

U-TAPIS: AUTOMATIC SPELLING FILTER AS AN EFFORT TO IMPROVE INDONESIAN LANGUAGE COMPETENCIES OF JOURNALISTIC STUDENTS

Niknik Mediyawati*, Julio Cristian Young, Samiaji Bintang Nusantara
Universitas Multimedia Nusantara, Indonesia
*e-mail: niknik@umn.ac.id

Abstract: The problem of Indonesian language errors among students is of particular observation. This problem becomes an important concern for students majoring in journalism because one day the graduates will become journalists. A language error filtering application has been developed that can be used quickly and accurately in journalists' work. This application, which involves statistical analysis, computational language, and artificial intelligence, is named U-Tapis. This study was aimed at finding out the feasibility and effectiveness measures of the U-Tapis model by focusing on the language of students' journalistic works such as opinions, news items, and news articles. The study is quantitative research involving 30 students majoring in Journalism, a private university in Jakarta, Indonesia. Data collection was carried out involving pre-test and post-test administration and descriptive statistical analyses. It was found that the students' error rate decreased after the use of the model. It can be concluded that, in addition to eligibility which reaches 92.31%, the U-Tapis application can help effectively increase students' proficiency in the use of the Indonesian language.

Keywords: *communication, Indonesian, journalism, language filter application.*

U-TAPIS: PENAPIS EJAAN OTOMATIS SEBAGAI UPAYA PENINGKATAN KEMAHIRAN BERBAHASA INDONESIA MAHASISWA JURNALISTIK

Abstrak: Masalah kesalahan berbahasa Indonesia di kalangan mahasiswa menjadi perhatian khusus. Masalah ini menjadi penting bagi mahasiswa jurusan jurnalistik karena kelak lulusannya akan menjadi seorang jurnalis. Telah dikembangkan aplikasi penapis kesalahan bahasa yang dapat digunakan secara cepat dan tepat saat jurnalis bekerja. Aplikasi yang melibatkan analisis statistik, bahasa komputasi, dan kecerdasan buatan ini diberi nama *U-Tapis*. Penelitian ini bertujuan untuk menentukan kelayakan dan keefektifan model aplikasi *U-Tapis* dengan memfokuskan pada bahasa karya jurnalistik mahasiswa seperti opini, berita, dan artikel berita. Penelitian ini merupakan penelitian kuantitatif yang melibatkan 30 orang mahasiswa jurusan Jurnalistik pada suatu universitas swasta di Indonesia. Pengumpulan data dilakukan melalui tes awal dan akhir melalui statistik deskriptif. Hasil penelitian menunjukkan bahwa tingkat kesalahan berbahasa pada mahasiswa menurun setelah menggunakan model *U-Tapis*. Dapat disimpulkan bahwa selain kelayakan yang mencapai 92,31%, aplikasi *U-Tapis* juga dapat membantu peningkatan kemahiran berbahasa Indonesia pada mahasiswa secara efektif.

Kata Kunci: *komunikasi, bahasa Indonesia, jurnalisme, aplikasi penapis bahasa.*

INTRODUCTION

Students' competencies in using the Indonesian language are below the expected standard. The Indonesian Government, through the Regulation of the Ministry of Education and Culture Number 70 Year 2016 (Permendikbud RI 2016 No. 70), stipulates that students' competencies in using Indonesian should be at the level of Excellence on the Test of Indonesian Language Competencies (TOILC).

Research finds repeated mistakes in the use of the language. The mistakes in the use of the Indonesian language are caused by inaccuracy, as the first rank, in-economy, as the second, and in-harmony, as the third. The inaccuracy of use is related to the carelessness in the application of the grammar of the language against the rules in the General Guide of Spelling in the Indonesian Language (GGSIL) and Grand Dictionary of the Indonesian Language (GDIL) such as in the cases

of writing affixed words, bound morphemes, compound words, prepositions, standard words, and particles and in typing mistakes. In-economy of use is found in overuse of conjunctives, double subjects, and redundancy of word uses. Meanwhile, in-harmony occurs in sentence grammatical structures and logics.

This finding on the pattern of the repeated mistakes in language use is in line with the finding of a survey on the reasons why students make the repeated mistakes; namely students' habits in using non-standard words and their unwillingness to open the dictionary to look up for word uses. Another reason is related to students' habits in adopting the language of media, both print and digital, taking these as correct examples of language use. The question remains whether the language of media can become a reference for the correct use of the language. As journalists would-be, students are trained to use Indonesian well and correctly since, in their future work, they will become models for the society in the use of the language. In the cases where students are not keen in using standard references, they have the inclination to imitate the language of the media which may not be in line with GGSIL and GDIL.

The Presidential Decree Number 63 Year 2019 (Perpres RI 2019 No. 63) about the use of the Indonesian language stipulates that Indonesian must be used in all elements of the Indonesian society. With the presidential decree, the Indonesian language becomes more dignified, both within and outside the country. In Chapter 2, items 1, 2, and 3, it is pronounced that Indonesian must be used appropriately and correctly; it must be used in line with the linguistic contexts and societal norms; and it must be used in accordance with the appropriate and correct grammar of the language and the situations and conditions of the language speakers. There is the question, again: what about the language used by newspaper reporters, news readers, masters of ceremony, radio/TV announcers, and other public speakers; whether or not their language use has been in accord with the Regulation of the Ministry of Education and Culture Number 70 Year 2016 about the standard competencies in the use of the Indonesian language (Permendikbud RI 2016 No. 70).

According to the data owned by the Body of Language and Book Development

(presently Body of Language Development and Nurturance), there have been 37,893 TOILC takers between 2005 and 2017, domestic and foreign citizens, among whom 19,313 are Indonesian language teachers, 964 are university students, 469 are school students, 74 are government employees, 100 are office workers, 46 are language envoys, 44 are language mentors, 7,647 are teachers of general subject-matters, 26 are university lecturers, 21 are bloggers, 19 are teachers of Indonesian for Foreign Speakers (IFS), 1,158 are foreign citizens, and 8,012 are other denominators. By looking at the data, it can be said that, compared to the populations of Indonesian and foreign citizens, people's interest in taking TOILC is still very low. The Government, represented by the Language Body, needs to improve the quality of the TOILC, primarily in the insistence of prescribing TOILC certification for native speakers such as, for example, journalists. This is because the mass media take an important role in the dissemination of information to the public. They also take an important role in making a frame for opinions and ideas from the public. Most of all, they play an important role in educating the public through the use of language. The mass media develop in a simultaneous way with the development of language.

In the development, bodies and practitioners in media in Indonesia use a number of regulations produced by the Press Council as their references. This council also supports journalist professionalism by applying the standard journalistic competencies that are in effect nationally. Referring to the Regulations of the Press Council Number 01/Regulation-PC/X/2018 about journalist standard competencies, mastery of language is one of the 11 key competency categories that must be mastered by journalists (Peraturan Dewan Pers No. 01/Peraturan-DP/X/2018). A key competency, still according to this regulation, is one that must be mastered by journalists in order to achieve performance that is required to carry out a task in a certain competency unit.

In practice, the language competencies and skills of journalists have not fully satisfied the competency standards. This is doubled by the challenges of the journalists in the digital era. Online media practitioners have to race in distributing news to the public; but they often have to ignore accuracy in the use of the language,

accuracy of information, and facts. Competition in the speed of the news distribution and gaining readers in the Internet often compel online media to practice clickbaiting on the distributed news titles. Hadiyat (2019) states one of the media that is widely accessed by the public because of the practice of clickbaiting is *tribunnews.com*. According to the data in the site alexa.com, one rank maker of the world websites, *tribunnews.com* occupies the third rank in Indonesia in the number of accessors (alexa.com, 2020). Looking at this *Alexa* data, in the case of *tribunnews.com* upload news items using Indonesian inaccurately, or even incorrectly, the public readers may procure incorrect understanding.

Inaccuracies and errors in news scripts by the media have become a concern among researchers (Maier, 2005; Maier 2009). Some journalists consider grammatical errors to be minor errors (Appelman & Schmierbach, 2018). However, writing errors in online news scripts cannot be seen as ordinary mistakes or a form of editorial inadvertence that can be understood by news readers. A number of previous studies have revealed that inaccuracies and errors in news affect media reputation (Nemeth & Sanders, 2000; Appelman & Hettingga, 2020).

According to the data owned by the Press Council, there are presently more than 43,000 online media in Indonesia (Dewan Pers, 2018). However, use of Indonesian in the appropriate and correct ways is not fully done by these media. According to Ambarita (2020), *Tribunnews.com* Content GM, *Tribunnews.com* uploads 3,000 news articles daily. A news reporter is assigned to write 20 news items. The writing and editing processes of the news items depend much on the language competencies of the news writers, either as reporters or editors. Looking at the amount of the daily work of editors, the possibility of errors in each news article becomes relatively high.

Looking at these data, there is an urgent need to provide journalist would-be students with competencies of appropriate and correct Indonesian use in order that, when they enter the world of work, to be journalists, they can really become educators of the people through language. This condition has become an opportunity to carry out a research study that is intended to design a technological model with the objective of reducing the mistakes in the spelling and diction in the use of the Indonesian language

as an effort to improve the competencies of journalists in the use of the language.

It is high time to utilize technology in language use. In line with the increasing use of the computer in the society, research in the application of the computational algorithm to identify and correct mistakes in spelling has been done since the 1960s (Zukarnain Abbas, Wayan, Trisetyarso, & Kang, 2020). Earlier research studies divide the computer programs for spelling mistakes into two, spelling checkers and spelling correctors. In spelling checkers, algorithm is designed to receive texts and identify spelling mistakes. In spelling correctors, the computer program is designed to detect spelling mistakes and give alternatives for correcting the mistakes (Peterson, 1980).

Scientists in the field of artificial intelligence continue to develop techniques and algorithms that are capable of correcting spelling mistakes by computer programs. One of the challenges of such enterprise is to develop a program in checking and correcting spelling mistakes in an interactive fashion and shortest possible time (Hládek, Staš, & Pleva, 2020). It is why a need is felt to devise research that produces a program in the form of an application to correct spelling mistakes. A program, called *U-Tapis*, has been developed as a tool to correct mistakes in Indonesian language use as a method in improving students' language competencies. The present study is aimed to describing the effectiveness and feasibility of the *U-Tapis* application model focusing on students' journalistic writings such as opinions, news items, and news articles. It is expected that the *U-Tapis* computer program can be found as being able to improve students' language competencies and help them in writing various news articles or reports in appropriate and correct Indonesian.

METHOD

The study used quantitative research approaches. The preparation of the study was as in the use of observation and interviews during collecting data for needs analyses. The quantitative approach was used to analyze the results obtained from the questionnaires of needs analysis and implementation of the developed instructional materials to find significant differences in the use of the *U-Tapis* application. The research design of the study was that of one

group pretest-posttest, conducted to test whether there was a significant difference in the scores of the samples before and after being subjected to the *U-Tapis* research treatment.

The study was also computational in the sense that it applied a computational algorithm to solve a complex problem using a cluster of big data. A computational approach is regarded as an appropriate problem-solving device since it is capable of relating theoretical models, mathematical functions, and statistics to empirical research (Corten, 2014). Algorithm is used solve a problem by computer processes. In a simple image, an algorithm is a set of instructions that must be executed to convert inputs into outputs. The algorithm is then changed into a computer program through programming language to optimize its performance using data samples or experiences obtained in the past; this is referred to as computers that have the capability of learning or machine learning (Alpaydin, 2010).

The method used in the *U-Tapis* application is a text analysis that is also known as a text mining. This analysis involves statistical analytical processes, computational linguistics, and artificial intelligence (Gandomi & Haider, 2015). It is aimed at analysing unstructured text data to become more structured to construct a pattern so that the data give insights and make it easy for the user to take decisions and conclusions (Sarkar, 2019).

Based on the results of the validity tests by experts, it was found that the *U-Tapis* had the feasibility to be used as an application to improve language competencies. The two experts, in linguistics and educational technology, used

six elements of feasibility; namely general dimension, material substance, learning, software engineering, visual communication, and language. The total criterion score obtained by the linguistic and educational technology experts was 208. With the highest possible score of 4, six dimensions of A-F, number of items = 26, and number of experts = 2, and total score from the obtained data = 192, it can be concluded that both the experts gave the appraisal that the *U-Tapis* was highly feasible as a software application of language correction capable of improving language competencies. According to the two experts, the percentage of the validity was 92.31%.

The research subjects were 30 students of the Journalistic Department of a private university in Jakarta, Indonesia (10% of the population). The instrument for testing the feasibility of the use of the *U-Tapis* was one adapted from the model of Wahono (2006) referring to a multimedia development with six dimensions of general dimension, material substance, learning, software engineering, visual communication, and language. As it has been shown above, the *U-Tapis* software is highly feasible as an application for language correction. Meanwhile, the instrument for appraising the effectiveness of the *U-Tapis* application was one based on the seven indicators of language structures of frequently-made mistakes; namely affixed words, bound morphemes, compound words, prepositions, standard words, and particles and typing mistakes using GGSIL and GDIL as reference criteria. The screen display of the web *U-Tapis* is presented in Figure 1.



Figure 1. Screen Display of the Web *U-Tapis*

The following are the steps for using *U-Tapis* in learning. First, students were given the task of writing opinions in a mastered language. Second, the opinion assignment file was entered into the *U-Tapis*. Third, the students analyzed the number of language errors in the *U-Tapis* version. Fourth, students received feedback and were asked to directly correct language errors. Finally, the lecturer gave an assessment, both in terms of language and content.

Data were collected by subjecting the research participant students to a pre-test in the form of writing journalistic texts of news and opinion articles. The response texts were graded using seven language indicators based on GGSIL and GDIL. The scores obtained were used to show students' language competencies before they received any research treatment. Students were then given the research treatment, the use of the *U-Tapis* application to minimize language mistakes. The post test was conducted as the next step to the product effectiveness. Data were subjected to statistical analyses using SPSS 25 statistical software program; *t*-test for differences in scores between the pre-test and post test and normality and homogeneity tests for prerequisite provisions.

FINDINGS AND DISCUSSION

Findings

The research pre-test was given prior to the experimental treatment of the use of the *U-Tapis* computer software. The test contained the seven indicators of grammatical structures frequently made by students; affixed words, bound morphemes, compound words, prepositions, standard words, and particles and typing mistakes. The test results showed 14% of mistakes in capitalization, 86% in italics, 38% in the use of affixes, 25% in prepositions, 21% in compound words, 10% in bound morphemes, 4% in particles and 2% in the use of numbers, and 70% in the use of commas, 23% full stops, and 7% brackets. Mistakes in the use of effective sentences were inaccuracy = 55%, in-economy = 37%, and in-harmony = 8%. The design of the *U-Tapis* application was of the earlier version so that the filtering capacity was limited to the word level in affixed words, bound morphemes, compound words, prepositions, standard words, particles, and typing mistakes.

Based on these data, a treatment needed to

be performed on the use of the *U-Tapis* in order to minimize the mistakes. The post test was used to appraise students' language competencies with the help of the *U-Tapis* application. In the 30 sheets of the students' journalistic writing, the system was able to find 130 of the 137 mistakes made (approximately). Then, from the 130 identified mistakes, the system was able to make 87 corrections (approximately 66.92%). Even though the system was able to correct about 33.08% of the mistakes, from the display of the system, students were given the advantages of being able to correct the mistakes by themselves by looking at the GGSIL and GDIL.

Further analyses showed that the *U-Tapis* system failed to correct the following mistakes: (1) typing mistakes in that borrowed words were not italicized, (2) typing mistakes in the absence of a space between two words, and (3) typing mistakes in that the initial letter of a name entity was not capitalized. Meanwhile, the system failed to identify mistakes in rare words that were not in the system's dictionary and in capitalization of certain words since the system had assumed that capitalization was effected only on words that signified name entities.

From the post test data, it was found that the students' level of mistakes decreased as evident from the obtained indicators and scores. Furthermore, a significant difference was found in the students' average scores between the pre-test and post test. The next step was comparing the pre-test and post test scores. Tables 1 shows the descriptive statistical data of the pre-test and post by *paired samples statistics* and the normality test by *one-sample Kolmogorov-Smirnov test*. Table 2 presents the results of Test of Homogeneity of Variances.

Table 1. Results of Normality Test

<i>M</i>		Pre	Post
<i>N</i>		30	30
Normal parameters ^{a,b}	Mean	24.97	53.00
	Std. deviation	3.44	2.91
	Std. error mean	.628	.532
Most extreme differences	Absolute	.144	.134
	Positive	.144	.134
	Negative	-.144	-.087
Test statistic		.144	.134
Asymp. sig. (2-tailed)		.111 ^c	.176 ^c

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

Table 2. Results of Homogeneity Test

	Levene Statistic	df1	df2	Sig.
Score Based on mean	2.210	1	53	.143
Based on median	1.208	1	53	.277
Based on median and with adjusted <i>df</i>	1.208	1	50.22	.277
Based on trimmed mean	2.289	1	53	.136

The significance statistics $.111 > .05$ and $.176 > .05$ show that the data distribution before and after *treatment* is normal. Results of the homogeneity test of Variance also show that the variances of the two score sets are not significantly different ($Sig. = .143 > .05$).

Results of the *t*-test show $t(29) = -35.420$, $Sig. = .000 < .05$ indicating that there is a significant difference between the scores before and after the *U-Tapis* treatment. It can be taken to mean that students' level of mistakes decreased and that the *U-Tapis* use helped in effective writing.

Discussion

It has been shown that the language competence low level of the Journalistic students of a private university in Jakarta, Indonesia (approximately under the TOILC average level) is related to the repeated mistakes that they make in their writing. It is most probable that these repeated mistakes are caused by students' habits. They are in the habit of using non-standard words, of doing what they did in high school, and of imitating the language of the Internet media that they think as correct. Besides, students lack initiatives to check uses according to GGSIL and GDIL (impractical, lazy, time consuming), do not care, are reluctant to search for correct answers, and, even, are ignorant towards grammatical rules. All the causes that make students repeat mistakes are plotted in a pie chart as Figure 2.

Considering the aforementioned description, an effective method is needed to help students have the conscience to use Indonesian well and appropriately and attractively and in coherence with their world of work which is always fast and practical. It is high time to use the Indonesian language in synergy with technology in order to satisfy the demands of the youth such

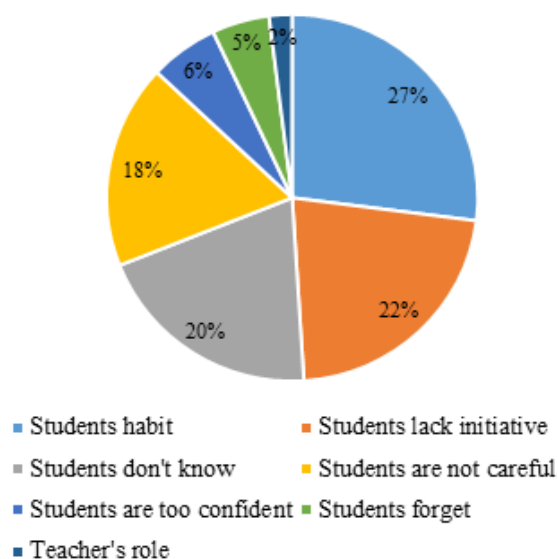


Figure 2. Causes that Make Students Repeat Mistakes in the Indonesian Language Use

as the journalistic students as journalists would-be who must become good models in the use of the Indonesian language as a tool to educate the society through the use of language. Research in the field of language and computation has been conducted by experts in language and computers. In accord with the development of the Internet technology which makes possible the fast flow of information and big data, research in language and computation continues to grow. In the Internet era, everyday the society holds communication and information sharing in the forms of texts, audio, or visual using digital applications. Experts in language and computation develop computation and analysis models that are able to identify linguistic forms and patterns that can help members of the society in communicating (Lai, Chui, Sah, Chung, & Liu, 2018).

Research in text analytics most generally uses computational algorithm that automatically becomes more intelligent or machine learning. Research studies that use machine learning for text analytics often focus on big themes. Examples of such big themes are text mining on the Internet sites or social media, text classification, text summarization, and text categorization (Sarkar, 2019).

The mass-media industry is one of the sectors that make a great use of language and informational technology to disseminate news items and other information pieces daily. Accuracy of information becomes an important

element of the media in order that the public can immediately understand the news contents (Kovach & Rosenstiel, 2010). Accuracy of information is interpreted as objective coherence (Kovach & Rosenstiel, 2001) and accuracy in language use including spelling, word forms, sentence grammar, syntax, and linguistic logics.

Digitalization of media brings along the changes in the production and consumption patterns of information. In the production processes of news, a number of foreign media and news headquarters try to develop automated journalism practices. This is a form of the combined process of journalistic work carried out by human beings by making use of computational algorithm and machine learning (Thurman, Lewis, & Kunert, 2019). A number of parties refer to automated journalism as “robo-journalism” since the computational technology works like a robot in writing news items from big data in the machine learning (Carlson, 2015). The ability of the machine learning to help journalistic work to develop sentences into a news text cannot surpass human’s skills (Clerwall, 2014). These machines, however, have been employed in helping the production processes of news items by media editors (Lindén, 2017).

In previous research, automated journalism is reported to be conducted generally by foreign media or news headquarters. The development of automated journalism in the Indonesian language is relatively rare. Media Beritagar, before being closed, made use of artificial intelligent in news item writing (Aditya, 2016). Different from Beritagar, the present study is intended to develop a machine learning-based application that is capable of correcting mistakes in language use.

In accordance with the development of the domain in natural language processing, both in the academic and industrial fields, there are a number of methods that can be used to overcome the problems of mistakes in writing with the urgencies that have been described above, especially for automated journalism. Symspell (Garbe, 2019) is one technique that can be used to correct mistakes in writing. Preceding research (Ljunglöf & Kjellberg, 2018) shows the potentials that Symspell has and its performance in correcting English words with an accuracy of 84%. This figure is 2% better than the technique

of word correction offered by machine learning-based models (Word2Vec). Furthermore, other studies also mention that SymSpell is often used during the pre-processing stage in natural language processing for correcting spelling mistakes, to be then followed up by other tasks such as classification (Cambray & Podadowski, 2019) or clustering (Guo, Liu, Cai, Chen, Qian, & Li, 2019).

There are other studies that are related to applications in the academic world such as the one conducted by Qumillaila, Susanti, & Zulfiani (2017). In this study, they developed instructional media for biology learning using augmented reality technology of the Android version. The study produced an AR application of the Android version in the instructional topic of the body excretion system. Another study in the development of project-based learning media to provide foundational knowledge and improve scientific literacy was conducted by Muskania & Wilujeng (2017). The study showed feasibility of use of the project-based instructional media in Physics Education Learning, rated very good in the criterion. In its statistical analyses, the independent-sample *t* test score showed that the project-based learning media gave a significant impact on the students’ scientific literacy.

In line with the study by Hládek *et al.* (2020), scientists in artificial intelligence continue to develop techniques and algorithms that are capable of identifying and correcting mistakes in a computer program. The presence of the *U-Tapis* aims at responding to these challenges in an interactive way and shortest possible time so that it can be used as a learning media in improving students’ competencies in using the Indonesian language.

Indonesian is an official national language that must be used in all the territory of the Republic of Indonesia. In the Law RI 2009 No. 24, Chapter 40, Indonesian is an identity symbol of the country. The Law reflects the importance of the appropriate and correct use of the Indonesian language by all layers of the society. Following the implementation of the Law, the President of the republic of Indonesia issued the Government Regulations Number 16 Year 2010 and Number 63 Year 2019 about the Use of the Indonesian Language to regulate the use of the language in all fields including documents and official state speeches, administrative services by government

institutions, national or international forums conducted in Indonesia, writing of academic works, dissemination of news items by the mass media, and others. After being tested for effectiveness with an effective result and for feasibility with a very feasible result, it is expected that the *U-Tapis* computer application software will be able to help in elevating the competencies of using Indonesian of students' as journalists would-be who will be working in the mass-media field in line with the Regulation of the Minister of Education and Culture Number 70 Year 2016 (Permendikbud RI 2016 No. 70) about the competency standards of the use of the Indonesian language. Students must achieve the Excellence level of the TOILC. The screen display of the web *U-Tapis* is presented in Figure 3.

Looking at the regulations and policies concerning the dissemination of information by the mass media as described above, it can be understood that all parties that work in the field must have the responsibility that the dissemination of information must be free from language mistakes. In the practice, however, mistakes in the use of the language are still frequently found in items and articles that are issued and read by the public. These are most frequently mistakes in spelling, word choice, and clause and sentence constructions. In the conditions where the practices of language use are still inconsistent with what is expected by the Government, other policies are needed in order to improve the conditions. In the space of the last decade, informational technology has developed fast. In all fields of work and services, informational technology has taken an important role in the business processes. Following this development, in the field of information dissemination, a need is felt to develop a language mistake correction system that will help the related parties in preparing and producing news items and articles. It is in this frame of thoughts that the *U-Tapis* system can be one of the methods in detecting language mistakes. By using the *U-Tapis*, journalistic students are trained to be ready to enter the world of work so that many language mistakes that often occur in the new articles can be avoided.

In addition to the efforts in the aspect

of human resources, such as the training of journalists would-be, it is certainly appropriate for the Government to issue policies that will encourage the mass-media enterprises to adopt language mistake detection systems such as the *U-Tapis*. By the implementation such policies, it is highly possible that the Government's expectations concerning the use of the Indonesian language can be achieved and will not much burden media business enterprises. The use of a language mistake detection system demands the involvement of a number of actors; namely those from the Government, professionals, and society members. The Government part takes the role of the initiator, supporter, and evaluator in the implementation of the program. This can be executed by the Press Council and the Language Body of the Ministry of Education and Culture.

Meanwhile, the professional actors play their roles in the implementation of the use of the system, giving periodical inputs and feedbacks for the betterment of the functions of the system. These may consist of news reporters, editors, and staff of the news quality assurance. And, members of the society participate in reporting mistakes they find in news articles and giving opinions concerning the mistakes. Based on the reports and opinions from the members of the society, the mistake detection system can be improved and this can become useful learning for all the parties involved in the use of the system.

The initial target of the present study consists of the journalistic students in a private university, in Indonesia. In the subsequent stages, however, the target of the study is expected to expand to journalistic students in other universities, and, ultimately to the citizens of the country at large in line with the Government Law RI 2009 No. 24, Chapter 40. With the implementation of the Government's policies in encouraging media business enterprises to adopt the language mistake detection system, it is expected that, in the end, the society will read news items and articles which do not contain language mistakes. On its turn, by consuming news items and articles published in appropriate and correct language, the society will be educated to use Indonesian in the appropriate and correct manners.

CONCLUSION

This experimental study, use of the *U-Tapis* computer application by journalistic students, has shown that there is a significant difference in the students' scores before and after the *U-Tapis* treatment. The feasibility level of the use of the *U-Tapis* system as a tool for identifying language mistakes reaches up to 92.31%. It can be concluded that the *U-Tapis* application is feasible and effective to be used as a tool for identifying language mistakes.

The study is preliminary research. Subsequent research studies are expected to follow this up with a primary objective of developing the *U-Tapis* application to become more feasible and more effective. In the median term of reach, the development of the *U-Tapis* system will be designed to fill in the needs of the users in educational institutions, schools, universities, and press populations. Writing of academic and journalistic items or articles should satisfy the appropriate and correct use of Indonesian. In this relation, development of the application of the *U-Tapis* tool will help in the construction of quality texts in the language.

The *U-Tapis* application can also be used as an instructional tool or medium of attractive learning processes in the educational field in order to improve the language competencies of the students and teacher alike. In the long-range term, the *U-Tapis* application can satisfy the needs of various Government and private institutions to support communication purposes using the Indonesian language. In the final stage, it is expected that the *U-Tapis* system, in line with the inspiring philosophy, will be able to be used as a filtering tool for language mistakes in all needs and activities involving the rapid dan skillful use of Indonesian in the midst of the fast-flowing information and communication.

REFERENCES

- Aditya, R. (2016, 28 Februari). Teknologi robot untuk memburu berita. [Robot technology for news hunting]. *Beritagar.id*. <https://beritagar.id/artikel/sains-teknologi-robot-yang-memudahkan-kerja-pemburu-berita>.
- Alexa.com. (2020). *Top sites in Indonesia*. <https://www.alexacom/topsites/countries/ID>.
- Alpaydin, E. (2010). *Introduction to machine learning* (2nd ed). Cambridge, Massachusetts: MIT Press. https://kkpatel7.files.wordpress.com/2015/04/alppaydin_machinelearning_2010.pdf.
- Ambarita, D. (2020, Mach). *Kegiatan penyuntingan di Tribunnews.com* (Interview at Kompas Gramedia).
- Appelman, A., & Schmierbach, M. (2018). Make no mistake? Exploring cognitive and perceptual effects of grammatical errors in news articles. *Journalism and Mass Communication Quarterly*, 95(4), 930–947. <https://doi.org/10.1177/1077699017736040>.
- Appelman, A., & Hettinga, K. (2020). Correcting online content: The influence of news outlet reputation. *Journalism Practice*, 0(0), 1–18. <https://doi.org/10.1080/17512786.2020.1784776>.
- Cambray, A., & Podsadowski, N. (2019). Bidirectional recurrent models for offensive tweet classification. *arxiv.org*. <https://arxiv.org/pdf/1903.08808v1.pdf>.
- Carlson, M. (2015). The robotic reporter. *Digital Journalism*, 3(3), 416–431. <https://doi.org/10.1080/21670811.2014.976412>.
- Clerwall, C. (2014). Enter the robot journalist: Users' perceptions of automated content. *Journalism Practice*, 8(5), 519–531. <https://doi.org/10.1080/17512786.2014.883116>.
- Corten, R. (2014). *Computational approaches to studying the co-evolution of networks and Behavior in Social Dilemmas*. West Sussex, UK: John Wiley & Sons.
- Dewan Pers. (2018, 25 September). *Progam pendataan menjadi ujung tombak*. [The data collection program is the spearhead]. <https://dewanpers.or.id/berita/detail/1066/Progam-Pendataan-Menjadi-Ujung-Tombak>.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal*

- of Information Management*, 35(2), 137-144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>.
- Garbe, W. (2019). Wolf Garbe. *Github*. <https://github.com/wolfgarbe/SymSpell>.
- Guo, C., Liu, X., Cai, E., Chen, Y. V., Qian, Z. C., & Li, R. (2019). Topic ink: Visualizing disaster-related textual data using LDA topic modeling: Vast challenge 2019: Honorable mention for clear articulation of methodology. In R. Chang, D. A. Keim, & R. Maciejewski (Eds.). *IEEE Conference on Visual Analytics Science and Technology (VAST)*. Vancouver, BC: IEEE, pp. 124-125. <https://binxia.github.io/publication/2018-SEKE-liu-noisy.pdf>.
- Hadiyat, Y. D. (2019). Clickbait di media online Indonesia. [Clickbait in Indonesian online media]. *Jurnal Pekommas*, 4(1), 1-10. <https://doi.org/10.30818/jpkm.2019.2040101>.
- Hládek, D., Staš, J., & Pleva, M. (2020). Survey of automatic spelling correction. *Electronics*, 9(10), 1-29. <https://doi.org/10.3390/electronics9101670>.
- Kovach, B., & Rosenstiel, T. (2001). *The elements of journalism: What newspeople should know and the public should expect*. New York, NY: Crown Publishers.
- Kovach, B., & Rosenstiel, T. (2010). *Blur: How to know what's true in the age of information overload*. New York, NY: Bloomsbury.
- Lai, H. -L., Chui, K., Sah, W. -H., Chung, S. -F., Liu, C. -L. (2018). Language communities, corpora, and cognition. In S.-H. Chen (Ed.). *Big data in computational social science and humanities*. Taipei: Springer Nature, pp. 195-206. https://doi.org/10.1007/978-3-319-95465-3_9.
- Law RI 2009 No. 24. *Bendera, Bahasa, dan Lambang Negara serta Lagu Kebangsaan*. [Flag, Language, and National Emblem and National Anthem]. http://badanbahasa.kemdikbud.go.id/lamanbahasa/sites/default/files/UU_2009_24.pdf.
- Lindén, C. -G (2017). Algorithms for journalism: The future of news work. *The Journal of Media Innovations*, 4(1), 60-76. <https://doi.org/10.5617/jmi.v4i1.2420>.
- Ljunglöf, P., & Kjellberg, J. M. (2018, 7-9 November). *Interactive correction of speech recognition errors: Implementation and evaluation for English and Swedish*. Paper presented at The 7th Swedish Language Technology Conference, Stockholm. <https://gub.ub.gu.se/file/207625>.
- Maier, S. R. (2005). Accuracy matters: A cross-market assessment of newspaper error and credibility. *Journalism & Mass Communication Quarterly*, 82(3), 533-551. <https://doi.org/10.1177/107769900508200304>.
- Maier, S. R. (2007). Setting the record straight: When the press errs, do corrections follow? *Journalism Practice*, 1(1), 33-43. <https://doi.org/10.1080/17512780601078845>.
- Muskania, R. T., & Wilujeng, I. (2017). Pengembangan perangkat pembelajaran project-based learning untuk membekali foundational knowledge dan meningkatkan scientific literacy. [Developing a project-based science learning kit to provide foundational knowledge and improve scientific literacy]. *Cakrawala Pendidikan*, 36(1), 34-43. <https://doi.org/10.21831/cp.v36i1.8830>.
- Nemeth, N., & Sanders, C. (2009). Number of corrections increase at two national newspapers. *Newspaper Research Journal*, 30(3), 90-104. <https://doi.org/10.1177/073953290903000307>.
- Peraturan Dewan Pers No. 01/Peraturan-DP/X/2018. *Standar Kompetensi Wartawan*. [Journalist Competency Standards]. https://dewanpers.or.id/assets/ebook/buku/1901191536_Buku_2018_SKW.pdf.
- Permendikbud RI 2016 No. 70. *Standar Kemahiran Berbahasa Indonesia*. [Indonesian Language Proficiency

- Standards.]. https://jdih.kemdikbud.go.id/arsip/Permendikbud_Tahun2016_Nomor070.pdf.
- Perpres RI 2019 No. 63. *Penggunaan Bahasa Indonesia*. [The use of Indonesian]. https://jdih.setkab.go.id/PUUdoc/175936/Perpres_Nomor_63_Tahun_2019.pdf.
- Peterson, J. L. (1980). Computer programs for detecting and correcting spelling errors. *Communications of the ACM*, 23(12), 676–687. <https://doi.org/10.1145/359038.359041>.
- Qumillaila, Q., Susanti, B. H., & Zulfiani. (2017). Pengembangan augmented reality versi android sebagai media pembelajaran sistem ekskresi manusia. [Developing android augmented reality as a learning media of human excretory system]. *Cakrawala Pendidikan*, 36(1), 57-69. <https://doi.org/10.21831/cp.v36i1.9786>.
- Sarkar, D. (2019). *Text analytics with python: A practitioner's guide to natural language processing*. Bangalore, India: Apress.
- Singh, S., & Singh, S. (2020). Systematic review of spell-checkers for highly inflectional languages. *Artificial Intelligence Review*, 53, 4051–4092. <https://doi.org/10.1007/s10462-019-09787-4>.
- Thurman, N., Lewis, S. C., & Kunert, J. (2019). Algorithms, automation, and news. *Digital Journalism*, 7(8), 980-992. <https://doi.org/10.1080/21670811.2019.1685395>.
- Wahono, R. S. (2006, 21 June). Aspek dan kriteria penilaian media pembelajaran. [Aspects and criteria for assessing instructional media]. *RomiSatriaWahono.net*. <https://romisatriawahono.net/2006/06/21/aspek-dan-kriteria-penilaian-media-pembelajaran/>.
- Zukarnain, N., Abbas, B. S., Wayan, S., Trisetyarso, A., & Kang, C. H. (2019, 19-20 August). Spelling checker algorithm methods for many languages. Paper presented in the *International Conference on Information Management and Technology (ICIMTech)*, Jakarta/Bali, Indonesia. <https://doi.org/10.1109/ICIMTech.2019.8843801>.