

## CrossCheck usage in a journal publication

Jaeseok Choi<sup>1</sup>, Sohee Park<sup>2</sup>, Ungjin Oh<sup>1</sup>

<sup>1</sup>Department of Electrical Engineering, Gyeongsang National University, Jinju; <sup>2</sup>Korean Institute of Electrical Engineers, Seoul, Korea

### Abstract

Since the inclusion of the Journal of Electrical Engineering and Technology (JEET) published by the Korean Institute of Electrical Engineers in the Science Citation Index Expanded on the Web of Science by Thomson Reuters, the journal has recorded a considerable increase in the number of submitted articles (i.e., from 400 articles in 2009 to 2,000 articles in 2015). This work explores the use of CrossCheck as a tool to prevent and provide protection against plagiarism in the JEET. Since 2011, the JEET has been using CrossCheck and has adopted implicit and latent review guidelines internally. In this study, we investigate the function of CrossCheck by considering two types of similarity levels for published and rejected articles, namely, integrated similarity index (ISI) and maximum similarity index (MSI). The Minitab tool is used for statistical analysis. The JEET employs a blind CrossCheck system, in which ISI and MSI information is supplied only to the associate editor and not to the reviewers. Positive results are obtained even under the blind CrossCheck system. An exception is the group of “red” articles with ISI and MSI scores of above 50%. The ISI and MSI information of such red articles is supplied to the editors and reviewers of the JEET. The results of this work could serve as a reference for establishing a guideline or criterion for rejecting suspicious plagiarized articles during the review process.

### Keywords

Article; CrossCheck; Plagiarism; Similarity index; Plagiarism guideline

### Introduction

In recent years, a similarity checking technology based on a large data linked web system has been created to find and cross-check plagiarism in articles (also called “papers” for convenience in this study) that are published or reviewed. CrossCheck is an available, useful, and valuable tool not only for checking plagiarism but also for preventing and protecting from attempts of plagiarism in advance. This study presents an actual experience of the CrossCheck tool for the Plagiarism Detection Service System.

CrossCheck was created on October 2008 by iParadigm [1,2]. The service has been provided to CrossRef members and the tool is called iThenticate since 2008. In recent, the CrossCheck

**Received:** January 18, 2016

**Accepted:** February 3, 2016

**Correspondence to** Jaeseok Choi  
jschoi@gnu.ac.kr

#### ORCID

Jaeseok Choi

<http://orcid.org/0000-0003-0867-6251>

Sohee Park

<http://orcid.org/0000-0001-9110-6254>

Ungjin Oh

<http://orcid.org/0000-0001-9105-1791>

This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

service system is developed into DB plus iThenticate tools. The main output of the system is the similarity index (SI), which consists of integrated SI (ISI) and maximum SI (MSI). The SI does not necessarily mean plagiarism index exactly. CrossCheck only helps to protect the original authors' copyrights and to improve authors' behavior by identifying instances of academic plagiarism. The CrossCheck tool can provide a reliable journal by comparing texts. However, this tool is unavailable for crosschecking figures, tables, and equations and is available only for checking text. Therefore, this tool is not perfect; nonetheless, it is very informative in terms of preventing and protecting from authors committing plagiarism.

This study presents the experiences of preventing and protecting against plagiarism using CrossCheck on articles submitted to Journal of Electrical Engineering and Technology (JEET) over the past three years. The first issue of JEET is published on March 1, 2006 by Korean Institute of Electrical Engineers (KIEE). Since the JEET has been registered as SCIE (Science Citation Index Expanded) on the Web of Science by Thomson Reuters in 2009, the number of submitted articles has increased considerably from 400 articles in 2009 to 1,600 articles in 2014 and 2,000 articles in 2015. In this study, two types of similarity levels, i.e., ISI and MSI, are used for published and rejected articles using CrossCheck.

Based on the CrossCheck statistical analysis on the relationship of similarity indices of articles submitted to JEET since late 2011, the JEET editorial board has made a basic and internal guideline and three categories (normal, warning, and red) depending on article similarity scores. The DB of similarity indices in two viewpoints of JEET has been upgraded annually since 2012. One is the SIs of the relationship of published and rejected papers, and another is the SIs of the relationship of domestic and overseas articles. The similarity analysis results using CrossCheck on JEET are used as a prior step before reviewers search for plagiarism suspicions. Articles were checked using CrossCheck in the submission site desk. The JEET Review Implicit Guideline has been used internally since late 2011. This is an internal guideline in latent and will be upgraded annually in the JEET.

### Methods

As soon as a paper is submitted to JEET, JEET checks the paper's similarity by using Crosscheck before reviewing the paper. Using CrossCheck, the JEET searches two indices (scores) of ISI and MSI. The ISI presents the total value (score) of the paper relative to the registered papers in the CrossCheck database, whereas the MSI describes the index (score) of the

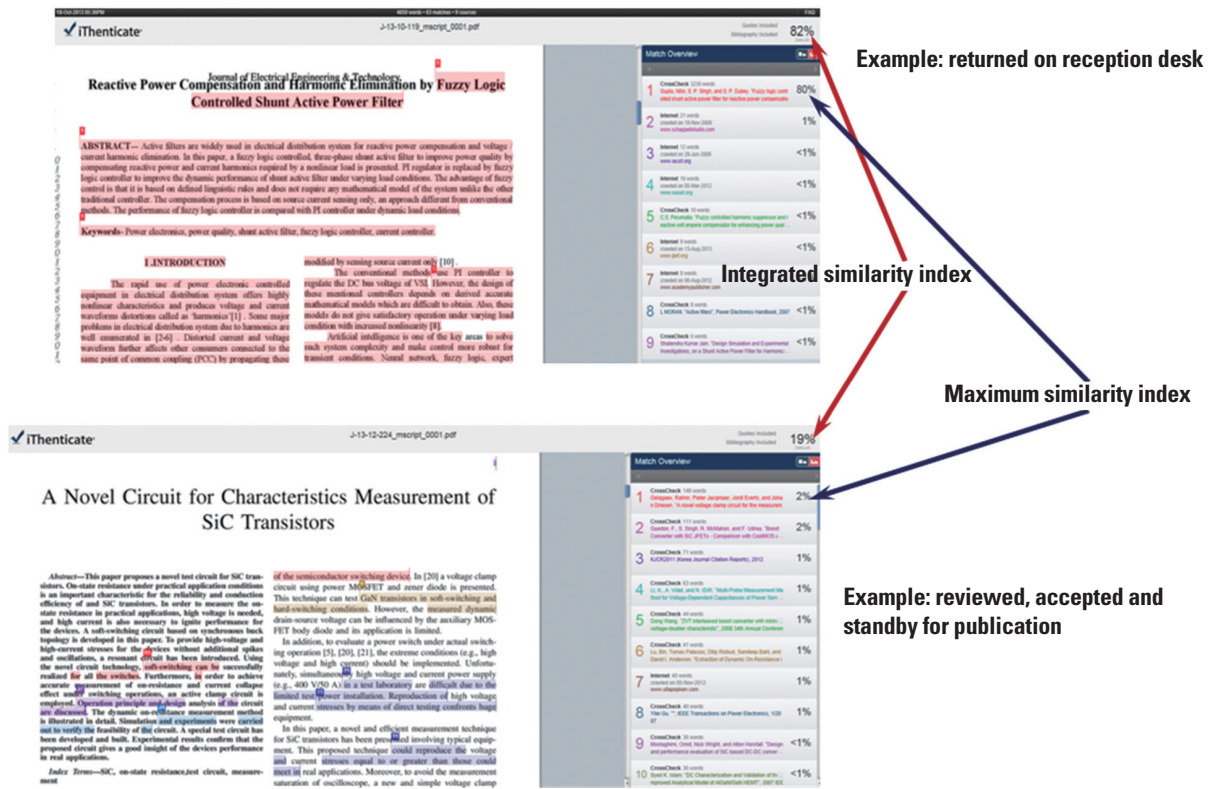


Fig. 1. Two examples: integrated similarity index and maximum similarity index on CrossCheck for two papers.

maximum similarity scored paper in the CrossCheck database. Fig. 1 shows two examples: ISI and MSI on CrossCheck for two papers. “Default” for setting option of CrossCheck is selected in this study.

Therefore, the ISI and MSI in CrossCheck can be formulated as equations (1) and (2).

$$ISI = \sum SI_k \quad k \in \Omega_i \quad (1)$$

$$MSI = \text{maximum} \{ SI_k \} \quad k \in \Omega_i \quad (2)$$

where  $\Omega_i$ : set of articles checked by CrossCheck for #*i* article  
 $SI_k$ : similarity index of #*k* article

Minitab (Minitab Inc., State College, PA, USA) was used for statistical analysis.

## Results

### Probabilistic density function of ISI and MSI of papers published and rejected on JEET

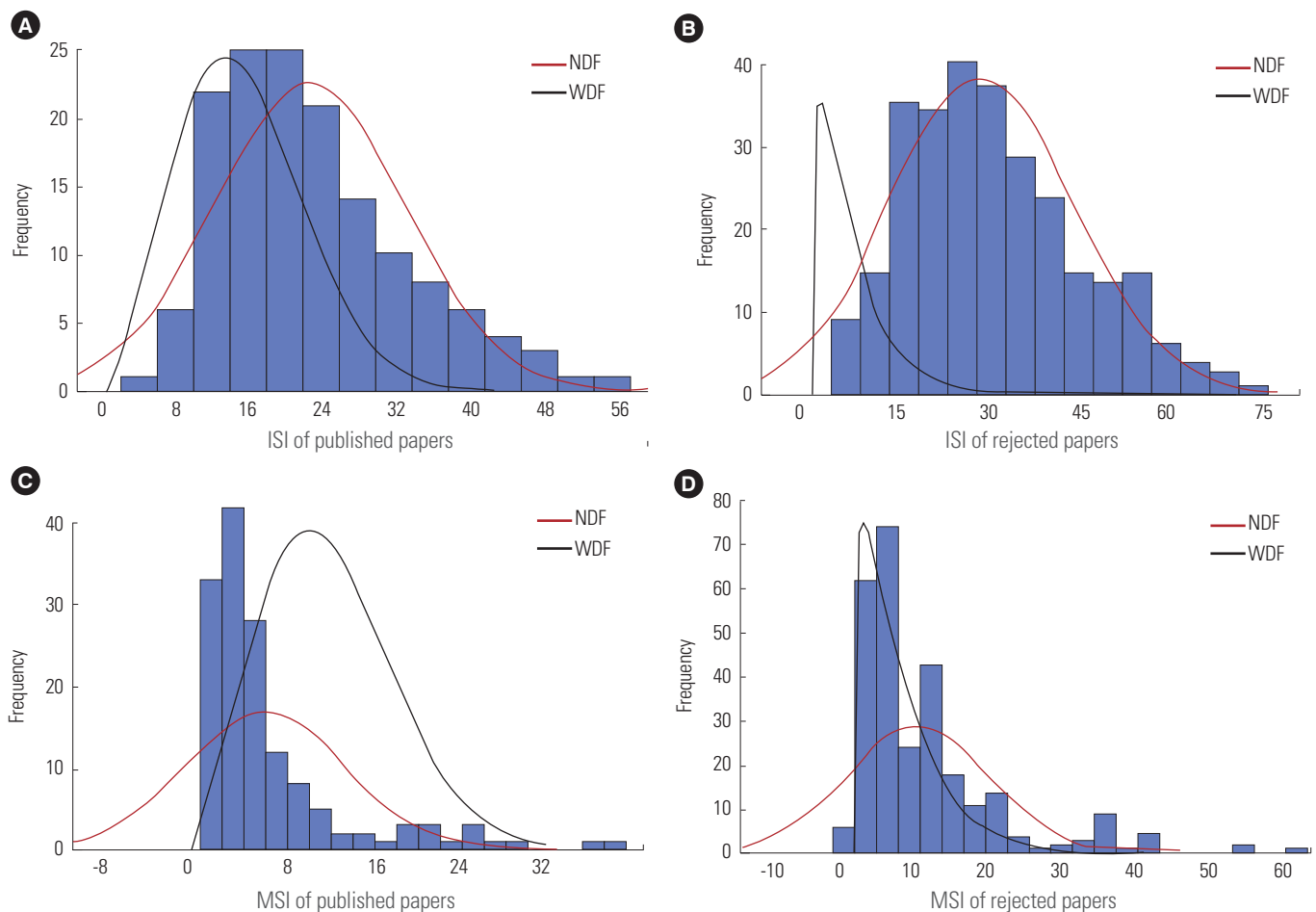
Fig. 2 shows the probabilistic density functions (*pdfs*) of ISI

and MSI in published and rejected papers on JEET in 2014. Where, the y axis is the frequency (number of papers). The *pdf* approaches Weibull distribution function rather than nor-

**Table 1.** Summary of ISI and MSI probabilistic density function parameters of Journal of Electrical Engineering and Technology in 2014

2014	A (ISI <sub>2014</sub> )	B (ISI <sub>2014</sub> )	C (MSI <sub>2014</sub> )	D (MSI <sub>2014</sub> )
Total papers	147 (2)	285 (39)	147 (0)	285 (3)
Average of ISI or MSI	22.527	30.667	7.011	9.978
SD of ISI or MSI	10.315	14.571	5.942	8.648
Shape parameter ( $\beta$ )	2.318	2.224	1.184	1.157
Scale parameter ( $\lambda$ )	25.425	34.626	7.428	10.503

ISI, integrated similarity index; MSI, maximum similarity index; SD, standard deviation. Value in parenthesis denote the number of papers with ISI and MSI scores of over 50% (A) ISI of published papers, (B) ISI of rejected papers, (C) MSI of published papers, and (D) MSI of rejected papers.



**Fig. 2.** Integrated similarity index (ISI) probabilistic density function (*pdf*) and maximum similarity index (MSI) *pdf* of Journal of Electrical Engineering and Technology in 2014. (A) ISI of published papers, (B) ISI of rejected papers, (C) MSI of published papers, and (D) MSI of rejected papers. WDF, Weibull distribution function; NDF, normal distribution function.

mal distribution function. The *pdf* parameters of CrossCheck indices (ISI and MSI) of JEET in 2014 are summarized in Table 1. In the Table, the numbers in parentheses in the row which is labeled as “total papers” denote the number of papers with ISI and MSI scores of over 50%. To avoid reviewer bias, the JEET employs a CrossCheck blind system, in which the information of ISI and MSI is not supplied to reviewers but supplied only to the associate editor. However, a similarity score of over 50% is announced to the associate editor, and the article is given to the reviewer with a plagiarism warning message. As shown in Table 1, the scale and shape parameters of published papers are lower than rejected papers. Thus, positive results have been obtained even if the ISI and MSI information was not given to reviewers (CrossCheck blind system). The same aspect is also obtained for 2012 and 2013.

### Relationship between ISI and MSI in JEET

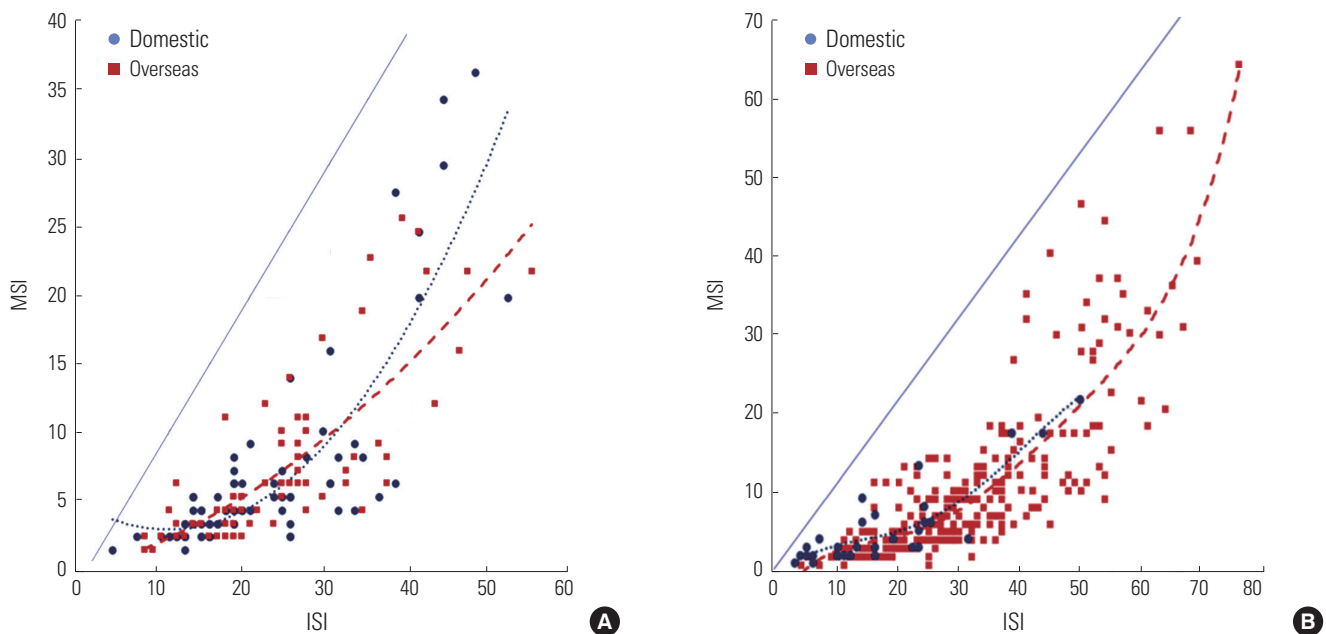
#### Relationship and comparison between ISI and MSI of domestic and overseas articles in viewpoint of published and rejected on JEET in 2014

First, the relationship between ISI and MSI of domestic and overseas articles published on JEET in 2014 is shown in Fig. 3A. The MSI of all articles is in a lower location than ISI because ISI is calculated as total similarity scores. MSI presents the similarity score of the highest similarity ranked paper in the CrossCheck list. Most articles published on JEET are located within 40% ISI similarity score and 10% MSI similarity score. Second, the relationship between ISI and MSI of do-

mestic and overseas articles rejected on JEET in 2014 are shown in Fig. 3B. Most articles rejected on JEET are dispersed in 60% ISI similarity score and 20% MSI similarity score. The ISI and MSI scores of articles rejected by the JEET are dispersed widely and are relatively higher than the indices of published articles in Fig. 2. Additionally, the ISI and MSI of overseas rejected articles are dispersed more widely and are higher than domestic rejected articles. The similar aspect is also obtained for 2012 and 2013.

#### Grouping according to plagiarism level from the relationship and comparison between ISI and MSI of articles published and rejected on JEET

To investigate plagiarism levels and make a looking for a plagiarism basic guide line from the relationship and comparison the ISI and MSI of articles published and rejected on JEET, the ISI and MSI of the published and rejected articles on JEET in 2013 are presented together in Fig. 4. A group articles have lower MSI and higher ISI compared with other groups. The sentences of the A-group articles are cited many times in other articles but have no plagiarism suspicions because of low MSI. On the other hand, B-group articles not only have higher ISI but also have higher MSI. In addition, the MSI of this group is almost the same as its ISI. Therefore, the articles are prone to plagiarism suspicions. The B group articles thus come with a plagiarism suspicious domain. The JEET editorial board has decided to set the similarity score of the B-group to 50%. The B group is called “red group” (domain). Finally, C



**Fig. 3.** Relationship between integrated similarity index (ISI) and maximum similarity index (MSI) of domestic and overseas papers on Journal of Electrical Engineering and Technology (JEET) in 2014. (A) Papers published on JEET and (B) papers rejected on JEET.

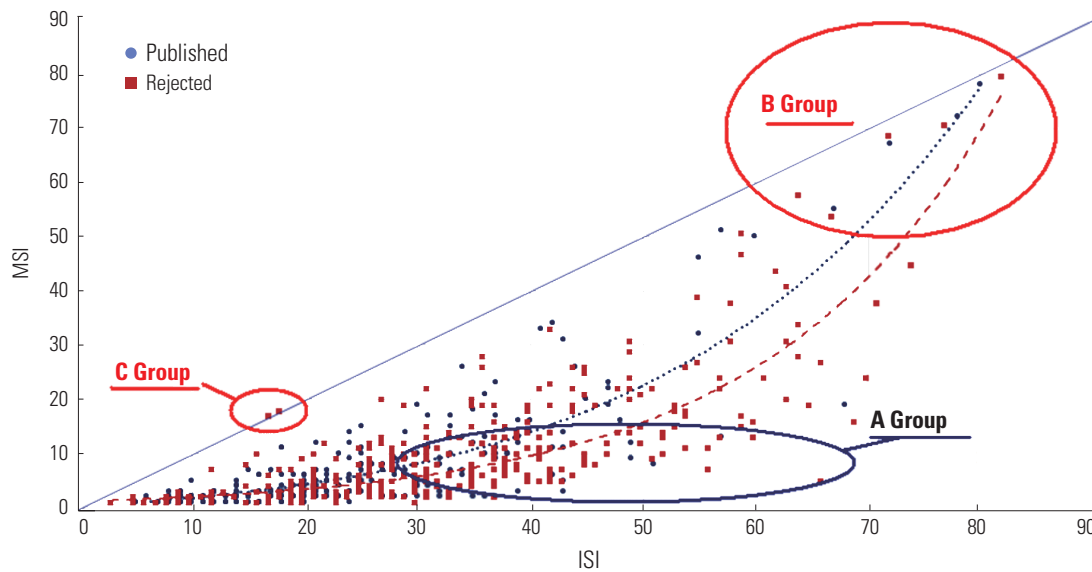


Fig. 4. Relationship comparison between integrated similarity index (ISI) and maximum similarity index (MSI) on papers published and rejected in 2013.

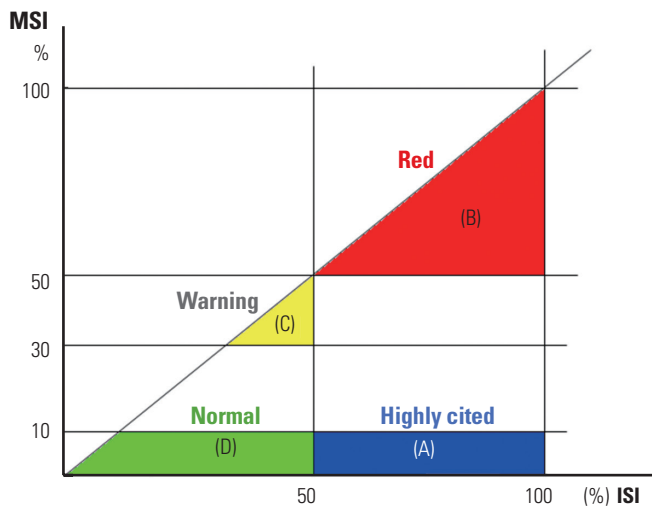


Fig. 5. Three domains defined internally from relationship between integrated similarity index (ISI) and maximum similarity index (MSI) on Journal of Electrical Engineering and Technology.

group articles are prone to use a few similar sentences, but the similarity level of plagiarism is low. Therefore, the difficulty and ambiguity of plagiarism level guideline will be used for C group. The JEET editorial board has decided to set the similarity score of the C group to 30%. The C group is called “warning group” (domain). These scores are included in the latent guideline of the JEET. The aspects are also obtained similarly for 2012 and 2014.

Fig. 5 shows three domains (normal, warning, and red) defined internally in latent from the relationship between ISI and MSI on JEET. This concept in JEET has been developed

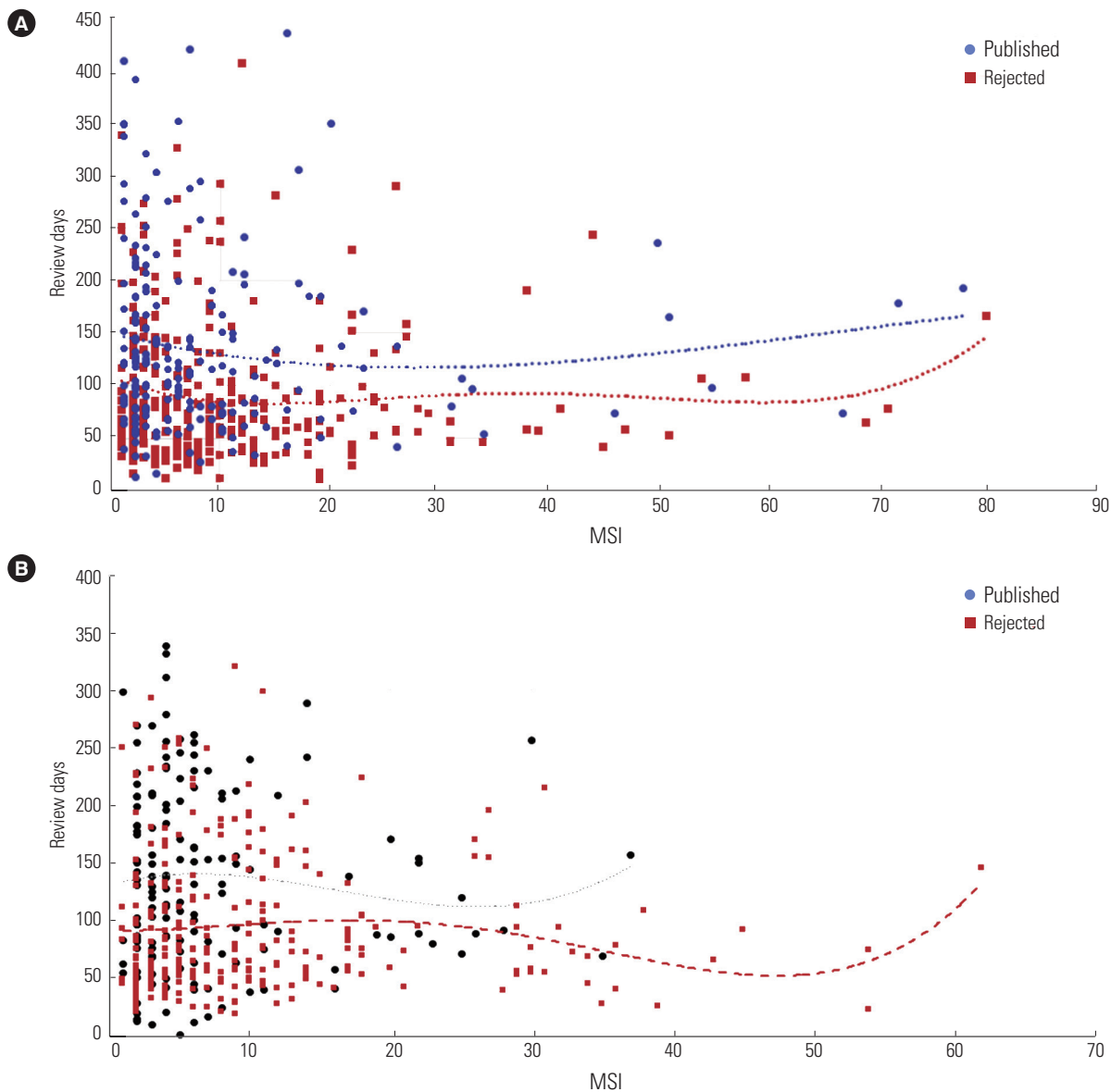
newly in 2012. In Fig. 5, highly cited papers (i.e., A domain) means excellent articles as the articles have low MSI and high ISI. The definition of the scores should be upgraded annually as it is not absolute and depends on submission number, acceptance rate, and quality level of articles in JEET. This definition is only an internal guideline in JEET.

**Relationship comparison between review period (days) and MSI of papers published and rejected on JEET**

Fig. 6 shows the relationship comparison between review period and MSI on all papers published and rejected in 2013 and 2014. Although the review period (days) is dispersed very widely (20 to 400 days), as shown in the figures, the review period has slowly decreased annually. Therefore, positive results have been obtained even if the ISI and MSI information was not given to reviewers (CrossCheck blind system). A similar aspect is also obtained for ISI. Additionally, the review days for published papers are higher than those of rejected papers.

**Summary of statistical parameters of ISI and MSI of JEET in 2012, 2013, and 2014**

Table 2 shows the average and standard deviation of ISI and MSI of JEET in 2012, 2013, and 2014. From Table 2, the ISI and MSI indices of JEET in view point of total, domestic, and overseas articles have slowly decreased annually. Therefore, positive results have been obtained even if the ISI and MSI information was not given to reviewers (CrossCheck blind system). The quality of the JEET also increased annually because the CrossCheck similarity levels (ISI and MSI) of papers sub-



**Fig. 6.** Relationship comparison between review period and maximum similarity index (MSI) on all papers published and rejected on Journal of Electrical Engineering and Technology in (A) 2013 and (B) 2014.

**Table 2.** Summary of average and standard deviation of ISI and MSI of Journal of Electrical Engineering and Technology in 2012, 2013, and 2014

		ISI (%)			MSI (%)	
		2012	2013	2014	2013	2014
Published	Total	26.85 (14.21)	26.09 (13.96)	22.527 (10.315)	8.70 (11.84)	7.011 (5.942)
	Domestic	23.2 (10.09)	23.16 (13.21)	21.827 (10.128)	8.19 (12.46)	6.568 (5.937)
	Overseas	31.94 (17.41)	30.14 (14.03)	23.877 (10.361)	9.41 (10.97)	7.496 (5.793)
Rejected	Total	36.85 (15.26)	33.02 (15.6)	30.667 (14.571)	10.73 (11.23)	9.978 (8.648)
	Domestic	32.63 (17.13)	20.86 (12.62)	18.05 (10.79)	5.64 (6.536)	5.271 (4.365)
	Overseas	37.92 (14.65)	34.75 (15.23)	32.283 (14.23)	11.46 (11.57)	10.581 (8.968)

ISI, integrated similarity index; MSI, maximum similarity index.

mitted to JEET decreased in the last years.

## Discussion

This study presents the experiences and effects of using CrossCheck in the prevention and protection of plagiarism in articles submitted to JEET, in the past three years. This study proposes three domains (normal, warning, and red) defined internally in latent from the relationship between ISI and MSI on JEET. This concept in JEET has been developed newly in 2012. Results from similarity indices supplied by CrossCheck on JEET are summarized as follows.

The ISIs of published papers on JEET in 2012, 2013, and 2014 are 26.85%, 26.09%, and 22.53%, respectively, whereas the ISIs of rejected papers on JEET in 2012, 2013, and 2014 are 36.85%, 33.02%, and 30.667%, respectively. The ISI of not only published but also rejected papers in JEET has decreased slowly in the last three years.

The ISIs of rejected papers on JEET in 2012, 2013, and 2014 are higher than those of published papers. The ISIs of domestic papers are lower relatively than those of overseas papers for both published and rejected papers in the last three years. The MSI has a generally positive monotonic relationship with the ISI. A higher SI (ISI or MSI) not only means higher rejection probability but also shorter review time (days).

Positive results have been obtained even if the ISI and MSI scores were not given to reviewers, except for the “red” papers with over 50% of ISI and MSI. The ISI and MSI scores have been used by editors and reviewers since late 2011 in JEET.

CrossCheck is a useful tool even if the tool is limited only to checking sentences and determining language similarities. In addition, this tool is still unable to check similarity (plagiarism) of equations and figures and similarity between different languages. However, it is expected that the CrossCheck also helps to protect the original authors' copyrights and to improve authors' behavior by identifying instances of academic plagiarism.

### JEET plagiarism similarity guidelines

The Editorial Board of JEET has established a guideline policy for peer review process. This latent guideline has been in

place since 2011.

Guideline 1. JEET First Previous Review Plagiarism Detection Service System guideline: As soon as papers are submitted, they are checked by CrossCheck on the office reception desk. A paper with over 50% ISI and MSI should be marked as “red” paper. While the “red” paper may still be permitted to have a review process, the ISI and MSI scores of the “red” paper are informed and announced to editor in chief, editors, associate editor, and reviewers.

Guideline 2. JEET defers the definition of score for handling the warning group until ISI and MSI data are determined and accumulated sufficiently. Nonetheless, the warning group's relative data should be stored and informed to the editor in chief and corresponding editor only if necessary.

Given that the SI does not necessarily mean plagiarism index and CrossCheck does not check figures, tables, and equations but check only texts as previously commented, the policy would be only an internal guideline in JEET and should be upgraded annually by the JEET editorial board. This study will provide useful and relevant information when making a reasonable guideline or criterion of returning papers that are suspicious of plagiarism before having a normal peer review process.

## Conflict of Interest

No potential conflict of interest relevant to this article was reported.

## Acknowledgments

The authors would like to acknowledge the editors and reviewers of JEET publication and members of InfoLumi.

## References

1. Zhang H. CrossCheck: an effective tool for detecting plagiarism. *Learn Publ* 2010;23:9-14.
2. Zhang X, Huo Z, Zhang Y. Detecting and (not) dealing with plagiarism in an engineering paper: beyond CrossCheck. A case study. *Sci Eng Ethics* 2014;20:433-43.