Local Feature Extraction for Noise Detection in Software Requirements Specification Document

Ahmad Mustofa1, a**,** Daniel Siahaan1,b

1Informatics Department, Institut Teknologi Sepuluh Nopember, Indonesia

amustofa.ahmad12@mhs.if.its.ac.id, bdaniel@if.its.ac.id

**Keywords:** software requirements, noise detection.

**Abstract.** Requirements specification is the first step of a software development cycle. If errors occurred in this step, errors would automatically occur in the next step. Errors in software requirements consist of noise, ambiguous, conflict, and inconsistency. This paper introduces a classification model for automatically detecting noises in Software Requirement Specification (SRS) document. The model was build based on local features extracted from each requirements statement within an SRS document. The local features are the statistical properties calculated by measuring the similarities between requirement statements within the same SRS document. Then, the set of data from the document were combined with other set of data from different documents to make a single dataset. This dataset was then used to build a classification model. The best model was obtained using Support Vector Machine (SVM) as classification method. Its performance was relatively high, i.e. 80% accuracy, 73% precision, and 60% recall.

Introduction

Detecting errors in software requirements holds an important role in software development since the first step in software development is requirement specification. The sooner an error is detected, the cheaper cost is needed to correct. In [1] the authors represented error in software requirement into 7 groups, renowned as Meyer’s seven sins. Noise, one of Meyer’s seven sin, is an error that caused by the presence of an element in the text that does not carry information relevant to any feature of the problem. There are 2 types of noise, redundancy and remorse. Remorse is defined as a requirement that holds irrelevant information to the problem domain, while redundancy is defined as a requirement that holds exactly same information with another requirement. Type of noise that will be used in this paper is remorse.

In past several years, machine learning has made a rapid progress in text processing. Bag-of-word (BOW) is a common approach to represent a text document as a vector of unique term/word [2]. Each column of the vector holds the weight value of corresponding term. In [2, 3] the authors use TF-iDF (Term Frequency-invers Document Frequency) as term weighting method to categorize text document. TF-iDF is also used in [4] to classify an email is either spam or not. Depends on these facts, TF-iDF will also be used in this paper as term weighting method.

Noise in Software Requirement Specification (SRS) document can only be recognized locally and since we treated a single requirement as a single document, a global BOW approach is not applicable. In this study, we proposed a statistical approach to represent a single requirement so then it can be used to classify a requirement is either a noise or not.

The remainder of this paper is organized as follows. Section 2 presents related works on text document classification. In section 3, we elaborate the method used to detect a noise in SRS document. Section 4 presents the performance evaluation results. Finally, we conclude the paper in section 5.

Related Works dihapus, digabung ke

All manuscripts must be in English, also the table and figure texts, otherwise we cannot publish your paper.

Please keep a second copy of your manuscript in your office. When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question. Should authors use tables or figures from other Publications, they must ask the corresponding publishers to grant them the right to publish this material in their paper.

Use *italic* for emphasizing a word or phrase. Do not use boldface typing or capital letters except for section headings (cf. remarks on section headings, below).

Methodology diganti Noise Detection Model

All manuscripts must be in English, also the table and figure texts, otherwise we cannot publish your paper.

Please keep a second copy of your manuscript in your office. When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question. Should authors use tables or figures from other Publications, they must ask the corresponding publishers to grant them the right to publish this material in their paper.

Use *italic* for emphasizing a word or phrase. Do not use boldface typing or capital letters except for section headings (cf. remarks on section headings, below).

Result and Analysis

All manuscripts must be in English, also the table and figure texts, otherwise we cannot publish your paper.

Please keep a second copy of your manuscript in your office. When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question. Should authors use tables or figures from other Publications, they must ask the corresponding publishers to grant them the right to publish this material in their paper.

Use *italic* for emphasizing a word or phrase. Do not use boldface typing or capital letters except for section headings (cf. remarks on section headings, below).

Organization of the Text

**Section Headings.** The section headings are in boldface capital and lowercase letters. Second level headings are typed as part of the succeeding paragraph (like the subsection heading of this paragraph).

**Page Numbers.** Do *not* number your paper:

**Tables.** Tables (refer with: Table 1, Table 2, ...) should be presented as part of the text, but in such a way as to avoid confusion with the text. A descriptive title should be placed above each table. Units in tables should be given in square brackets [meV]. If square brackets are not available, use curly {meV} or standard brackets (meV).

**Special Signs**. for example , α γ μ Ω () ≥ ± ● Γ {110} should always be written in with the fonts Times New Roman or Arial, especially also in the figures and tables.

**Macros**. Do not use any macros for the figures and tables. (We will not be able to convert such papers into our system)

**Language**. All text, figures and tables must be in English.

**Figures.** Figures (refer with: Fig. 1, Fig. 2, ...) also should be presented as part of the text, leaving enough space so that the capt­ion will not be confused with the text. The caption should be self-contained and placed *below or beside* the figure. Generally, only original drawings or photographic reproductions are acceptable. Only very good photocopies are acceptable. Utmost care must be taken to *insert the figures in correct alignment with the text*. Half-tone pictures should be in the form of glossy prints. If possible, please include your figures as graphic images in the electronic version. For best quality the pictures should have a resolution of 300 dpi(dots per inch).

Color figures are welcome for the online version of the journal. Generally, these figures will be reduced to black and white for the print version. The author should indicate on the checklist if he wishes to have them printed in full color and make the necessary payments in advance.

**Equations.** Equations (refer with: Eq. 1, Eq. 2, ...) should be indented 5 mm (0.2"). There should be one line of space above the equation and one line of space below it before the text continues. The equations have to be numbered sequentially, and the number put in parentheses at the right-hand edge of the text. Equations should be punctuated as if they were an ordinary part of the text. Punctuation appears after the equation but before the equation number, e.g.

c2 = a2 + b2. (1)

Literature References

References are cited in the text just by square brackets [1]. (If square brackets are not available, slashes may be used instead, e.g. /2/.) Two or more references at a time may be put in one set of brackets [3,4]. The references are to be numbered in the order in which they are cited in the text and are to be listed at the end of the contribution under a heading *References*, see our example below.

Summary

If you follow the “checklist” your paper will conform to the requirements of the publisher and facilitate a problem-free publication process.

References

[1] J. van der Geer, J.A.J. Hanraads, R.A. Lupton, The art of writing a scientific article, J. Sci. Commun. 163 (2000) 51-59.

Reference to a book:

[2] W. Strunk Jr., E.B. White, The Elements of Style, third ed., Macmillan, New York, 1979.

Reference to a chapter in an edited book:

[3] G.R. Mettam, L.B. Adams, How to prepare an electronic version of your article, in: B.S. Jones, R.Z. Smith (Eds.), Introduction to the Electronic Age, E-Publishing Inc., New York, 1999, pp. 281-304.

[4] R.J. Ong, J.T. Dawley and P.G. Clem: submitted to Journal of Materials Research (2003)

[5] P.G. Clem, M. Rodriguez, J.A. Voigt and C.S. Ashley, U.S. Patent 6,231,666. (2001)

[6] Information on http://www.weld.labs.gov.cn