

CrossCheck – interpreting the similarity reports



Introduction

One of the questions that comes up a lot from CrossCheck users is this: 'I've run a paper through iThenticate and I can view the similarity report – now what?' As you'll see when using the service, the reports that iThenticate produces are comprehensive, with matched text highlighted in the manuscript that has been uploaded and the sources it matches. It's the 'now what' part that we aim to cover in this document – what do the matches mean in different contexts, what is cause for concern and what is not.

Interpreting Reports and Reading Percentages

Interpreting the results of the report is the most difficult part of using CrossCheck, as unfortunately the report can't say 'yes' or 'no' to whether something is plagiarism, it can only present you with what it finds. Over time you will get more familiar with the sort of results that the system displays and find that the process of evaluating reports will become quicker.

Here are some key points to remember when looking at the reports:

- Experience is key. Over time each journal will learn what is a 'usual' percentage match for them and what is unusually high. This differs from subject to subject and journal to journal.
- A good way of getting an idea of whether or not something is suspicious is to look at the individual percentage matches that are being flagged up for a particular paper. If the overall similarity index for a report shows 40% but this is made up of 40 x 1% matches to sources then it's probably fine, but if it's one 40% match or two 20% matches to a particular source then this you will want to look into it in more detail.
- Along with the individual percentage for each item, the number of words matched between the papers can be found beside the individual percentages, and sometimes gives a better indication of the extent of a match than the percentage.
- The percentage shown in cases of full copying is unlikely to be 100% due to differences in journal page headers and footers and possible re-wording of sections of text. Also, a match may look artificially high due to standard phrasing, page numbering etc. This is why it is always important to look at the report itself before drawing conclusions. This is discussed further in the 'Checking a Similarity Report' section of this document.
- **A submitted paper with a high similarity percentage does not mean that the author has intended to copy work.** For example, re-use of standard phrases within a subject area may give an article a high percentage match when there is a valid reason for the sections of text being the same or similar to other published work. These are discussed in more detail in the next section.

Checking a Similarity Report

The following steps can be taken when looking through a similarity report:

- Look through the matches that the report shows and scan the full paper to get a rough idea of the extent of similarity between the two papers. Is it just page headers that have matched? Does it look like there is standard phrasing used to describe common concepts within the subject area? There are certain standard phrases and explanations that are commonly used within academic fields to explain certain aspects of a discipline. Using these standard explanations helps create consistency within a field. So although text may be similar, it may take a subject specialist to look at the matches to determine if standard text is being replicated and would not constitute plagiarism.
- Remember that the number of words matched between papers can often be a more useful guide than the percentages. Matches that are over 1000 words should be checked particularly carefully. Are the matches large and/or in sequence? This can be a concern and may suggest the author's work is not original.
- Check where the matched paper comes from. Click on the link to the matched paper. Is it from a publisher site, author homepage or a university repository? If the paper is from an author homepage or a university repository, it could be that the paper has been posted as a pre-print. Other open access repositories such as arXiv allow authors to post pre-prints. Check for copyright on the articles. A paper from a publisher site often copyrighted to that publisher unless expressly stated on the article.
- Referencing:
 - Check the references in the submitted paper. If the matched paper is referenced, is it referenced sufficiently for the extent of similarity? Does the submitted manuscript explain that the paper is an extension of the matched paper? Some papers are extensions building upon previous work, however this must be clear in the submitted paper and the matched paper should be referenced appropriately throughout i.e. in all appropriate sections where material is used, not just once at the start of the submission. It may be the case that the author is using sections of their own previous work in their paper but again, this must be referenced correctly. The section on self-plagiarism later in this document explains this in a little more depth.
 - Similarly with referencing, authors should have ensured that any text quoted verbatim from another source should be clearly identified as such. A reference alone is not enough if a section of text is being reproduced word-for-word. Use of quotation marks is encouraged in this situation to make it clear to readers which text has been directly re-used.
 - Some authors may replicate passages of text they feel are well-known enough in their discipline for an expert reader to know where they have come from originally. Or an author may say they have re-used text (without explicitly noting that this is the case) from another author by way of paying

homage to them – remember that the re-used text must be referenced correctly.

In all cases editors should consider the consequences of the duplicated passages and its potential to mislead readers.

- Take a final, more in-depth look at the similarity report. Look through the paper again and determine if the paper is acceptable, if changes need to be made by the author, if it needs to be rejected from the journal, or if it requires a subject specialist to look through the report – getting a second opinion can often be useful. Obtaining a PDF or a pre-print of the matched article is useful for comparing equations or figures within the text and may be found by searching the matched article title using a search engine.

Reviews: points to bear in mind

Review articles often bring back higher similarity scores because they are commenting on or developing ideas on already published work, so it may be useful to consider the following if the article you are checking is a review:

- Although review articles are expected to cover ideas and information from previously published works, it is important to note that they should still contain original thoughts and ideas from the author of the review article. A higher similarity score may come back, but you the journal should still be satisfied that the author has added their own unique contribution in writing the article.
- The other item to note is that the authors of systematic reviews may change between different versions of the review. As such, the iThenticate report may show a large overlap between two reviews that have different authors. This may look problematic, but can be explained by the authors. Again, it is always helpful if they make this clear upon submission of the updated systematic review.

Self-plagiarism and ‘Patch writing’ or ‘microplagiarism’

Self-plagiarism: (the verbatim or near verbatim re-use of significant portions of one’s own work without citing the original source) is also an infringement against publication ethics. However, this does not apply to legitimate uses of an author’s own material, as sanctioned by the copyright holder, e.g. in collected works and where an explicit reference is made to the prior publication. The section below from the *Springer CrossCheck Guide*¹ puts it:

This may not seem a problem, but dependent on the amount of duplication could lead to copyright problems with publishers. Further, every journal has the right to expect that new and novel information is being presented to further the science.

¹ How to interpret the results using CrossCheck/iThenticate: Springer Guide V1: 2012 [Internet]. Provided by Springer [cited 2014 Jan 12]. Available from: http://static.springer.com/sgw/documents/1370303/application/pdf/Springer+CrossCheck+Guide_2013.pdf

'Patch writing' or 'microplagiarism': It is recognized that authors who are not writing in their native language are experiencing difficulties in correctly paraphrasing the work of other authors. "This may result in so-called 'patch' (or 'patchwork') writing." iThenticate will then show many 'dislocated' matches, e.g., many individual sentences or parts of sentences coming from many different sources. Each copied section is short, but will result in a high similarity percentage. The sources of the copied text are usually not referenced. This form of duplication is considered somewhat acceptable if it is minor and when it concerns materials, methods and/or procedures. But is not acceptable if the majority of the paper has been lifted from elsewhere, with only a few word changes here and there, or it leads to misrepresentation, e.g., the phrases copied are inappropriate or there is the attempt to deceive.

Limitations of the reports

Not all publishers have signed up to CrossCheck so their material will not be covered in the database. Once a Publisher has joined it takes time to index all of their content and get it added to the CrossCheck database so it becomes searchable. It may be helpful to look at the list of CrossCheck members here: http://www.crossref.org/crosscheck_members.html or at the list of titles that have been indexed in the CrossCheck database which are updated regularly on the CrossRef website: <http://www.crossref.org/crosscheck/index.html>

If you notice a database/website that you feel should be included in the CrossCheck database then you can complete these forms: [Submit a database](#), [Submit a URL to crawl](#).

iThenticate can only compare the manuscript you upload against the full CrossCheck database. It cannot compare one document against another unpublished document – at least one of the items for comparison should be a published paper.

Where language is concerned, iThenticate can only check a language against other content in that same language. If an editor uploads a paper written in Spanish to check it, it will be compared against other papers written in Spanish. The system cannot translate the paper in order to check it against papers written in another language. It is worth noting that an author should always let the journal know if they are submitting a paper for review that has previously been published in another language. The original publisher may need to grant permission for the paper to be re-used in this way. It may also be misleading to readers, so it is always best for authors to provide clarity as to any previous uses of their article.

Equations often do not match up correctly as the symbols used are often treated as figures.

It is also worth noting that figures are not compared by the reports. The system is not able to match figures to other figures found online.

What next?

If a paper is acceptable, no action is required and the paper can carry on through the peer review or production process, as appropriate.

If you do suspect misconduct, the CrossCheck Guidelines on Handling Plagiarism (available at: http://www.crossref.org/crosscheck/crosscheck_guidelines_on_plagiarism.pdf) suggests the following approach:

Different journals and publishers vary in their approaches to dealing with misconduct. You should therefore refer to your publisher (if you have one) for their guidelines if you identify cases of possible misconduct and those guidelines should control your actions.

Raising an allegation of plagiarism is a very serious matter and has the potential to cause damage to an author's, editor's or publisher's reputation and career. In some cases, an accusation that is not well-founded could create possible liability for defamation.

Tools like iThenticate can identify possible instances of misconduct through showing similar text, but whether a particular overlap in content rises to the level of plagiarism or even copyright infringement is something that must be determined on a case-by-case basis and evaluated in the light of all the available facts. It is almost always advisable to contact the author about the issue that has been identified and invite him or her to respond. The nature of the response will aid you in deciding about appropriate next steps. While ignorance of publishing ethics is not an excuse, it may influence the way in which the matter should be handled. All necessary precautions should be taken to ensure that communications exchanged with an author about a possible plagiarism situation are kept confidential, secure, in writing, and are carefully phrased to avoid accusatory or possibly defamatory language.

Some publishers share the similarity reports with their authors, but some choose just to make reference to their concerns via email and not send the full report.

The publisher will be responsible for handling any correction of the literature, and each publisher may have a slightly different procedure for doing so, based on their own internal guidelines.

If you have any queries regarding account set up or use, billing or technical issues/faults then please contact ccsupport@ithenticate.com.

Acknowledgements: This document has been compiled from input from CrossCheck contacts at Taylor & Francis, Wiley and Springer. If you have further comments or would like to add your own input to the document, please contact rlammey@crossref.org.