

INTISARI

AUTHOR OBFUSCATION DENGAN MENGGUNAKAN WORD2VEC DAN LANGUAGE MODEL

Oleh

FAHRUL ROZI MEIDUAN

15/379619/PA/16677

Author obfuscation merupakan suatu cara untuk melakukan parafrase dokumen dengan mengubah *writing style* dari dokumen. *Author obfuscation* merupakan salah satu cara untuk menjaga anonimitas *author* terhadap analisis *stylometry*. *Stylometry* sering digunakan untuk domain *authorship verification*, dimana pendekatan *stylometry* dapat menemukan fitur-fitur yang terdapat dalam sebuah dokumen, seperti *writing style*, maupun gender dari *author*. Proses keberhasilan *author obfuscation* bergantung pada pendekatan dari *author verification*. Semakin baik *author verification*, maka pendekatan yang sama dapat diaplikasikan untuk membuat *author obfuscation* yang baik.

Pada penelitian ini, digunakan Word2Vec dan *language model* sebagai metode dari *author obfuscation* yang dilakukan. Proses yang dilakukan adalah dengan melakukan *modulizing text* dengan memecah teks menjadi kalimat. Kalimat digunakan untuk menghitung posisi karakter dari setiap kalimat. Untuk mengubah data dari dokumen, dilakukan tokenisasi dan dilakukan penggantian sinonim yang didapatkan dari Word2Vec. Hasil dokumen yang didapatkan akan dilakukan pengujian dengan menggunakan tiga kondisi.

Hasil dari *author obfuscation* harus memenuhi tiga kondisi, yaitu *safety* dimana *author verification software* tidak dapat mengetahui *author* aslinya, *soundness* dimana dokumen yang di-*obfuscate* berhubungan secara tekstual dengan yang aslinya, dan *sensibleness* dimana dokumen tidak mencolok dari kata dan *style* yang digunakan. Hasil dari pengujian *safety* menggunakan GLAD dan mendapatkan penurunan akurasi sebesar 3.6%. Hasil pengujian dari *sensibleness* menggunakan pengecekan *language model score* dimana *perplexity* yang didapatkan masih mendekati nilai dokumen *original*. Hasil pengujian *sensibleness* menggunakan *peer review* dimana menunjukkan bahwa dokumen masih dapat diterima dan tidak terlihat seperti di-*generate* oleh mesin.

Kata-kata kunci : *natural language processing, word embedding, author obfuscation, language model*

ABSTRACT

AUTHOR OBFUSCATION USING WORD2VEC AND LANGUAGE MODEL

By

FAHRUL ROZI MEIDUAN

15/379619/PA/16677

Author obfuscation is an act of paraphrasing document by changing the writing style of the document. Author obfuscation is a way to maintain author anonymity from stylometry analysis. Stylometry is a form of authorship identification that relies on linguistics information in a document. Stylometry has been a well-known approach in authorship verification, where documents attribute detected, such as writing style and gender of the author. The successful from author obfuscation and author verification depends on each approach used. The better the author verification, then the same approach can be applied to create a good author obfuscation.

In this research, we used Word2Vec and language model as a method to make author obfuscation. The process is done by modifying text by breaking text into sentences. Sentences are used to calculate the character's position of each sentence. In order to change the data from dokumen, we used tokenization and synonym replacement that is obtained from Word2Vec. The results of this generated document will be evaluated using three conditions.

Result of author obfuscation must satisfied three conditions, that is safety which author verification can not tell who is the original author, soundness which paraphrased document can be accepted in human knowledge and the content is preserved, and sensibleness where document is inconspicuous. Safety evaluation is used by using GLAD and the accuracy dropped by this approach is 3.6%. Sensibleness evaluation is used by generating language model score and perplexity from each document, where this approach is still giving out results that obfuscated document is still near the original. Sensibleness evaluation is used by peer review and document generated is looks like not generated by machine.

Keywords : natural language processing, word embedding, author obfuscation, language model, Word2Vec