Education and Information Technologies (2023) 28:13287–13321
https://doi.org/10.1007/s10639-023-11718-4

Testing of support tools to detect plagiarism in academic Japanese texts

Tolga Özşen[1] · İrem Saka[2] · Özgür Çelik[3] · Salim Razı[4] · Senem Çente Akkan[5] · Dita Henek Dlabolova[6]

Received: 28 December 2022 / Accepted: 6 March 2023 / Published online: 24 March 2023

Abstract
Plagiarism has been among the top forms of academic misconduct. Detective, reactive and proactive measures are taken to mitigate plagiarism in scholarly works. Text-matching tools play a significant role in the detection of plagiarism. Many studies have tested the performance of text-matching tools in detecting plagiarism from various perspectives. However, no study addressed the performance of such tools in ideographic languages, particularly Japanese. Considering the sharp increase in the number of academic Japanese text and plagiarism incidents in the Japanese context, it is essential to explore to what extent text-matching tools catch similarities in Japanese texts and respond to the needs of Japanese users. Within this scope, this study set out to explore the coverage and usability performance of text-matching tools in the Japanese language. We tested the coverage performance of 10 text-matching tools with five types of intentionally plagiarized documents. Also, we tested the usability performance via a feature checklist. The testing results suggested that the tools generally give a relatively higher performance on the usability side rather than the coverage aspect. Most tools have minimal coverage performance in the Japanese language. In the end, we provided takeaways for vendors, policymakers and educators.

Keywords Plagiarism · Text-matching tools · Ideographic languages · Japanese · Coverage · Usability

1 Introduction

In recent years, the interest of academia in academic misconduct (AM) issues, particularly plagiarism, one of the "cardinal sins" (Foltýnek et al., 2020, p. 17) of AM, has been growing. Moreover, research regarding plagiarism has been expanding its borders

✉ Tolga Özşen
tozsen@gmail.com; ozsen@comu.edu.tr

Extended author information available on the last page of the article not only methodologically or theoretically but also going beyond the lingua franca in terms of the languages as well (Abdelhamid et al., 2022; Chen & Macfarlane, 2016; Ehrich et al., 2016; Perkins et al., 2020). It has been discussed theoretically in previous works related to AM issues, including plagiarism, that the detection, prevention, and reaction towards it are counted as essential principles (Kier & Ives, 2022; McGowan, 2005). One of the important elements in realizing the principles mentioned above is text-matching tools,[1] widely known as plagiarism detection tools and/or plagiarism software. In this paper, we prefer to use the term "text-matching" tools not to give the impression that the software is capable of detecting plagiarism automatically without human involvement.

Text-matching services are tools that report the identical and/or similar parts of the original document with quantitative measurements such as numbers, rates, etc. However, in some cases, the similarities that the tool finds may be matches that should not be counted as plagiarism or actual similarity. In such cases, those who are not trained to interpret the reports produced by text-matching tools may draw false conclusions. Therefore, an expert evaluation is also required to interpret the results accurately. Currently, there are more than a hundred tools and tools on the market. Some of them are online tools, while some are downloadable tools (Foltýnek et al., 2020). Besides, several studies have questioned how far these tools are capable of detecting similarities. Considering that these tools have not been developed mainly for ideographic writing tools/languages, such as Japanese, most of the research has focused on texts written in alphabetic languages, regardless of whether it is English or other non-English languages (e.g., Abdelhamid et al., 2022; Ahmed, 2015; Ali et al., 2011; Elkhatat et al., 2021; Jadhav Sunayana & Lihitkar Shalini, 2021; Nadhri et al., 2021; Naik et al., 2015).

Ideographic languages such as Japanese can be positioned as 'local/regional' or 'minor' languages in terms of numbers compared to the English user population. However, the reality is that the number of Japanese language users, even only in higher education at the 'local level', is considerable. As of 2020, there are 323 junior colleges in Japan, 795 universities, and 643 graduate schools with nearly 200,000 academic staff and 4 million students (Japan Statistical Yearbook, 2022). Even as a foreign/second language, today about 4 million people, and 80,000 teachers in 20,000 institutions in approximately 140 countries are involved with Japanese (Japan Foundation, 2018). Needless to say, today's resources are also digital in Japan. The digital environment for academic research has been expanding and diversifying. On the other hand, studies carried out on tool testing in Japan so far are limited and generally focused on a specific tool (Fukaya et al., 2003; Odaka et al., 2003; Suzuki et al., 2009; Takahashi et al., 2007; Ueta & Tominaga, 2010). Yet, Japanese language texts have not been tested on tools that are not specifically developed based on the writing tool of Japanese.

One of the studies on text-matching tools, which was conducted by members of the European Network for Academic Integrity (ENAI) working group named'**T**esting of **S**upport **T**ools for **P**lagiarism Detection' (TeSToP hereinafter), is inspiring both in

[1] In this manuscript, the word "tool" refers any service, system, machine, website or downloadable program related to text matching and is grouped under one word for consistency only.

terms of its methodology, approach and findings. The TeSToP simply aimed to answer the question of "How far can these tools reach in detecting text similarities and to what extent are they successful?". The study focused on eight languages (Czech, English, German, Italian, Latvian, Slovak, Spanish, and Turkish), and provided a comparison based on detailed criteria with the authentic systematic methodology that was originally developed by Debora Weber-Wulff (Foltýnek et al., 2020).

This paper was motivated by the approach and methodology of the original TeSToP work. The methodology and protocols used in the original TeSToP project were revised and developed in accordance with the characteristics of the Japanese language. This study aims to analyze Japanese-written documents compiled from four different sources by comparing 10 regional (Japan-based) and international text-matching tools using two main criteria (coverage and usability) as in the original TeSToP. We expect to reveal the current status of the performance of text-matching tools for various stakeholders, such as vendors,[2] professionals (academics), students, and decision-makers in educational institutions, while creating opportunities, particularly for non-Japanese language vendors, to improve their algorithms related to checking similarities in the Japanese language. More importantly, we expect that the results of this study will be a source of inspiration for other ideographic and/or Asian languages, too.

2 Previous tests

With the digitalization of education, inappropriate use of sources and plagiarism have become more widespread (Masic, 2012) and visible. This digitalization also led to the emergence of text-matching tools, and these tools have been widely used to prevent plagiarism by detecting similarities in texts (Mostofa et al., 2021). These tools varied in their scope, performance, coverage, usability and many other criteria. Therefore, several studies attempted to compare these tools from different perspectives to reveal their functionality for the end users. However, as outlined in the original TeSToP project, most of these comparative studies only offer a straightforward overview of the tools or categorize them without addressing their evaluation performance and usability (Foltýnek et al., 2020).

Pre-TeSToP studies on text-matching tools either classify the tools or conduct a comparative analysis on their functional features (Foltýnek et al., 2020). The major classification categories are document vs source code, free vs private, online (web-based) vs offline (desktop version) and operating intra-corpally vs extra-corpally (Foltýnek et al., 2020). For example, the study of Lancaster and Culwin (2005) offers a classification based on the types of metrics that these tools use. They argue that classifying text-matching tools based on the type of corpora they operate allows these tools to be discussed and compared more accurately. Pertile et al.

[2] The vendor in this manuscript is referred not only to a commercial organization that provides a sort of platform to detect similarities on the texts but also the stakeholder contributes to educational and research activities on plagiarism and academic misconduct issues.

(2016) conducted a more comprehensive classification study on 17 tools under eight categories, including open source, free, private, platform, citation analysis, content analysis, structure analysis and paraphrasing plagiarism. They also classify the tools based on the detection methods. Nahas (2017) categorizes the text-matching tools in two categories (free and commercial) and provides a brief description of inspected tools. Although Nahas concludes with a suggestion for the best free and commercial text-matching tool, he does not provide any rationale for this claim. Lastly, Chowdhury and Bhattacharyya (2018) give a short overview of 31 text-matching tools and classify the tools into three categories; textual plagiarism detection tools, source code plagiarism detection tools, and both. Also, they briefly report on the usability functions of the tools from two aspects; namely, whether they are user-friendly and allow single/multiple submissions.

The comparative studies mainly focus on the performance and usability features of certain tools based on pre-determined criteria. The study of Bull et al. (2001) compared five text-matching tools from their functional features and performance. They used plagiarized texts in four categories (verbatim plagiarism, essay-mill texts, and collusive texts produced using online and offline documents) to test the performance of the tools. They found that the tools performed well in detecting similarities in the texts. Maurer et al. (2006) conducted a similar comparative study to test the similarity detection performance of the tools. They created a wide range of plagiarized documents and tested three tools using these documents. It was concluded that the tools were successful in detecting similarities in plagiarized texts from sources available online. However, these tools failed to detect similarity when the texts were not stored online or translated. Kakkonen and Mozgovoy (2010) conducted a comprehensive comparative study on eight tools with 84 documents which were created using various plagiarism-disguising techniques such as synonym replacement, intentional spelling errors, paraphrasing, etc. They found that the tools failed to detect any similarity in synonym-replaced and paraphrased texts. In their study, Birkić et al. (2016) compared four tools based on usability features such as application programming interface support, integration to learning management tools, database scope, etc. Also, they tested the performance of the tools for verbatim plagiarism without applying disguising techniques on the texts and presented their performance testing by quantifying a test score out of ten for each tool. The comparative testing of Krizkova et al. (2016) adopted a different methodology and tested the performance of five tools in two steps. In the first step, they tested the performance of the tools using verbatim texts, and in the second step, they conducted the testing by reordering the texts used in the first step. Then they suggested a ranking based on the testing scores. Vani and Gupta's (2016) small-scale comparative study tested three tools using the modified versions of an abstract. They found that the tools failed to detect any similarity in the translated and summarized texts.

Among the studies conducted to compare text-matching tools before the TeSToP project, the works of Debora Weber-Wulff come to the fore with their strong methodology and rigor in the testing process. In their most comprehensive study, Weber-Wulff et al. (2013) tested 15 tools by generating texts using different types of plagiarism techniques such as disguised plagiarism, translation plagiarism, structural plagiarism, shake & paste and pawn sacrifice. They also evaluated the usability features of the selected tools and combined the effectiveness and usability results to categorize the tools according to their usefulness for academic institutions. They noted that the tools suffer from two major problems, false positives and false negatives, which can lead to wrong interpretations of similarity reports. They conclude that such tools cannot detect plagiarism but offer some indicators for text matching. A more detailed inspection is needed by humans to decide whether the text match is plagiarism or not.

The methodology of the original TeSToP project relies on the methodology developed by Weber-Wulff et al. (2013), with some minor tweaks. Considering the number of tools tested and the number of testing documents along with language variety, TeSToP describes itself as the largest text-matching tool testing study ever conducted. In this study, Foltýnek et al. (2020) tested the tools based on a wide range of usability features and coverage performance. In terms of coverage, they tested the tools based on the language of the text for eight languages, types of plagiarism sources (Wikipedia, open access papers, student theses, online articles), plagiarism methods (copy-paste, synonym-replaced, paraphrased, translation) and source types (single-source, multi-source). At the end of the usability and coverage testing results, they classified the tools into four categories as useful, partially useful, marginally useful and not suited for academic institutions. They also proposed some takeaways to vendors, users and educators.

Several studies were conducted by other researchers to test the performance of text-matching tools after the TeSToP project. Building on the disguising techniques used by students to trick text-matching tools, Elkhatat et al. (2021) tested nine text-matching tools by creating intentionally plagiarized documents. They used four disguising techniques to test the documents: imaged-texts, using invisible quotation marks, using letter-like symbols and using invisible letters. They classified the tools into two categories as non-functional and partially functional against disguising techniques. They concluded that among the tested disguising techniques, the tools mainly failed to detect similarity in imaged texts, and they recommended the vendors employ OCR technology to detect imaged-text plagiarism. The study of Kulkarni et al. (2021) focused on semantic analysis rather than syntactic analysis, as most studies did, and presented a detailed taxonomy and methodologies for the identification of plagiarized contents. However, in terms of performance testing, they tested 11 free tools by using four sample journal papers without applying any disguising techniques but simply reporting the matching percentages of each tool. In the end, they listed ten challenges that should be considered in plagiarism detection by text-matching tools. Vrbanec and Meštrović (2021) approached the issue from a different perspective and conducted a toolatic study of corpus-based deep learning models in determining paraphrased plagiarism. They presented a comprehensive overview of corpus-based models in paraphrasing detection and tested the models by employing different approaches. They also compared corpus-based models with traditional approaches and concluded that corpus-based deep learning models perform similarly to traditional text-matching approaches and can be developed more effectively. Condurache and Bolboacă (2022) tested whether free or commercial text-matching tools perform better in detecting copy-paste plagiarism in the medical field. They tested the performance of four commercial and three free tools using a copy-paste document compiled from eight different sources. They concluded that free software performed better than commercial software in detecting copy-paste plagiarism. Lastly, Wahle et al. (2022) investigated whether text-matching tools can detect similarity in machine-paraphrased texts. They created 160 documents using randomly chosen paragraphs from Wikipedia, akXiv and theses. They paraphrased the texts using three neural language models, and concluded that the performance of text-matching tools in detecting similarity in machine-paraphrased texts varies.

Many of the comparative studies focus on the source or disguising technique-based performance of the text-matching tools. However, few studies explore the language-based differences in the coverage performance of these tools. The most comprehensive language-based testing was conducted in the original TeSToP project. Foltýnek et al. (2020) compared the coverage performance of 15 tools in eight languages (English, Italian, Spanish, German, Latvian, Slovak, Czech, and Turkish). They found that the performance of the tools dramatically changed according to the source language of the texts. They also aggregated the language-based results to explore in which language families the tools are better at identifying plagiarism. The study found that the tools performed better in Germanic and Romanic languages than in Slavic languages. In another study, Nadhri et al. (2021) tested three tools for two languages (English and French). The French corpus was composed of copy-paste documents, while the English corpus included paraphrased texts. It was found that the tools fell short in identifying plagiarism in the French corpus, which was due to the encoding problem. The tools performed better in identifying paraphrased texts in the English corpora. Other studies did not conduct a language-based comparison but proposed some new methods to detect similarities in non-English languages, such as French (e.g., Elamine et al., 2020, 2021) and Arabic (e.g., Alotaibi & Joy, 2021; El Bachir Menai & Bagais, 2011; Hussein, 2015; Kahloula & Berri, 2016; Nagoudi et al., 2018).

2.1 Aim of the study

In the literature, to our knowledge, no study has dealt with the performance of plagiarism-catching tools in the Japanese language. Many of the Japanese-related studies on text-matching tools concern new algorithmic or practical models to detect similarities in Japanese texts (Fukaya et al., 2003; Odaka et al., 2003; Suzuki et al., 2009; Takahashi et al., 2007; Ueno et al., 2006; Ueta & Tominaga, 2010). It is known that the prevalence of academic dishonesty, particularly plagiarism and difficulties establishing academic integrity, is on the rise in Japan (Wheeler, 2016). This posits that text-matching tools have great potential to be widely used in helping the detection of plagiarism in Japanese texts. However, the coverage performance of text-matching tools in Japanese texts remains a big question. From this perspective, our study aims to explore the current status of text-matching tools in Japanese texts.

3 Methodology

The methodology of our study was originally developed by Debora Weber-Wulff (Weber-Wulff, 2010; Weber-Wulff et al., 2013), and revised and developed in accordance with the characteristics of the Japanese language by the team members. In addition to the original framework, transitions between Japanese writing systems (Kanji, kana systems), Japanese-specific punctuation, and numeral systems (Chinese and Arabic) in the texts were also tested. We also tested whether the tools support encoding systems such as Shift-JIS and UTF-8 which support the Japanese language. In terms of sources, Japanese academic and government databases were used in addition to the sources used in the original work. The methodology is summarized under three subtitles: Documents and sources, Procedures, and Evaluation.

3.1 Documents and sources

3.1.1 Documents tested

In order to test the text-matching tools, we prepared five types of intentionally-plagiarized documents (Table 1). Document Type 1 is the text totally copied from Wikipedia and pasted onto a word document. Hyperlinks on the original Wikipedia source were removed.

Document Type 2 is a text automatically paraphrased by a free paraphrasing tool called "paraphraser"; which is often used among students.

The third type of testing document is a unique feature of this present study, contributing to the original TeSToP methodology. The Japanese language has three unique writing tools (Hiragana, Katakana, and Kanji). Those writing tools, which can be considered as an "alphabet" to people unfamiliar with ideographic languages, are intertwined and used together usually. However, in order to reveal whether the transition between writing tools is recognized by the text-matching tools, the original text was intentionally written only in Hiragana. Here, it should be mentioned that in the Japanese language, a text, especially an academic text, is never written only

Table 1 Documents to be tested for plagiarism

Source*	Type 1 (copy and paste)	Type 2 (rephrasing)	Type 3 (different writing tool)	Type 4 (translation)	Type 5 (disguising techniques)
Wikipedia Source A	A1	A2	A3	A4	A5
OAJP Source B	B1	B2	B3	B4	B5
Non-online book chapter Source C	C1	C2	C3	C4	C5
Mixed source Source D	D1	D2	D3	D4	D5

in Hiragana or Katakana, but such texts are common practice in primary schools. When faced with such a written text in high school or higher education, no special training is required to understand that it is something abnormal.

Document Type 4 is a translation (Japanese to English) of the original text. The translation was made by Google Translate on the web. The reason that Google Translate was preferred is that it is the most common and free tool for both non-native and native Japanese users or learners.

The last type of testing document is a text that contains five types of possible disguising techniques. Three of them are unique to the Japanese language, and two of them are universal techniques. In Japanese, numbers can be written in both wide and narrow text. For example, the number "340" can be written as "３４０" or as "340". Moreover, numbers can be displayed as numbers as well as in Kanji (Chinese characters). The number 340 can be displayed as "三四〇" too. On the other hand, the punctuation marks used in Japanese can also be different. For example, the dot "." in the alphabet can be written as "。" in Japanese text as well. These three unique possible disguising techniques were applied to the text to evaluate whether the tools could catch the differences in Japanese. As for the universal disguising techniques, an image of part of the text and white characters were also added to document Type 5.

3.1.2 Document sources

The documents explained in detail above were created by using four different sources. As a single source, Wikipedia (Source A), Open Access Journal Paper (Source B), and a non-online book chapter (Source C) were used. The mixed source document (Source D) was created as a combination of three different documents: Wikipedia, Open Access Journal Paper (OAJP hereinafter), and a Government White Paper (annual report). Source B was retrieved from the Japanese academic publication platform "J-STAGE", developed and managed by the Japan Science and Technology Agency (JST). Source C is a book chapter written in Japanese, published in Japan, which has never been online. Sources B and C are the products of one of the authors of this study and were included in this study with the author's consent. The 20 documents (texts) were created between 17 January and 28 March 2022 (see Table 1 and Appendix 1 for details).

3.2 Procedures

3.2.1 Sampling protocols, testing and data collection

Between February and May 2022, 83 tool vendors were emailed asking for their permission to be included in the test. Out of these, 10 agreed to participate in the testing, whereas others indicated their disinterest as their algorithms were not developed by taking ideographic writing systems/languages into consideration. The detailed workflow regarding the sampling process is as follows.

3.2.2 Forming the rater groups and inter-reliability tests

Three groups (A, B, and C) were formed to analyze the data and interpret the findings. Each group was composed of two raters: one a Japanese language expert (JLE) and the other a TeSToP expert. Both were members of the original TeSToP team. The distribution of the tools to the groups is as follows.

Group A: StrikePlagiarism.com, OXSICO, SmallSEOTools
Group B: Plagiarism Detector.net, Dupli Checker, Plagiarism Checker.co
Group C: Docol©c, CopyContentDetector, chiyo-co, Plagiarism Checker X

Before the analysis began, a TeSToP expert member of the team trained all team members between 29 June and 4 July 2022 to establish common ground for the evaluation and interpretation of the test results. Afterwards, the JLEs practiced their skills by means of pilot documents. The results from the pilot analyses were used to ensure inter-rater reliability among the JLEs. For this purpose, JLEs discussed their evaluations at a Zoom meeting through which it was confirmed that all raters tended to think alike in the evaluation and interpretation. Since the texts were in Japanese, the JLEs of each group firstly evaluated the tools in accordance with the systematic criteria explained in detail below between 4–17 July 2022. Evaluation results calculated by the JLEs were explained to the TeSToP expert's counterpart, who responded to the TeSToP expert`s questions and discussed cases, until they reached a consensus. Then, they finalized the decision for each tool they were assigned. All results of the test were sent on 16 August 2022 to each vendor to avoid any mistakes with the analyses. As of 1 September, responses from 4 tool vendors were received. Some points related to the technical features were addressed in accordance with feedback from the vendors.

3.3 Evaluation

The analysis framework has two main aspects with several sub-subjects to evaluate and interpret the tools (Fig. 1). In the coverage section of the test, the coverage ability of the tool was considered from five different aspects. In order to evaluate the output that the tool provides quantitatively, a metric from 0 (none) to 5 (all) was applied. To evaluate the other main aspect, usability, 20 items were tested in two subsections. The first subsection, *Usability: process and presentation of the results* (UP hereafter), aims to reveal how much the tool is functional/easy to use from the viewpoint of simple/basic users, via seven items. The second subsection, *Usability: technical features* (UT hereafter), tries to confirm whether the tool has these 13 technical features listed. All evaluation criteria for the coverage and usability sections were developed by the TeSToP-J research team, based on the original TeSToP project.

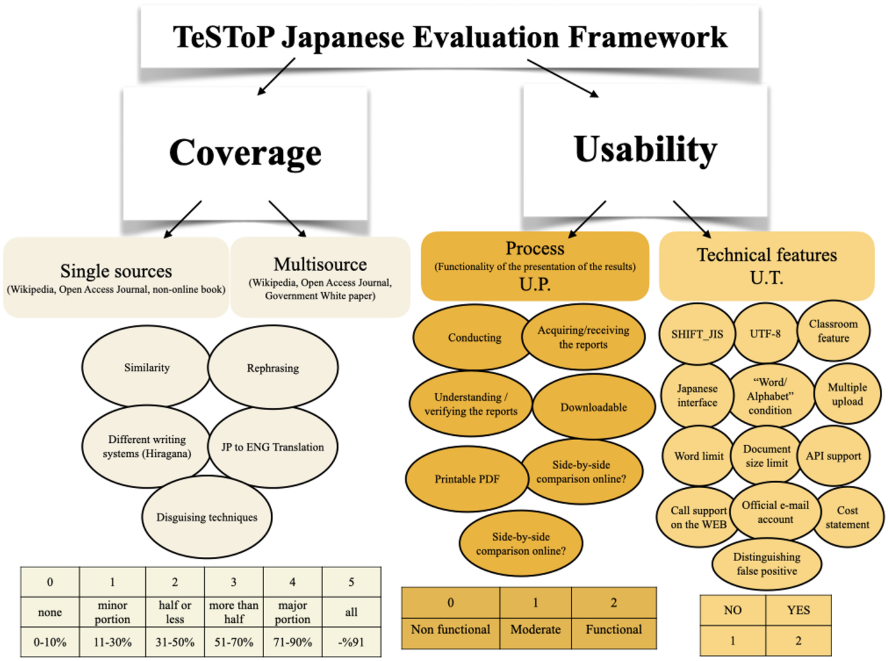

Fig. 1 TeSToP Japanese evaluation framework

4 Findings

4.1 Coverage

This section discusses how much the tools could quantitatively detect the similarities in the texts intentionally plagiarized. As stated in the Document and Sources section, five different text types (copy-paste text, paraphrased text, same text in different writing systems, translated text, text with disguising techniques) based on four sources (Wikipedia, OAJP, Non-online book chapter, and mixed text composed of Wikipedia, OAJP, and government reports), in a total of 20 texts, were used for similarity detection (See Tables 1 and 2). All tables in this section show the average scores of the evaluation (ranging from 0 to 5). Since the similarity reports are demonstrated by percentage, scores used in this section are also classified by the percentage that the tools report. Details regarding the scoring are as follows:

0 ("none") refers to 0–10%
1 ("minor part") refers to 11–30%
2 ("half or less") refers to 31–50%
3 ("more than half") refers to 51–70%
4 ("major part") refers to 71–90%
5 ("all") refers to 91% and above of similarity detection that the tool provides

Table 2 Sampling workflow

Date	Process	Details
February 2022	Listing the tools	94 tool vendors were listed from the database of the previous TeSToP project (10 out of 94 tools could not be reached (contact info))
February 2022	Composing invitation letter	The invitation letter was composed by the research teams
26 March 2022	E-mail to vendors	We failed to send an email to 1 of the 84 tools because of a word limit for the mail set by the tool. The invitation letter was therefore sent to 83 tool vendors, but a "failure mail" was returned from 6 tools). In the end we obtained 77 tools
26 March—4 April 2022	Responses from vendors	12 out of 77 tools responded. 9 were positive, and 3 were negative
5 April 2022	Reminder email to tools	Reminder emails were sent to the 65 tools that did not respond
5 – 21 April 2022	More responses	8 tools responded. 5 were positive, 2 were negative, and 1 was positive on payment conditions
Summary: between 26.03.2022 and 22.04.2022, 20 tool vendors responded with 14 positive and 6 negative responses. 57 tools did not return with any reply		
22 – 27 April 2022	Analysis process	10 tools were successfully investigated and data were obtained. However, downloaded applications of 4 tools failed to run
21 September 2022	Late response	One tool responded positively to the invitation letter but we could not include the tool in the study because it was too late for the data collection phase
Total Summary: Between 26.03.2022 and 21.09.2022, a total of 21 tool vendors responded with 15 positive and 6 negative responses. However, one of the tools could not be included due to the late response. 56 tools did not reply in any way		

Table 3 Coverage of overall results according to the type of sources

Sources	chiyo-co	Copy Content Detector	Docol©c	DupliChecker	OXSICO	Plagiarism Checker.co	Plagiarism Checker X	Plagiarism Detector.net	Small SEO Tools	StrikePlagiarism.com	Average of tools
Wikipedia	1.0	1.0	0,2	1.0	2.8	0.6	0.0	0.0	1.8	2.0	1.0
Open Access Journal Paper	0.0	0.6	0,0	1.0	2.4	0.0	0.0	0.0	1.8	0.2	0.6
Non-online book chapter	0.0	0.0	0,0	1.6	1.2	0.0	0.0	0.0	0.6	0.2	0.4
Multi-source	0.3	0.6	0,0	0.6	2.6	0.0	0.0	0.0	1.2	1.2	0.7
Overall average of sources	0.3	0.6	0,1	1.1	2.3	0.2	0.0	0.0	1.4	0.9	0.7

4.1.1 Coverage findings by sources

The results are presented according to the sources in Table 3. The average score of the coverage test for all tools is 0.7 points from 0 to 5. As expected, the tools did not demonstrate adequate performance. The average score of all tools shows us that Wikipedia is relatively the most detectable source among all sources. However, results also show that most of the similarities found by the tools refer to irrelevant websites/links, rather than the original text. In contrast, the non-online book chapter has the lowest matching score in the test, which makes the non-online documents in Japanese the most non-detectable source among others tested in this research.

We also aimed to explore to what extent the tools were able to detect similarities in documents created using the "patch-writing" technique, which is a form of plagiarism frequently used by students (Howard, 1992). Texts with short parts from different sources were compiled to test this. In addition, to see how much the single source is different from the mixed sources in terms of matching, the average score of A (Wikipedia), B (OAJP), and C (non-online book chapter) was used. It seems that the average scores of all tools in terms of both single and mixed sources are the same. However, the results in Table 4 indicate that the average scores are higher for the single source, when it is examined on a tool basis. Only two tools were unsuccessful in the single source, while half of the tools failed to detect the mixed source.

4.1.2 Coverage findings by plagiarism techniques

Regarding plagiarism methods, five different texts with different plagiarism techniques (copy-paste text, automatically paraphrased text, text written in a different writing system, translated text, and text with disguising techniques) were created

Table 4 Coverage scores for single-sources and multi-source

Scores by source (source types)	chiyo-co	Copy Content Detector	Docol©c	DupliChecker	OXSICO	Plagiarism Checker.co	Plagiarism Checker X	Plagiarism Detector.net	Small SEO Tools	StrikePlagiarism.com	Average of tools
Average score of single-sources (A+B+C)	0.3	0.5	0.1	1.2	2.1	0.2	0.0	0.0	1.4	0.8	0.7
Multi-source score	0.3	0.6	0.0	0.6	2.6	0.0	0.0	0.0	1.2	1.2	0.7

Table 5 Coverage of overall results according to the type of plagiarism technique

Plagiarism Techniques (text types)	chiyo-co	Copy Content Detector	Docol©c	DupliChecker	OXSICO	Plagiarism Checker.co	Plagiarism Checker X	Plagiarism Detector.net	Small SEO Tools	StrikePlagiarism.com	Average of tools
Copy-paste	1.6	2.8	0.3	2.5	3.0	0.3	0.0	0.0	2.8	2.3	1.5
Paraphrase	0.0	0.0	0.0	1.3	2.5	0.3	0.0	0.0	2.0	0.0	0.6
Hiragana	0.0	0.0	0.0	0.0	1.5	0.0	0.0	0.0	0.0	0.0	0.2
Translate	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.8	0.2
Disguising techniques	0.0	0.0	0.0	1.5	3.5	0.3	0.0	0.0	2.0	1.5	0.9

in four different sources: Wikipedia, OAJP, non-online book chapter, and a mixed source composed from Wikipedia, OAJP, and the government report. A total of 20 documents were tested to reveal how much the tools could detect/find the tampered excerpts from the texts. Table 5 shows that as an average score, the tools are still relatively strong in detecting copy-paste texts compared to the other document types. The average scores of copy-paste texts in terms of sources are Wikipedia: 2.9, OAJP: 1.2, non-online book chapter: 0.8, mixed source: 1.3.

These results lead us to consider that these tools have successful detection abilities to a certain extent, even if they are not specifically developed for ideographic languages, such as Japanese. However, the same interpretation does not apply to the other techniques. From the average score viewpoint, regardless of the source, automatically paraphrased texts and the texts in Hiragana in particular were the most unsuccessful to be detected. In terms of disguising techniques, scores may not reflect the actual situation. As aforementioned, disguised texts have five different techniques related to writing styles of numbers (in narrow-wide form in Chinese characters), punctuation marks, images, and white characters. Even though some tools have relatively high quantitative scores, output reports show that most matches are false-positive findings, while the rest are random or nonsignificant catches. Therefore, the matches regarding disguised texts are basically inconsistent and insignificant.

4.2 Usability

The Usability of the tools was evaluated using 20 objective criteria, which were divided into two groups whether they related to UP or UT. The main aim of the Usability criteria group is to see whether the tool is easily understandable and usable from the viewpoint of an average user. In order to discuss how much the tools are usable, therefore, the following questions were evaluated. Raters were asked to rate the usability of the tool from 0 (non-functional) to 2 (functional) as a basic user. Most of the evaluators of this study were using these tools for the first time. Accordingly, it can be said that the evaluation of the tool's process, presentation, and technical features was done literally by "beginner users". The following seven items refer to the criteria related to UP.

1) Conducting the tool is… (interface usage, uploading procedures, etc.)
2) Acquiring/receiving the reports is…
3) Understanding and verifying the reports is…
4) Downloading the results is…
5) Printing the results as a PDF is…
6) Allowing side-by-side comparison online…
7) Allowing side-by-side comparison offline…

4.3 Usability: Process and presentation of the results (functionality)

Since the text-matching tools provide similarity scores or ratios, not plagiarism results, the outputs should be carefully evaluated in detail. In order to proceed on solid ground, the results of the tools should then be available for further analysis. Output reports should be available for offline comparison, if necessary. Accordingly, tool outputs should be downloadable and printable. Moreover, outputs should be understandable and verifiable for the average user or the inexperienced interpreter who evaluates the results in order to decide whether the output of the student is questionable in terms of misconduct. These are some of the key points for assessing whether the tool is usable and functional for the main user.

As summarized in Table 6, the average score of 10 tools covering all criteria is 1.2 out of 2.0, and 6 tools have a higher score than the average. Even though each tool has such limitations, the overall results allow us to interpret that tools are relatively functional. Moreover, considering the technical features illustrated in Table 7, few tools have availability for side-by-side comparison offline. Non-functional "offline

Table 6 Usability evaluation: Process and presentation of results (Functionality)

Presentation of Results (UP)	chiyo-co	Copy Content Detector	Docol©c	DupliChecker	OXSICO	Plagiarism Checker.co	Plagiarism Checker X	Plagiarism Detector.net	Small SEO Tools	StrikePlagiarism.com	Total
Conducting the tool is… (interface usage, uploading procedures, etc.)	1	1	0	2	2	2	2	2	2	2	16
Acquiring/receiving the reports is…	0	0	0	2	2	2	2	0	2	2	12
Understanding and verifying the reports is…	1	1	0	2	1	1	0	0	1	1	8
Downloading the results is …	0	0	0	2	2	2	2	2	2	2	14
Printing the results as a PDF is	0	0	0	2	2	2	2	2	2	2	14
Allowing side-by-side comparison online	2	2	2	0	2	0	2	0	1	2	13
Allowing side-by-side comparison offline	0	0	0	0	2	0	2	0	1	0	5
Average	0.6	0.6	0.3	1.4	1.9	1.3	1.7	0.9	1.6	1.6	
Total	4	4	2	10	13	9	12	6	11	11	7

comparison" means that tools either do not have downloadable, printable processes or their downloadable and printable reports cannot display Japanese characters.

4.3.1 Usability: Technical features

In order to position the tool in terms of usability, the second criteria group relating to UT is composed of 13 items, as listed below. The tool:

1) supports Shift-JIS
2) supports UTF-8
3) has classroom features
4) has a Japanese language interface
5) has alphabet/word precondition
6) allows multiple uploads
7) has no word limit
8) has no size limit
9) has API support
10) states related costs clearly
11) has call support on the webpage
12) has/uses an official e-mail account
13) substantially distinguishes false positive findings

A two-step measurement was employed to evaluate the technical features. In the first step, the criteria listed above were checked by the raters by logging into the tools.

Table 7 Usability evaluation: Technical features (UT)

Technical features	chiyo-co	Copy Content Detector	Docol©c	DupliChecker	OXSICO	Plagiarism Checker.co	Plagiarism Checker X	Plagiarism Detector.net	Small SEO Tools	StrikePlagiarism.com	Total
supports Shift-JIS	0	1	0	0	0	0	0	0	0	0	**1**
supports UTF-8	1	1	1	1	1	1	1	1	1	1	**10**
has classroom features	0	0	0	0	1	0	0	1	0	0	**2**
has a Japanese language interface	1	1	0	0	1	1	0	0	1	0	**5**
has alphabet/word precondition	1	1	1	0	1	0	0	0	0	1	**5**
allows multiple uploads	1	1	1	1	1	1	1	1	1	1	**10**
has no word limit	0	1	1	0	1	0	0	0	0	1	**4**
has no size limit	0	1	1	1	0	1	0	1	1	1	**7**
has API support	0	1	1	0	1	0	0	0	1	0	**4**
states related costs clearly	1	1	1	1	1	1	1	1	1	1	**10**
has call support on the webpage	0	0	0	0	0	0	1	0	0	1	**2**
has/uses an official e-mail account	0	0	1	0	1	1	1	1	1	1	**7**
distinguishes false positive findings	0	0	0	0	0	0	0	0	0	0	**0**
Sum	**5**	**9**	**8**	**4**	**9**	**6**	**5**	**6**	**7**	**8**	**13 \ 10**

Since this section only requires ‘Yes’ or ‘No’ as an answer, it was scored either as “1” if the criterion is met or “0” if it is not. In the second step, these results were sent to the tool vendors for their confirmation of technical results. Feedback from the tool vendors was addressed when necessary and the data set was finalized. The main purpose of this criteria group is to see how much of the tool has technical features to be able to cover the misconducted texts. Five criteria out of 13 are directly related to the Japanese language (Shift-JIS encoding, UTF-8 encodings, language interface, alphabet precondition, and false positive finding), and the rest are general (See Table 7).

With respect to Japanese language-related criteria, it turned out that almost none of the tools accepted/allowed uploading a file with Shift-JIS encoding that supports the Japanese language. Besides, five tools require ‘alphabet-based text’ as a precondition to upload a file. This precondition means that these tools set a minimum requirement of a certain word count, such as “minimum 30 words”. For an alphabet-based language and its user, this is just an issue of ‘number’. However, when it comes to an ideographic language, this prerequisite prevents uploading a text that is only in Japanese unless it meets the word count. Four of these tools (SmallSEOTools, Dupli Checker, Plagiarism Checker, and Plagiarism Detector) required at least 50 'words', and Plagiarism Checker X required at least 15 'words'. After we noted that this criterion was '0' for the tools given above, we created a 'sentence(s)' containing the requested number of words, which was: "*This is a trial text for Japanese. Since the tool requires words we input this sentence*". We then inserted them into the texts to meet the prerequisite word count. In this way, the prerequisite hurdle was overcome and uploads could be made to the tool. However, it should not be forgotten and strongly emphasized that even though it helps to clear the software hurdle, adding alphabet-based words to the full Japanese text that is required to be tested makes it highly possible that the similarity test results will be greatly affected. From the viewpoint of Japanese speakers, accessing the tool in their own language can also be counted as another beneficial feature, and half of the tools have a Japanese interface on their websites.

Additionally, from the pedagogical perspective, the feedback process is one of the most important elements to develop an understanding of academic integrity and prevent misconduct behaviors. Some tools allow instructors to use text-matching tools as part of learning management. Such tools allow instructors to assign homework, grade student works, and give feedback with the tool. Also, instructors can create virtual classrooms and manage the homework process via the text-matching tool. Regarding the learning environment, only two of the tools have classroom features, and four of them have API (Application Programming Interface).

To sum up, none of the tools was able to meet all the criteria. Of the Japanese-related criteria, only supporting UTF-8 encoding was fulfilled by all tools. Only five tools were able to meet more than half of the criteria-defined features. The features that are supported by all tools are multiple uploads and cost statement. The problematic features, except for the Japanese-related points, are not mentioning call support clearly and not sufficiently distinguishing false positive findings.

In regard to distinguishing the false positive findings, none of the tools gave a distinguished performance. This does not mean that the tools did not catch the 'similarities'. Each tool detected similarities at different rates. However, false positives refer to similarities that text-matching tools mark as a text match, but the matches are out of context. For instance, the tool might report a 50% text match in a text; however, this match may not always be counted as plagiarism and actual similarity. In such cases, those who are not trained to interpret the reports produced by text-matching tools may draw false conclusions. In terms of test results, the tests showed that this commonly happened in the Japanese language as well. In most cases, the similarities that the tools detected included mostly punctuations or numbers, and in some cases, matched the Japanese characters in the text without lexical or semantic grounds.

5 Discussion

The majority of previous research studies on testing text-matching tools has mainly focused on alphabetic languages. Investigating ideographic (non-alphabetic) languages in text-matching studies has been carried out far less than alphabetical language-matching tests (El Bachir Menai & Bagais, 2011; Kahloula & Berri, 2016; Nagoudi et al., 2018; Wu et al., 2021).

With respect to the Japanese language, studies on academic misconduct of Japanese speakers and users have started to increase over the past decades. Most of the academic integrity studies in the Japanese context focus on students' perceptions of academic misconduct issues and problems with academic writing (i.e., Kamimura, 2014; Teeter, 2014; Wheeler, 2009, 2014; Yamamoto, 2016; Yamamoto & Nitsū, 2015; Yamamoto et al., 2014; Yoshimura, 2015). Studies on detecting similarities in Japanese texts using text-matching tools are quite limited. The intersecting points of most of these studies are the target population and the materials/sources they focus on. The majority of the studies regarding text-matching tools focus on university students` writing assignments aiming to identify similarities based on words (syllables/characters) (Fukaya et al., 2003; Odaka et al., 2003) or sentences (Suzuki et al., 2009) to reach the plagiarized web source from the paraphrased texts (Takahashi et al., 2007), and developing detection tools (Ueta & Tominaga, 2010). Apart from these studies, Weber-Wulff's (2010) emphasis on the importance of encoding variables (i.e., JIS-Shift and UTF-8) in plagiarism detection, particularly for the Japanese language, apart from linguistic variables, is an important point that should be taken into account. However, in general, it seems that an inclusive model does not exist. As emphasized by Foltýnek et al. (2020) in the TeSToP Project, most studies approach coverage from only one perspective. Based on this, they generally applied the testing framework developed originally by Weber-Wulff (2010, 2013), which includes not only coverage but also a usability perspective.

5.1 Coverage testing in accordance with sources

As one of the four sources used in this study, Wikipedia seems to be the most common source among undergraduates as novice academic writers. It has free access and is easy to use, and this makes the platform even more attractive to students. Hence, it may be one of the first places to visit in case of academic misconduct behaviors. On the other hand, since it is online and free, it is expected that plagiarism will be detected more easily, which was confirmed in this study as well. Today, online and open access platforms are an indispensable part of academic studies for both students and researchers.

In terms of the Japanese language, both native and non-native Japanese language speakers and users use CiNii and J-Stage platforms for academic purposes apart from international services. CiNii (Scholarly and Academic Information Navigator of the National Institute of Informatics) is one of the main academic digital, online-open access databases. The J-Stage is another database that was developed and is managed by the Japan Science and Technology Agency (JST). For Japanese language students, especially those outside Japan, those platforms are the main gateway to academic knowledge written in Japanese. Therefore, these sources play a major role in testing whether text-matching tools can match the texts in Japanese.

In the specific case of this study, it is not possible to say that very promising results were obtained in terms of OAJP. Since the text-matching tools basically scan the digital databases, non-online publications such as books, documents, old periodicals, etc. are the most difficult types of the source to detect. This reality is likely

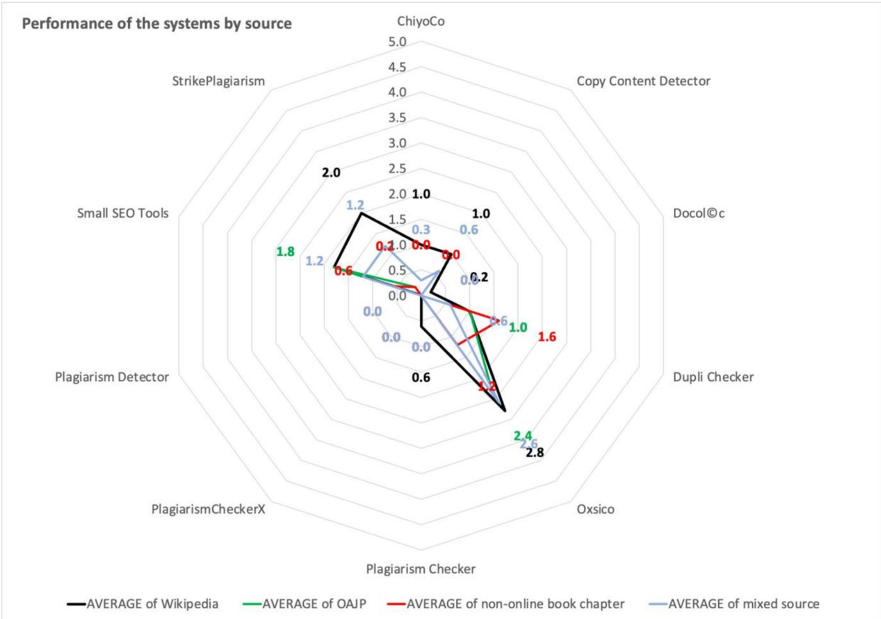

Fig. 2 Performance of tools by source

to be even more difficult for texts (books) written in Japanese and published only in Japan. Despite the momentum of digitalization, ‘paper culture’ is still strong in Japan. The paper culture here not only refers to the importance attached to traditional paper and writing culture, but also the fact that procedures are still carried out with old-fashioned paper as much as digital media. Depending on this kind of sociocultural habit, printed books, especially in the social sciences and humanities, are as much a part of academic life as digital studies. Locating sources may become even more difficult and complex when the text (book) was published, for example, in the 1940s.

Therefore, in order to interpret the performances of tools, non-online sources are also considered essential in this study. However, the text-matching tools tested in this study were found to perform the most inadequately with offline books, among all the sources. Lastly, as students show a tendency to disguise essays by mixing texts from different sources (patch-writing), we aimed to learn how much the tools can detect plagiarism in those texts as well. Five different (plagiarized/disguised) documents for each source were created and applied to the tools. The performance of each tool for various sources is illustrated in Fig. 2.

Oxsico, SmallSEOTools, and StrikePlagiarism stand out for identifying output from Wikipedia. As for the open access journal paper, Oxsico, Dupli Checker, and SmallSEOTools are above the average score. In terms of the non-online book chapter, apparently, DupliChecker and Oxsico are the tools that scored clearly above the average. Lastly, for the mixed source, Oxsico, SmallSEOTools and StrikePlagiarism have higher scores than the overall average score. Since the highest scores were seen

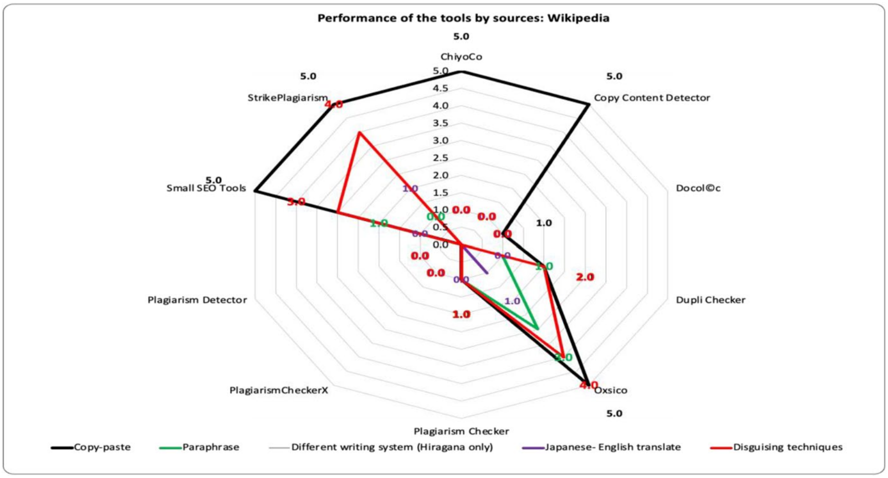

Fig. 3 Performance of tools in Wikipedia

in Wikipedia among these four sources, and moreover, Wikipedia was assumed to be a widely-used source in the original TeSToP study, it might be worth scrutinizing the performance of Wikipedia in detail (Fig. 3).

Even though Wikipedia is an online and fully open database, it was apparent that over 90% similarity was detected on only five tools. As for the automatically paraphrased texts, most tools fell short in detecting the similarities. The similarity rate in Oxsico was 68%, of which 50% referred to the original Wikipedia page. Quantitatively, it can be said that Oxsico is the most successful tool in this section with a score of 3.0 (51–70%/more than half). However, there were 31 other sources that matched the text created. Regarding this case, Weber-Wulff et al. (2013) highlight that this can lead to the appearance of many smallish text matches, instead of one large one. In particular, this can happen if the copy of the ever-changing Wikipedia in the database of the software tool is relatively old and the copies on the internet are from newer versions (Weber-Wulff et al., 2013). It is not easy to state whether the results in this study have the same (technical) background as Weber-Wulff et al. mentioned. However, it is also obvious that the performance for the Japanese language either on Wikipedia or other sources does not yet fully reflect the results and needs to be thoroughly reconsidered from various perspectives.

5.2 Coverage testing in accordance with plagiarism techniques

The second aspect of the coverage section is the analysis of data through documents that have been intentionally tampered with. Regarding the plagiarism methods, five documents were created using five different plagiarism techniques; copy-paste text, automatically paraphrased text, text written in a different writing system, translated text, and text with disguising techniques. In copy & paste documents, the overall average score is 1.5 out of 5.0. Except for two tools (Plagiarism CheckerX and Plagiarism Detector), all tools produced moderately acceptable results, given the fact

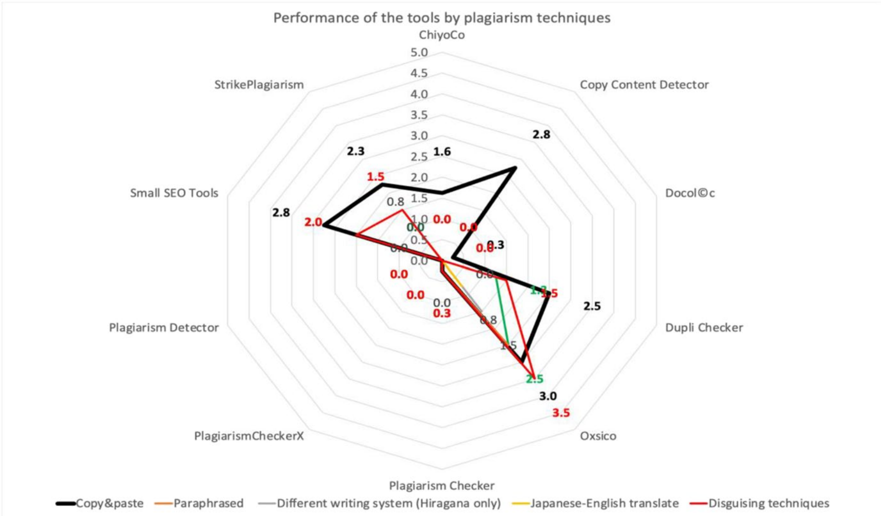

Fig. 4 Performance of the tools by plagiarism techniques

that those tools were not developed specially for the Japanese language. CopyContentDetector, built specifically for the Japanese language (2.8), DupliChecker (2.5), Oxsico (3.0), SmallSEOTools (2.8), and StrikePlagiarism (2.3) had the relatively highest performance in terms of copy & paste documents.

For the automatically paraphrased text, it is hard to conclude that there is a promising tendency of the results obtained in the test. The strongest score was accomplished by Oxsico with 3.5 out of 5.0, whereas SmallSEOTool (2.0) and DupliChecker (1.3) were the only tools with a score above the overall average of 0.7. The remaining eight tools had scores under the overall average of 0.7.

One other unique aspect of this test is evaluating the tools by changing the same content to another officially used writing system. In order to test it, the original texts were changed to the other writing system of Japanese, namely, Hiragana. As mentioned earlier, composing the text in only one writing system (i.e., Hiragana) is not a common practice. Any Japanese speaker without any special training can understand that something is wrong with such a text. However, the purpose of establishing such a criterion is to understand whether the transformation of the writing systems can be identified by the tools. Consequently, all tools failed to recognize the change in legitimate writing systems in Japanese. Quantitatively, only Oxsico showed a score of 1.5 (16% on Wikipedia, 39% on OAJP, 16% on the book chapter, and 37% on mixed-source). However, when the similarity reports were scrutinized in detail, no semantically or lexically meaningful matches were found (Fig. 4).

5.3 Usability

The main purpose of this part of the analysis is to see whether the text-matching tools are usable and useful for simple or basic users. Outputs earned from test criteria will be interpreted and discussed in two aspects from the following two subsections: How *functional* are the tools from the viewpoint of users? What kind of *technical features* do the tools have, and at which levels?

5.3.1 Usability: Process and presentation of outputs

5.4 How functional are the tools from the viewpoint of users?

Table 8 shows a summary of the results in which "average" refers to the score that the tool received whereas "performance" means the percentage of the average. For the process and presentation performance, six tools (OXSICO, Plagiarism Checker X, SmallSEOTools, StrikePlagiarism.com, Dupli Checker, and Plagiarism Checker.com) are above the overall average score. Among the criteria,'offline comparison' is the least supported function, while tool

Table 8 Usability: Process and presentation of outputs (UP)

Degree of functionality of tools for simple users		
Tool	Average Min. 0 Max. 2	Performance % of full score
OXSICO	1.9	92.9
Plagiarism Checker X	1.7	85.7
SmallSEOTools	1.6	78.6
StrikePlagiarism.com	1.6	78.6
Dupli Checker	1.4	71.4
Plagiarism Checker.com	1.3	64.3
Plagiarism Detector.net	0.9	42.9
chiyo-co	0.6	28.6
CopyContentDetector	0.6	28.6
Docol©c	0.3	14.3

conduction (interface usage, uploading procedures, etc.) was evaluated as the most functional. Apart from downloadability and printability issues, the inability to display Japanese characters and/or garbled characters (*mojibake*) was another issue in the test reports. Process and results presentations of Japanese tools (chiyo-co, CopyContentDetector) were as detailed as possible compared to non-Japanese tools. Similarities with different dimensions (lexical, semantic, etc.) were discussed in detail. However, the excessive details in the process and presentation made the tool extremely complicated in use and in understanding and verifying the outputs. Even though the raters had high capability in comprehending Japanese, there were times during tests when the raters could not find their way and felt lost in the tool. Moreover, the inability to download and print outputs is another point that should be noted. In this sense, it can be said that the tools built for the Japanese language could not perform as well as expected.

5.4.1 Usability: Technical features of the tool

What technical features do the tools have? In order to evaluate the performance of the text-matching tools, the adequacy of features is essential, both the usability functions and coverage ability. Tools can provide more effective output as long as they have inclusive and adequate technical features. Outputs can be misleading, particularly for inexperienced interpreters, unless they are clear or have adequate features (Razı, 2015). In the second aspect of evaluation, the adequacy of the tool was interpreted through 13 technical features, five of which are directly related to the Japanese language, while others are common features of all languages and users. With regard to the features directly related to the Japanese language, all tools are compatible with UTF-8 encoding, while only CopyContentDetector, which is a Japanese tool, allows uploading a file with the Shift-JIS extension. No significant performance could be identified in terms of distinguishing false positive findings, as mentioned in *Usability: Technical features*. Some tools require a certain number of alphabet-based words as a precondition to upload the text for testing. This may affect the numbers/ratio of the output, which eventually forms the interpretation of one who is inexperienced in reading the similarity reports. Therefore, given the circumstances, it is possible to conclude that the technical performance of tools directly related to the Japanese language is very limited. On the other hand, from the perspective of the overall scores, it is a fact that OXSICO, Plagiarism Checker.com, Docol@c, and Dupli Checker showed a relatively higher performance compared to others in terms of technical features (Table 9).

The overall approach to tool performances: Integrated evaluation of scores Since not only numbers (coverage scores) but also the qualitative approach (two-aspect usability) is important to see the tool's performance, and interpret the outputs of the reports the tool provides correspondingly, an integrated approach was applied here. Coverage and usability results were

Table 9 Usability: Technical features of tool (UT)

Tool	Availability of technical features (N) Sum Min. 0 Max. 13	Performance % of full score
OXSICO	9	69.2
Plagiarism Checker.co	9	69.2
Docol©c	8	61.5
Dupli Checker	8	61.5
Plagiarism Detector.net	6	46.2
Plagiarism Checker X	6	46.2
chiyo-co	6	46.2
SmallSEOTools	5	38.5
CopyContentDetector	4	30.8
StrikePlagiarism.com	4	30.8

combined into a two-dimensional graph (Fig. 5). Coverage in the X-axis in Fig. 5 is the percentage of the average scores of all documents tested in the four types of sources. Usability in Y-axis in Fig. 5 is the sum of the percentage of the average scores of UP and UT.

Considering the overall assessment, the tools, in general, give a relatively higher performance on the usability side rather than the coverage aspect. Most tools have very limited coverage performance in the Japanese language.

Once again, it should be noted that as these tools were not designed specifically for the Japanese language, such results are expected. Yet, the results presented in this paper will provide some guidance for vendors to meet the needs of ideographic language users, such as Japanese.

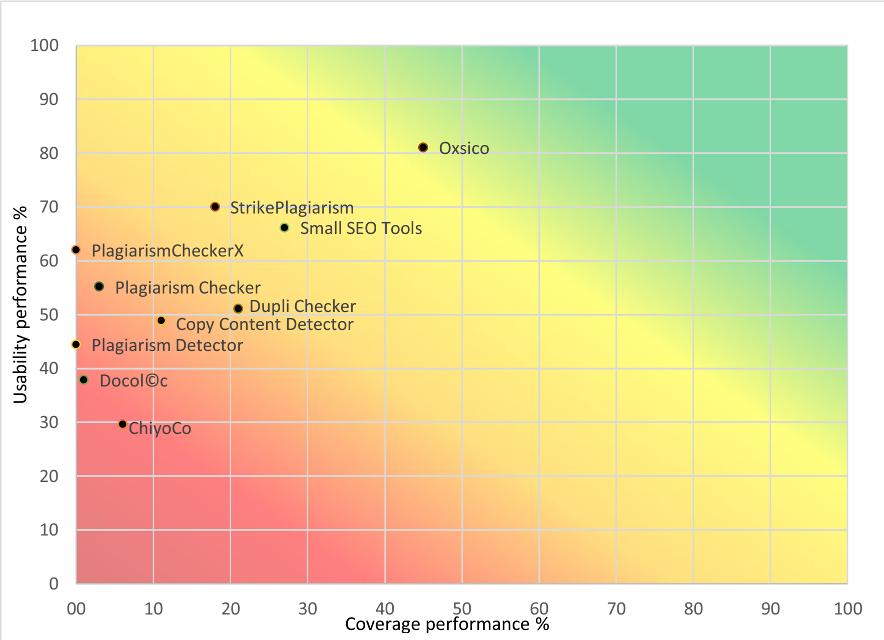

Fig. 5 Coverage and Usability combined. X-axis: Average percentage of total coverage scores (A + B + C + D); Y-axis: Average percentage of total usability scores (UP + UT)

6 Conclusions and recommendations

This is the first study in which the Japanese language has been extensively tested in terms of the performance of text matching tools. From the formulation of the research question to the analysis process, a detailed, multidimensional and well-organized methodology was employed for almost a year. Our recommendations are expected to be helpful for vendors to improve their algorithms for checking similarities in the Japanese language, and educators to identify the barriers involved in the performance of text-matching tools in the Japanese language.

Before providing recommendations, here we would like to highlight an important point. In this paper, we hypothesize that text-matching tools might perform worse in ideographic languages (in this paper, Japanese)compared to alphabetic languages. A performance decrease in Japanese is understandable because no text-matching tools have been developed based on the Japanese writing systems. However, these tools are used for Japanese text-matching purposes on some occasions, in which similarity reports may play an important role, such as a prerequisite for the approval of a postgraduate thesis/dissertation. Therefore, for the student who submits a thesis in Asian languages (in this case, Japanese), non-Japanese text-matching tools are mostly the only option to provide a similarity report. Within this scope, the current text-matching tools tested in this paper are expected to be limited in their performance on Japanese texts. For this reason, the approach of this paper toward the results will be more revealing rather than labeling their performance as good or bad.

A brief summary of the most ***important findings*** **is as follows**:

a. The text-matching tools are often not effective for the Japanese language.
b. In terms of coverage, results vary considerably across sources. Wikipedia is the most detectable among these sources, while OAJP, offline books and a mixed text have rather ineffective results.
c. Most of the matches, in the mixed source particularly, are mostly either false positives or semantically and lexically insignificant matches.
d. As for the performance of different text types, copied and pasted texts are those with which the tools mostly produce relatively satisfactory results.
e. Although the text types occasionally show quantitatively high scores, qualitative analysis shows that these matches are mostly either false positives or semantically and lexically meaningless matches. This is particularly noticeable in the text transformed to Hiragana and the tampered text.
f. Photographic images could not be identified in the tampered texts. Disguises in punctuation marks and numbers were rarely detected, and no pattern could be identified. Therefore, the detection performance of plagiarism in tampered texts is quite low.

Recommendations **for improvement of the tool** for Japanese (and possibly other ideographic) languages are to:

a. Solve the encoding problem: Shift-JIS based texts cannot be uploaded to the tool, let alone scanned.
b. Remove the lower WORD limit as a prerequisite for text uploading: If Japanese and other ideographic languages are to be added to the language pool of text-matching tools, they should remove the lower minimum word count prerequisite. Otherwise, the following proposals and improvements would be meaningless.
c. Expand the resource pool: Expanding the web-based sources from which scans are made will increase the coverage potential of the tool.
d. Improve the Japanese interface for reporting and documentation: With a Japanese interface, it will be possible to evaluate the matching results more effectively.
e. Add a hand-drawing-writing function to the learning environments for the feedback phase: In this way, more effective feedback can be provided in ideographic languages.
f. Make the outputs and reports readable and comparable offline: This is a feature that needs to be improved, not only for Japanese but in general.

Lessons learned from this study from a ***pedagogical*** **perspective (for the Japanese language teachers)** are:

a. It should be strongly emphasized that the text-matching tools, in general, can find text overlaps and similarities, but not plagiarism. Similarity and overlap do not always indicate plagiarism. Therefore, evaluators should be wary of judging plagiarism based only on quantitative results (similarity rates) in reports. This fact was confirmed once again in this study. Such decisions are sometimes life-changing for students considering that they may be sanctioned as the result of a wrong decision of the evaluator, and labelled as a plagiariser. Therefore, evaluators should have enough knowledge and be experienced in how to interpret the reports. Such improvements can be made in the short and long term in the following ways. In the short term, providing essential information about text-matching services to currently active Japanese language teachers may help. This can be done in two ways, either by the introduction of tools or texts in Japanese that were developed by the tools or their stakeholders on their Japanese interfaces, or by face-to-face or online workshops of institutions where Japanese is taught. In the long term, including courses on how to read and interpret text-matching reports in the program of teacher training faculties will also ensure that future teachers are better equipped with these skills. In fact, a concrete step in this regard has already been taken by Çanakkale Onsekiz Mart University, Türkiye. For the past two years, a course called "Avoiding Plagiarism in Academic Writing" has been taught by an expert faculty member in the curriculum of the Japanese language education master's program.
b. It is necessary to pay attention to the matched sources highlighted in the tool reports. This study confirmed that a source spotlighted as the source of plagiarism may actually be an irrelevant one.
c. The least detected text types were the texts written in only Hiragana, and the non-online book chapter. Teachers or evaluators should also pay attention to the details of the references used in student reports. Moreover, they should also take into consideration writing the original text in a different writing system. Hiragana or Katakana can be used for possible acts of misconduct other than its ordinary functions.

Lastly, by taking into account all previous works discussed in this paper and the results learned in this study, we can summarize the desirable future developments as a multi-layered roadmap regarding text-matching services for the Japanese language, and for all stakeholders from short-term to long-term, as follows (Table 10):

Taking into consideration the number of anti-plagiarism tools to be tested, the variety and number of testing documents, this study is the most inclusive test on the Japanese language ever carried out. With the results obtained from this study, it is expected to contribute to all stakeholders, such as vendors, faculty members, and decision makers in educational institutions. More importantly, we hope that this work, with its results, will be a source of inspiration for other researchers of ideographic and/or Asian languages.

Table 10 A roadmap for all stakeholders

	Short term	Medium term	Long term
Vendors	Developing new algorithms responsive to Japanese text-matching	R&D on Asian languages (in terms of both technically and linguistics)	Extending the experience gained in Asian languages to ideographic languages
Policy makers	Developing departmental, institutional policies and regulations specific to the Japanese language (e.g., evaluation and assessment criteria, etc.)	Developing national policies and regulations specific to the Japanese language, and adapting those to the National Qualifications Framework (NQF)	Establishing a common international framework on text-matching interpretation and evaluation as EQF by including international associations related to the Japanese language such as the Association of Japanese Language Teachers in Europe (EAJE), American Association of Teachers of Japanese (AATJ), Association for Japanese Language Education (NKG), Association for Japanese Language Teaching (AJALT)
Educators	Awareness by dissemination activities such as seminars and workshops for pre-service teachers	Directly related courses on teacher training programs (at MA, PhD levels)	International teaching & training events conducted by the above-mentioned organizations both for students, early career educators and Japanologists

Appendix 1

Details of sources used

	Type of text	Source of text	Link of text (if available), details (if any)	Date
A1	Original text	Wikipedia	https://ja.wikipedia.org/wiki/%E6%97%A5%E6%9C%AC%E3%81%AE%E9%AB%98%E9%BD%A2%E5%8C%96	17.01.2022
A2	Paraphrased text (automatically)	Web tool	https://www.paraphraser.io/ja/paraphrasing-tool	17.01.2022
A3	Paraphrased text (manually) (changed kana systems)	Manual	Applied by Senem Çente Akkan	17.01.2022
A4	Japanese-English Translation	Google translate	https://translate.google.com/	17.01.2022
A5	Disguising techniques	Manual	Numbering styles, OCR, punctuation, white characters, etc. are applied	17.01.2022
B1	Original paper (online & open access database)	J-Stage	https://www.jstage.jst.go.jp/article/jtje/21/0/21_3/_article/-char/ja Özşen, T. (2019).「日本学研究における日本語学習の意味と課題」[The meaning and issues of Japanese Learning in Japanology Studies],『専門日本語教育研究』[Journal of Technical Japanese Education], No.21:3–9, ISSN: 1345–1995	18.01.2022
B2	Paraphrased text (automatically)	Web tool	https://www.paraphraser.io/ja/paraphrasing-tool	18.01.2022
B3	Paraphrased text (manually) (changed kana systems)	Manual	Checked by Senem Çente Akkan	18.01.2022
B4	Japanese-English Translation	Google translate	https://translate.google.com/	18.01.2022
B5	Disguising techniques	Manual	Numbering styles, OCR, punctuation, white characters, etc. are applied	18.01.2022

	Type of text	Source of text	Link of text (if available), details (if any)	Date
C1	Original paper (Non-online, unpublished on the internet)	Book chapter	Özşen, T. (2015) 「生活構造論的視点からトルコの農村を読み直す」(Translation: Rereading the Turkish Rural Community from the viewpoint of Life Structure) 、 in Tokuno S., Makino A., Matsumoto T. (eds) 、『暮らしの視点からの地方再生—地域と生活の社会学』(Sociology of Community and Life) Kyushu University Press; 139–162	18.01.2022
C2	Paraphrased text (automatically)	Web tool	https://www.paraphraser.io/ja/paraphrasing-tool	28.03.2022
C3	Paraphrased text (manually) (changed kana systems)	Manual	Checked by Senem Çente Akkan	18.01.2022
C4	Japanese-English Translation	Google translate	https://translate.google.com/	18.01.2022
C5	Disguising techniques	Manual	Numbering styles, OCR, punctuation, white characters, etc. are applied	28.03.2022
D1	Multi-source text (Wikipedia, government white papers, OA journal paper)	Wikipedia, Japan Foundation webpage, CiNii for journal paper	Government White paper (Japan Foundation): https://www.jpf.go.jp/j/project/japanese/survey/result/dl/survey2018/text.pdf Wikipedia: https://ja.wikipedia.org/wiki/%E6%97%A5%E6%9C%AC%E8%AA%9E%E6%95%99%E8%82%B2#%E6%97%A5%E6%9C%AC%E8%AA%9E%E6%95%99%E8%82%B2%E3%81%AE%E6%AD%B4%E5%8F%B2 Online OA Journal Paper: https://ci.nii.ac.jp/naid/110009687716 1st paragraph from Wikipedia, 2nd and 3rd paragraphs are from Japan Foundation The last paragraph is from the Journal paper	18.01.2022
D2	Paraphrased text (automatically)	Web tool	https://www.paraphraser.io/ja/paraphrasing-tool	19.01.2022

	Type of text	Source of text	Link of text (if available), details (if any)	Date
D3	Paraphrased text (manually) (changed kana systems)	Manual	Checked by Senem Çente Akkan	18.01.2022
D4	Japanese-English Translation	Google translate	https://translate.google.com/	18.01.2022
D5	Disguising techniques	Manual	Numbering styles, OCR, punctuation, white characters, etc. are applied	18.01.2022

Appendix 2

Main contact URLs for the 10 text-matching tools evaluated in this paper

chiyo-co
CopyContentDetector
Docol©c
Dupli Checker
OXSICO
Plagiarism Checker.co
Plagiarism Checker X
Plagiarism Detector.net
SmallSEOTools
StrikePlagiarism.com

Acknowledgements We gratefully thank the ENAI TeSToP Working Group members for inspiring this work with the original TeSToP and allowing us to use the method; Debora Weber-Wulff for enabling us to use the methodology she developed, and for sharing her time to discuss encoding issues in Japanese from the software viewpoint; and Tomáš Foltýnek on behalf of the ENAI for letting us use their official e-mail account to communicate with the tools. We also would like to express our sincere thanks to Graham Lee for proofreading the manuscript.

Authors' contributions In light of the workload types given below, authors' contributions are as follow.

Work/contribution types:

1. managing the project
2. creating/establishing the evaluation framework
3. idea support
4. theoretical contribution and/or guidance
5. secretariat/communicating with tools
6. data input
7. data interpretation
8. designing/developing images & graphs & tables
9. chapter/section writing (if yes, which part)
10. improving the language
11. mentorship

TÖ: 1, 2, 3, 5, 6, 7, 8, 9 (entire manuscript), 11
İS: 5, 6, 7, 8
ÖÇ: 2, 3, 4, 7, 9 (section of "previous test"), 10
SR: 2, 3, 4, 7, 10, 11
SÇA: 6,7, 9 (partially contributed to "previous test" section)
DD: 3, 7, 11

Data availability Data and materials generated, used, and analyzed in this study are publicly available in the European Network for Academic Integrity (ENAI) repository. https://www.academicintegrity.eu/wp/testing-of-support-tools-for-plagiarism-detection-working-group/

Declarations

Competing interests Several authors of this article are involved in organizing the European Network for Academic Integrity annual conferences. The Academic Integrity Ph.D. Summer Schools receive funding from text-matching software vendors. This did not influence our research in any phase.

References

Abdelhamid, M., Azouaou, F., & Batata, S. (2022). A survey of plagiarism detection systems: Case of use with English, French and Arabic Languages. *arXiv preprint* arXiv:2201.03423.

Ahmed, R. K. A. (2015). Overview of different plagiarism detection tools. *International Journal of Futuristic Trends in Engineering and Technology, 2*(10), 1–3.

Ali, A. M. E. T., Abdulla, H. M. D., & Snasel, V. (2011, May 13). *Overview and comparison of plagiarism detection tools* [Poster presentation]. Annual International Workshop on DAtabases, TExts, Specifications and Objects (Dateso), 161–172, Pisek, Czech Republic. https://ceur-ws.org/Vol-706/poster22.pdf

Alotaibi, N., & Joy, M. (2021). English-Arabic cross-language plagiarism detection. *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)* (pp. 44–52). https://doi.org/10.26615/978-954-452-072-4_006

Birkić, T., Celjak, D., Cundeković, M., & Rako, S. (2016). *Analysis of software for plagiarism detection in science and education*. University of Zagreb. Retrieved January, 2022, from https://www.srce.unizg.hr/files/srce/docs/CEU/analysis_of_software_for_plagiarism_detection_in_science_and_education.pdf

Bull, J., Collins, C., Coughlin, E., Sharp, D., & Square, P. (2001). *Technical review of plagiarism detection software report*. University of Luton.

Chen, S., & Macfarlane, B. (2016). Academic integrity in China. In T. Bretag (Ed.), *Handbook of Academic Integrity* (pp. 99–105). Springer Reference.

Chowdhury, H. A., & Bhattacharyya, D. K. (2018). Plagiarism: Taxonomy, tools and detection techniques. arXiv:1801.06323. https://doi.org/10.48550/arXiv.1801.06323

Condurache, I.-A., & Bolboacă, S. D. (2022). Comparison of plagiarism detection performance between some commercial and free software. *Applied Medical Informatics, 44*(2), 73–86. https://ami.info.umfcluj.ro/index.php/AMI/article/view/904

Ehrich, J., Howard, S. J., Mu, C., & Bokosmaty, S. (2016). A comparison of Chinese and Australian university students' attitudes towards plagiarism. *Studies in Higher Education, 41*(2), 231–246.

El Bachir Menai, M., & Bagais, M. (2011). APlag: A plagiarism checker for Arabic texts. *2011 6th International Conference on Computer Science & Education (ICCSE)* (pp.1379–1383). IEEE. https://doi.org/10.1109/ICCSE.2011.6028888

Elamine, M., Mechti, S., & Belguith, L. H. (2020). Hybrid plagiarism detection method for French language. *International Journal of Hybrid Intelligent Systems, 16*(3), 163–175. https://doi.org/10.3233/HIS-200284

Elamine, M., Bougares, F., Mechti, S., & HadrichBelguith, L. (2021). Extrinsic plagiarism detection for French language with Word embeddings. In A. Abraham, P. Siarry, K. Ma, & A. Kaklauskas

(Eds.), *Intelligent systems design and applications* (pp. 217–224). Springer International Publishing. https://doi.org/10.1007/978-3-030-49342-4_21

Elkhatat, A. M., Elsaid, K., & Almeer, S. (2021). Some students plagiarism tricks, and tips for effective check. *International Journal for Education Integrity, 17*(1), 1–12. https://doi.org/10.1007/s40979-021-00082-w

Foltýnek, T., Dlabolová, D., Anohina-Naumeca, A., Razı, S., Kravjar, J., Kamzola, L., Guerrero-Dib, J., Çelik, Ö., & Weber-Wulff, D. (2020). Testing of support tools for plagiarism detection. *International Journal of Educational Technology in Higher Education, 17*(1), 1–31. https://doi.org/10.1186/s41239-020-00192-4

Fukaya, R., Yamamura, T., Kudō, H., Matsumoto, T., Takeuchi, Y., & Ohnishi, N. (2003). Hindo tōkei to gainen jisho wo mochiita bunshō no ruijisei no teiryōka [Measuring similarity between documents using term frequency and concept dictionary]. *Jōhō Shori Gakkai Kenkyū Hōkoku [IPSJ SIG Notes], 153*, 73–79.

Howard, R. M. (1992). A plagiarism pentimento. *Journal of Teaching Writing, 11*(2), 233–245.

Hussein, A. S. (2015). A plagiarism detection system for Arabic documents. In D. Filev, J. Jabłkowski, J. Kacprzyk, M. Krawczak, I. Popchev, L. Rutkowski, V. Sgurev, E. Sotirova, P. Szynkarczyk, & S. Zadrozny (Eds.), *Intelligent Systems'2014* (pp. 541–552). Springer International Publishing. https://doi.org/10.1007/978-3-319-11310-4_47

Jadhav Sunayana, D., & Lihitkar Shalini, R. (2021). Plagiarism detection software: A comparative evaluation. *Library Philosophy and Practice (e-journal)*, 6206. https://digitalcommons.unl.edu/libphilprac/6206/

Japan Foundation. (2018). *Survey report on Japanese-Language education abroad 2018*. Retrieved May, 2022, from https://www.jpf.go.jp/j/project/japanese/survey/result/dl/survey2018/text.pdf

Japan Statistical Yearbook. (2022). *Chapter 25 Education*. Retrieved September, 2022, from https://www.stat.go.jp/english/data/nenkan/71nenkan/1431-25.html

Kahloula, B., & Berri, J. (2016). Plagiarism detection in arabic documents: Approaches, architecture and systems. *Journal of Digital Information Management, 14*(2), 124–135.

Kakkonen, T., & Mozgovoy, M. (2010). Hermetic and web plagiarism detection systems for student essays—An evaluation of the State-of-the-Art. *Journal of Educational Computing Research, 42*(2), 135–159. https://doi.org/10.2190/EC.42.2.a

Kamimura, T. (2014). Citation behaviors observed in Japanese EFL students' argumentative writing. *Journal of Pan-Pacific Association of Applied Linguistics, 18*(1), 85–101.

Kier, C.A., & Ives, C. (2022) Recommendations for a balanced approach to supporting academic integrity: perspectives from a survey of students, faculty, and tutors. *International Journal for Educational Integrity, 18*(22). https://doi.org/10.1007/s40979-022-00116-x

Krizkova, S., Tomaskova, H., & Gavalec, M. (2016). Preference comparison for plagiarism detection systems. *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp.1760–1767). IEEE. https://doi.org/10.1109/FUZZ-IEEE.2016.7737903

Kulkarni, S., Govilkar, S., & Amin, D. (2021). Analysis of plagiarism detection tools and methods. *Proceedings of the 4th International Conference on Advances in Science & Technology (ICAST2021). International Conference on Advances in Science & Technology,* Bahir Dar, Ethiopia. https://doi.org/10.2139/ssrn.3869091

Lancaster, T., & Culwin, F. (2005). Classifications of plagiarism detection engines. *Innovation in Teaching and Learning in Information and Computer Sciences, 4*(2), 1–16. https://doi.org/10.11120/ital.2005.04020006

Masic, I. (2012). Plagiarism in scientific publishing. *Acta Informatica Medica, 20*(4), 208–213. https://doi.org/10.5455/aim.2012.20.208-213

Maurer, H., Kappe, F., & Zaka, B. (2006). Plagiarism—A survey. *Journal of Universal Computer Science, 12*(8), 1050–1084. https://doi.org/10.3217/JUCS-012-08-1050

McGowan, U. (2005). Plagiarism detection and prevention: Are we putting the cart before the horse. In A. Brew & C. Asmar (Eds), *Proceedings of Higher Education Research and Development Society of Australasia (HERDSA) conference* (pp. 287–293). HERDSA. https://www.herdsa.org.au/publications/conference-proceedings/research-and-development-higher-education-higher-education-82

Mostofa, S. K. M., Tabassum, M., & Ahmed, S. M. Z. (2021). Researchers' awareness about plagiarism and impact of plagiarism detection tools – does awareness effect the actions towards preventing plagiarism? *Digital Library Perspectives, 37*(3), 257–274. https://doi.org/10.1108/DLP-10-2020-0100

Nadhri, S., Elamine, M., & Belguith, L. H. (2021, December). *Automatic evaluation of existing plagiarism detection tools* [Paper presentation]. In *Tunisian Algerian Conference on Applied Computing (TACC)*. Tabarka, Tunisia: TACC. https://ceur-ws.org/Vol-3067/paper17.pdf

Nagoudi, E. M. B., Khorsi, A., Cherroun, H., & Schwab, D. (2018). A two-level plagiarism detection system for Arabic documents. *Cybernetics and Information Technologies, 18*(1), 124–138.

Nahas, M. N. (2017). Survey and comparison between plagiarism detection tools. *American Journal of Data Mining and Knowledge Discovery, 2*(2), 50–53. https://doi.org/10.11648/j.ajdmkd.20170202.12

Naik, R. R., Landge, M. B., & Mahender, C. N. (2015). A review on plagiarism detection tools. *International Journal of Computer Applications, 125*(11), 16–22.

Odaka, T., Murata, T., Gao, J., Suwa, I., Kuroiwa, J., & Ogura, H. (2003). n-gram wo mochiita gakusei repooto hyōkashuhōnoteisatsu [A proposal on student report scoring system using n-gram text analysis method]. *IEICE, J86-D-1*(9), 702–705.

Perkins, M., Gezgin, U. B., & Roe, J. (2020). Reducing plagiarism through academic misconduct education. *International Journal for Educational Integrity, 16*(1), 1–15.

Pertile, S. de L., Moreira, V. P., & Rosso, P. (2016). Comparing and combining Content- and Citation-based approaches for plagiarism detection. *Journal of the Association for Information Science and Technology, 67*(10), 2511–2526. https://doi.org/10.1002/asi.23593

Razı, S. (2015). Development of a rubric to assess academic writing incorporating plagiarism detectors. *SAGE Open, 5*(2), 1–13. https://doi.org/10.1177/2158244015590162

Suzuki, K., Takahashi, I., Shirai, H., Kuroiwa, J., Odaka, T., & Ogura, H. (2009). Hyōsetsu repooto hakkenn ni riyō suru 1 buntan`I de no kensaku kuerisakuseishushō [Web search query in detecting plagiarism reports]. *IEICE, 92*(11), 2072–2076.

Takahashi, I., Miyakawa, K., Odaka, T., Shirai, H., Kuroiwa, J., & Ogura, H. (2007). Web saito karano hyōsetsurepooto hakken shien shisutemu [A Computer Aided Detection System for Learners` Reports Plagiarism from Web-site]. *IEICE, 90*(11), 2989–2999.

Teeter, J. (2014). Deconstructing attitudes towards plagiarism of Japanese undergraduates in EFL academic writing classes. *English Language Teaching, 8*(1), 95–109.

Ueno, S., Takahashi, I., Kuroiwa, J., Shirai, H., Odaka, T., & Ogura, H. (2006). Fukusuu no web peeji kara hyōsetsu shita repooto no hakken shien shisutemu no jissō [Implementation of a support system to find out of the report plagiarized from several web pages]. *JŌhō Shori Gakkai Kenkyū Hōkoku [IPSJ SIG Notes], 87*, 41–46.

Ueta, K., & Tominaga, H. (2010). A development and application of similarity detection methods for plagiarism of online reports. *Proceedings of ITHET, 2010*, 363–371.

Vani, K., & Gupta, D. (2016). Study on extrinsic text plagiarism detection techniques and tools. *Journal of Engineering Science and Technology Review, 9*(4), 150–164. https://doi.org/10.25103/jestr.094.23

Vrbanec, T., & Meštrović, A. (2021). Corpus-based paraphrase detection experiments and review. *Information, 11*(5), 241.

Wahle, J. P., Ruas, T., Foltýnek, T., Meuschke, N., & Gipp, B. (2022). Identifying machine-paraphrased plagiarism. In M. Smits (Ed.), *Information for a better World: Shaping the Global future* (pp. 393–413). Springer International Publishing. https://doi.org/10.1007/978-3-030-96957-8_34

Weber-Wulff, D., Möller, C., Touras, J., & Zincke, E. (2013). *Plagiarism detection software test 2013.* Retrieved March, 2022, from http://plagiat.htw-berlin.de/software/2013/

Weber-Wulff, D. (2010). *Plagiarism detection test 2010.* Retrieved March, 2022, from https://plagiat.htw-berlin.de/software-en/2010-2/

Wheeler, G. (2009). Plagiarism in the Japanese universities: Truly a cultural matter? *Journal of Second Language Writing, 18*(1), 17–29.

Wheeler, G. (2014). Culture of minimal influence: A study of Japanese university students' attitudes toward plagiarism. *The International Journal for Educational Integrity, 10*(2), 44–59.

Wheeler, G. (2016). Perspectives from Japan. In T. Bretag (Ed.), *Handbook of Academic Integrity* (pp. 107–112). Springer Singapore. https://doi.org/10.1007/978-981-287-098-8_7

Wu, Z., Liang, J., Zhang, Z., & Lei, J. (2021). Exploration of text matching methods in Chinese disease Q&A systems: A method using ensemble based on BERT and boosted tree models. *Journal of Biomedical Informatics, 115*, 1–10. https://doi.org/10.1016/j.jbi.2021.103683

Yamamoto, F. (2016). Ronbun no 'itoteki dehanai hyōsetsu' no mondai: Modaritii no kondō to kaishaku no nai in`yō [Unintentional plagiarism in Japanese writing: Confusion of modalities and citation without interpretation]. *Global Communication, 6*, 117–132.

Yamamoto, F., & Nitsū, N. (2015). Ronbun no in`yō – kōzō: Jinbun & shakaikagakukei ronbun shidō no tame no kisoteki kenkyū [Quotation and interpretation structure of literature-analysis papers: Basic research on instruction for writing papers in humanities and social science]. *Nihongo Kyōiku [journal of Japanese Language Teaching], 160*, 94–109.

Yamamoto, F., Nitsū, N., Ohshima, Y., & Satō, S. (2014). In`yō kara kaishaku ni itaru in`yōbun no tayōsei [Varieties of citations from quoting to interpreting in the "literature-analysis-type" papers in humanity and social science]. *Dai 16 kai Senmon Nihongo Kyōiku Gakkai Ronshū [Conference proceeding of 16th conference of the society for technical Japanese education]*, 2–23.

Yoshimura, F. (2015). Japanese university students' experience with and perceptions of citations in academic writing. *Journal of Institute for Research in English Language and Literature, 40*(1), 37–62.

Authors and Affiliations

Tolga Özşen[1] · İrem Saka[2] · Özgür Çelik[3] · Salim Razı[4] · Senem Çente Akkan[5] · Dita Henek Dlabolova[6]

İrem Saka
sakaairemm@gmail.com

Özgür Çelik
ozgurcelik911@gmail.com

Salim Razı
salimrazi@gmail.com

Senem Çente Akkan
senemcente@gmail.com

Dita Henek Dlabolova
dita.dlabolova@gmail.com

1 Department of Japanese Language Teaching, Çanakkale Onsekiz Mart University, Çanakkale, Türkiye

2 Japanese Language Teaching Graduate Program Student, Çanakkale Onsekiz Mart University, Çanakkale, Türkiye

3 School of Foreign Languages, Balıkesir University, Balıkesir, Türkiye

4 Center for Academic Integrity, Çanakkale Onsekiz Mart University, Çanakkale, Türkiye

5 Graduate of Japanese Language Teaching Master's Program, Çanakkale Onsekiz Mart University, Çanakkale, Türkiye

6 European Network for Academic Integrity, Brno, Czech Republic