS), 2020 Publication	<1%
25	aclweb.org Internet Source	<1%
26	epdf.pub Internet Source	<1%

27	link.springer.com Internet Source	<1 %
28	www.nactem.ac.uk Internet Source	<1 %

Exclude quotes	On	Exclude matches	< 3 words
Exclude bibliography	On		

sltu

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7