Viewpoint

Update on Similarity Check usage and developments

Madeleine Watson
Product Manager, Crossref; mwatson@crossref.org

While the landscape of scientific publishing may have undergone an evolution of sorts over the last decade, with new technologies providing the capacity for alternative publishing models, publishing integrity remains a central tenet. The role of the journal editor is therefore imperative in helping to maintain the quality of scholarly outputs. In her recent publication, *Against Plagiarism: A Guide for Editors and Authors*, Helen Zhang describes the role of the journal editor as being able "to identify plagiarism when it occurs, to determine the nature and severity of the plagiarism, and then to take appropriate action".[1] Today there is a range of technology-led tools available on the market to help editors do just this. And although we know that technology alone cannot definitively determine true plagiarism, a view Zhang also shares, it is certainly advantageous for editors to have another tool in their workflow to help them isolate and investigate instances of suspected academic misconduct.

This is where the Crossref Similarity Check service (formerly CrossCheck) plays a role. Similarity Check comprises two key features: a full-text database containing copies of all current and archival content published by participating members; and access to Turnitin's powerful plagiarism checking tool, iThenticate.

Current statistics

Similarity Check has grown to become one of Crossref's most-subscribed services. With over 1100 publisher members currently, the service has indexed over 50 million full-text content items in the database from over 193,000 journal titles. With 82 million records currently registered with Crossref (as of June 2016), this means that 61.5% of all registered content is now available to Similarity Check members to compare their own manuscripts against. We also know that iThenticate usage is consistent with this growth in membership, with an average of 323,325 manuscripts screened in iThenticate per month by Similarity Check members during the first half of 2016.

What's new in 2016

Collaborating on the development of a new content intake system (CIS) in 2016, both Crossref and Turnitin are focused on developing technology solutions which best support methods of robust and reliable indexing. Due to the introduction of the CIS, the service is now transitioning to a single method of indexing via as-crawled URLs, located within the metadata member's deposit with Crossref. This will provide Turnitin with a faster and more reliable method for collecting full-text data links from Crossref, therefore reducing the time delay between a paper's initial publication and its subsequent inclusion in the Similarity Check full-text database. This will help to reduce the likelihood of users missing potential matches to recently published content.

Crossref and Turnitin are also committed to better supporting editors in their use of iThenticate. With this in mind, Crossref signed a new agreement with Turnitin in early 2016 which includes the addition of 2000 hours of dedicated development time each year by Turnitin's engineering team to focus on iThenticate feature developments to support Similarity Check members and their needs. In order to inform a feature development plan, Turnitin initiated a survey to find out more about how Similarity Check members use iThenticate. The survey was sent to all Admin account holders via email in May 2016, and Turnitin received 105 responses.

Survey results

Respondents represented a spread of experience with Similarity Check. The largest category (38%) reported to have been using the service for between one and three years, and 86% of all respondents reported that the service is either extremely or very helpful in identifying cases of academic misconduct. In terms of the number of cases identified, 41% of respondents reported that Similarity Check helped them to discern over 11 cases each year.
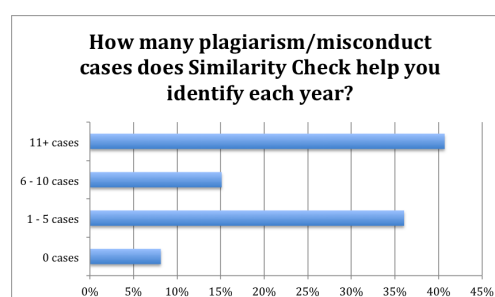

Figure 1 Number of plagiarism cases identified
*answered question: 86, skipped question: 19

A majority of this sample (87%) use Similarity Check through application programming interface (API) integration with their manuscript tracking system, and of this group only 5% reported being not so satisfied with their integration.

Regarding how editors are integrating Similarity Check into their workflow, over half of those surveyed (58%) reported using iThenticate daily, with a further 26% indicating weekly usage. Members are communicating their use of Similarity Check to authors, with 82% of respondents

notifying authors that their work will be checked for similarity. A majority of users (84%) take advantage of the fully formatted document report tool (Document Viewer) when reviewing matches, with only 16% using the screenshot view in the text-only report. Interestingly, just over half the respondents to this question (52%) confirmed that they review every similarity report they generate, regardless of the similarity percentage, whereas 37% only review reports which show a similarity score exceeding a specific threshold. The remaining 10.5% reported using other indicators to determine when to review.

In terms of understanding similarity report results, most respondents to this question reported they found it either very easy (43.0%) or somewhat easy (38%).
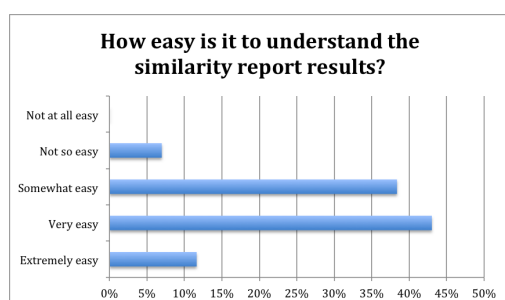

Figure 2 Ease of understanding similarity reports
 *answered question: 86, skipped question: 19

The data also showed that, while 54% of respondents were either extremely or very satisfied with the report exclusion options (eg the ability to exclude quotes), there was an indication that this may be an area for future improvement due to 38% reporting they were only somewhat satisfied.
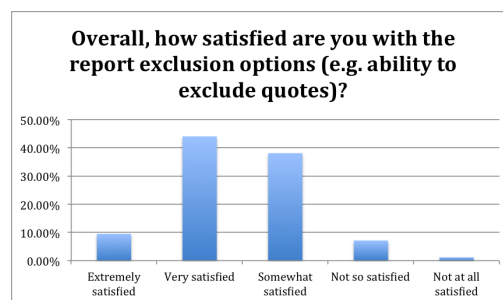

Figure 3 Satisfaction with exclusion options
*answered question: 84, skipped question: 21

Our future vision

It's encouraging to see that Similarity Check members are finding the service useful; however, both Crossref and Turnitin are keen for the service to adapt and improve so that it continues to help users ensure the integrity of papers. The survey results indicate there are opportunities to focus feature developments around items such as increasing readability of the similarity reports and improving report exclusion options in iThenticate.

If you have any feedback or suggestions for either iThenticate feature developments or general improvements to the Similarity Check service, please contact Madeleine Watson (Crossref Product Manager) at similaritycheck@crossref.org.

References

1 Zhang YH. *Against Plagiarism: A Guide for Editors and Authors.* Switzerland: Springer, 2016.

Updated EASE guidelines

The updated *EASE Guidelines for Authors and Translators of Scientific Articles to be Published in English* are now available as an e-article only.