Identification Word Spacing of Erroneous Sentences on Indonesian Scientific

Sunda Ariana¹, Hadi Syaputra², Tri Basuki Kurniawan³
Computer Science Faculty/Universitas Bina Darma Jalan Ahmad Yani No.3, Plaju, Palembang - Indonesia
E-mail: sunda_ariana@binadarma.ac.id¹, hadisyaputra@binadarma.ac.id², tribasukikurniawan@binadarma.ac.id³

Abstract
Scientific writings in Indonesian language must the rules of the General Guide to the Reformed Indonesian Spelling especially in terms of word spacing. good word spacing is proportional to the width of the typeface, generally use the align right facility in order to have article look neat in typing, but causes space from word to word over space. Additionally, according to the General Guide book for the updated Indonesian spelling, there are rules that use the distance of words with punctuation after and before punctuation. This research is intended to make for detecting applications using word spacing of the Regular Expression String algorithm and to comment on articles based on principles that are in accordance with the General Guidelines for Indonesian ReformedSpelling.

Keywords: word spacing, Regular Expression, Indonesian Scientific.

Copyright © 2020 APTIKOM - All rights reserved.

1. Introduction
Bahasa Indonesia is the official language in education [1]. In accordance with Article 36 of the 1945 Constitution [2], it is stated that the Indonesian language as the State language has the function of being the official language of the state, the language of instruction in educational institutions, the means of communication at the national level for the interests of running the government and development, utilization of science, art, and modern technology. Unfortunately, not all language users use good and correct Indonesian, according to the context of their usage. On the other hand, the writing of scientific papers must use good and correct Indonesian language.

In the writing of scientific papers, researchers generally use the align right facility in order to have article look neat in typing. Unfortunately, the use of align right facilities causes spaces from word to word more than one space because good word spacing is proportional to the width of typeface [3], and also as required in the General Guide to the Reformed Indonesian Spelling [4], there are rules that use the distance of words with punctuation after and before punctuation. Therefore, the purpose of this research is how to build an application to detect word spaces using Regular Regression String algorithms [5], and add comments suggestions or notices on articles based on principles in accordance with the General Guide to the Reformed Indonesian Spelling rule.

2. Research Method
Typographically, coercion align right(combined without word decoding) causing reading problems or reading comfort. The readability is disturbed when the word spacing is not the same [6]. Other then that in the General Guide to the Reformed Indonesian Spelling there are rules in the use of word spacing and punctuation, the following is a figure of the use of word spacing before and after the punctuation.
By utilizing computer technology in the form of software can help in applying General Guidelines to Reformed Indonesian Spelling especially in using word spacing. The outline is to open the document, read the document, find space in the document, comment on the document, and create a new document that contains comments on the use of word spacing (Figure 2).

![Figure 2. System Architecture](https://pbs.twimg.com/media/Cso-1l3UAAA81JN.jpg)

#### Identification Word Spacing of Errorneous Sentences on Indonesian Scientific

To open a document file in the application requires addition of references or functions. So that allows the application to find and open the file along with filter what files will appear when performing the search process. The OpenFileDialog component allows users to browse the folders of their computer or any computer on the network and select one or more files to open. The dialog box returns the path and name of the file the user selected in the dialog box. Once the user has selected the file to be opened, there are two approaches to the mechanism of opening the file. If you prefer to work with file streams, you can create an instance of the StreamReader class. Alternately, you can use the OpenFile method to open the selected file [7].

The process of changing the word document becomes readable by the app using an external library Free.Spire.Doc for .NET [8], by calling the word document that has been read by the OpenFileDialog component, counting the number of pages in the document, counting the number of paragraphs on the page, and getting the text or words that are then compared to the word search. This process is done repeatedly until the number of paragraphs and the number of pages in the document are met and marked with comments on the document if the article in the paragraph is found.
The process of detecting strings using a regular expression algorithm [5]. In accordance with the figure (Figure 1), the following tables use of regular expression in programming language C# (Table 1).

| Punctuation                  | Symbol | Formula |
|------------------------------|--------|---------|
| Multiple word spacing        | .      |_regex\(\s\)\) |
| Word spacing before the dot   | ,      |_regex\(\s\)\) |
| Word spacing before the coma  | :      |_regex\(\s\)\) |
| Word spacing before the colon | ;      |_regex\(\s\)\) |
| Word spacing before the question mark | ? |_regex\(\s\)\) |
| Word spacing before the exclamation mark | ! |_regex\(\s\)\) |
| Word spacing after the hyphen | -      |_regex\(\s\)\) |
| Word spacing after the dash   | _      |_regex\(\s\)\) |
| Word spacing before the dash  | \      |_regex\(\s\)\) |
| Word spacing before the slash | /      |_regex\(\s\)\) |
| Word spacing after the slash  | \      |_regex\(\s\)\) |

In general can be described flow diagram for identification word spacing (Figure 3), and the process of searching the string with regular expression (Figure 4).

3. Results and Analysis

Testing on the application is done by testing the needs of functionality. Test functionality using black box method with the result that all functional system can be run as it should be. Figure 5
4. Conclusion

Writing scientific papers of Indonesian language follow the rules of the General Guide to the Reformed Indonesian Spelling, word spacing is the most important part in a writing, because with word spacing will help readability of writing. Errors can be identified with the help of the software. The search method uses the Regular expression method, the result of this identification software is to comment on a paragraph where there is an error in using word spacing.

References

[1] Sari, I. P. Bahasa Indonesia Sebagai Pemersatu Negara Kesatuan Republik Indonesia (NKRI). In Seminar Nasional Bulan Bahasa. UNIB. 2015: 234-242.
[2] E. Z. Arifin. Implementasi Pasal 36 UNDANG-UNDANG BAHASA. In Pujanga. 2015: 1-23.
[3] Strizver Ilene, Word Spacing. 2013 : 1.
[4] Grasindo, Tim. Pedoman Umum Ejaan Bahasa Indonesia & Pembentukan Istilah. Gramedia Widadarana, 2016.
[5] Thompson, Ken. Regular Expression Search Algorithm. Communications of the ACM 11.6 (1968):419–422. Communications of the ACM. Web.
[6] Swardjomo. Aspek Tipografi dalam Penyajian Karya Tulis Ilmu/Akademik/Profesional. In Seminar-Loka. 2008:1-12
[7] Jones, Mike, At all. How to: Open Files Using the OpenFileDialog Component. Microsoft. 2017. https://docs.microsoft.com/en-us/dotnet/framework/winforms/controls/how-to-open-files-using-the-openfiledialog-component. 26 March 2018
[8] Nuget. FreeSpire.Doc for .NET. 2016. https://www.nuget.org/packages/FreeSpire.Doc/. 27 March 2018