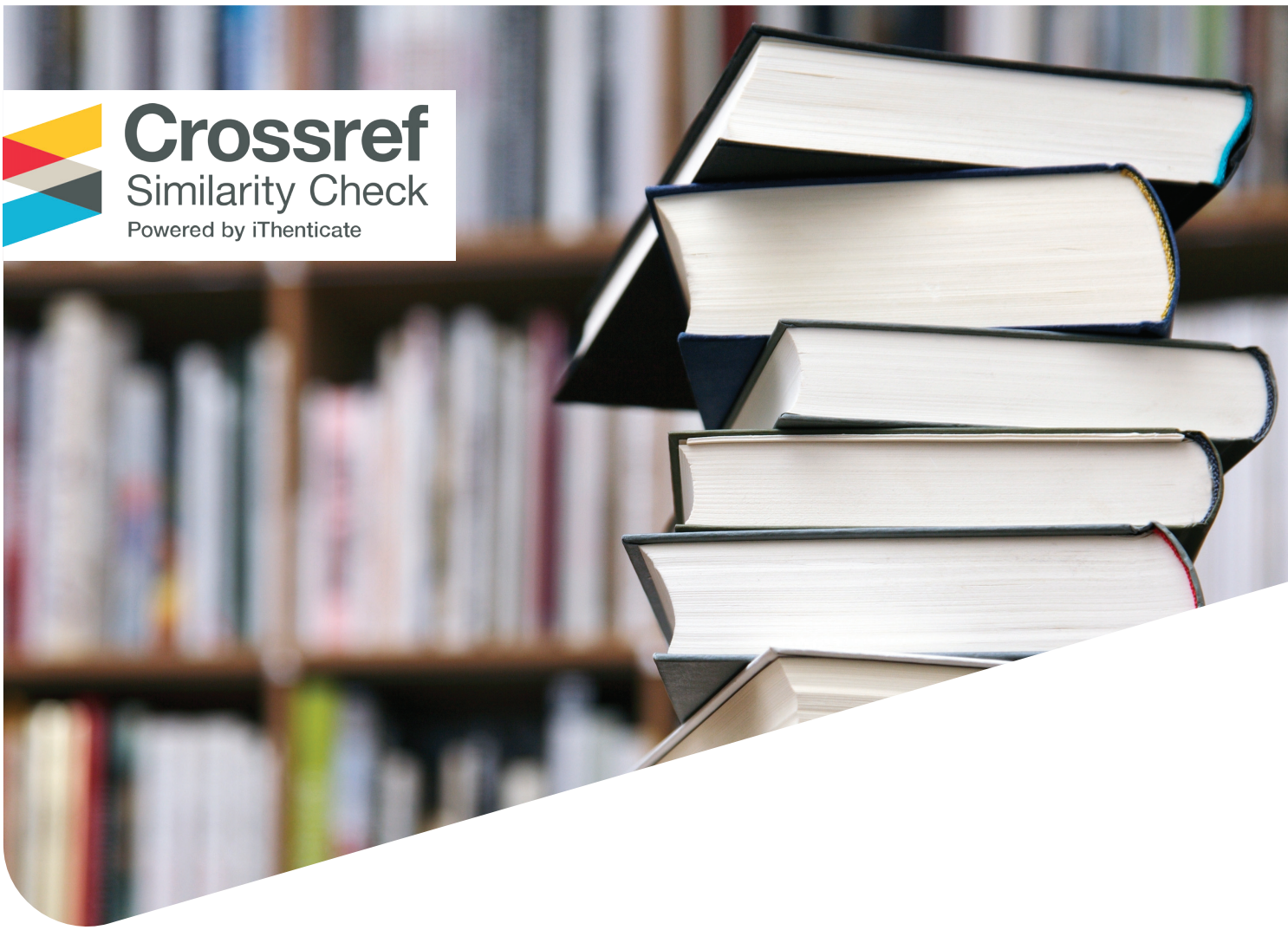
SPRINGER NATURE

springernature.com



Springer Nature Guide

HOW TO INTERPRET RESULTS FROM USING CROSSREF SIMILARITY CHECK

(Powered by iThenticate)

For journals, books and proceedings

Table of Contents:

Introduction 3

How to get started 4

The Document Viewer Report 5

Limitations 9

After peer review 9

How to read the percentages 10

Document to document comparison 11

Cut & Paste 12

Types of plagiarism 12

Potential pitfalls 13

Actions to take if a document appears to contain duplicated text 14

Minor overlap in text 14

Substantial overlap in text 15

Downloading a PDF version of a similarity report 15

Sharing results with external parties 15

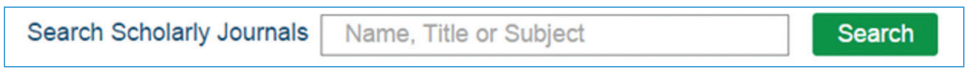
Queries 16

1 Introduction

Crossref Similarity Check (powered by iThenticate) can be used to cross-check documents against thousands of published content items in their database. It highlights text passages that a submitted document shares in common with previously published material.

Please note that only content from Crossref members is available in the Crossref database. Content is continually added.

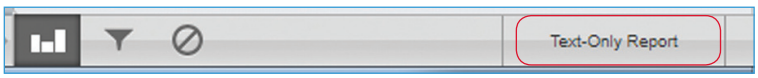
Generally **all major publishers** are members, but if you wish to know whether a particular publication is included in the database, please visit the **search** link and complete the 'search scholarly journals' in the middle of the page:



Important! Crossref Similarity Check simply checks for text similarity by comparing documents. It makes neither legal judgments nor legal recommendations.

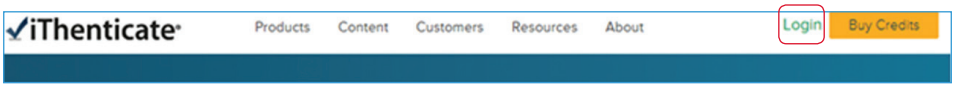
Using text similarity screening software can save authors embarrassment, if, for example, the duplication was unintentional. It is designed to make Journal, Proceedings and Book (Series) Editors aware of this important issue and to help identify situations that can be remedied in advance of publication. The results of the software and the guidelines in this document do not substitute for legal advice. If in doubt about the legal impact please contact your legal counsel or department.

Similarity Reports are available in two viewing modes: **"Document Viewer"** and **"Text-only"**. Document viewer is set as the default report view and shows the submission in its original format including figures and tables. If you would like to use the "text-only" view you can select this view in the menu on the lower right hand side. This guideline describes the Document Viewer. For guidelines on Text-only please view the following **guidelines** (pages 47-54).

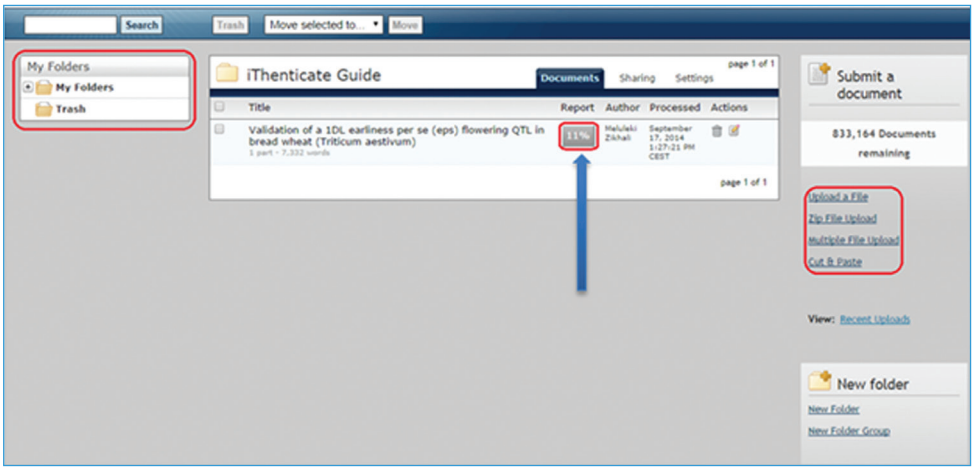


2 How to get started

First time users will have received an automatic message with **log-in** details. To log-in click on the Login button on **www.iThenticate.com**.



The below screenshot shows the homepage of the Crossref Similarity Check. On this page you can organise your **folders** and upload **files** you wish to screen for similarities. When you have uploaded a document and the file is processed by iThenticate, click on the percentage, i.e. the **similarity index**, to open the similarity report.



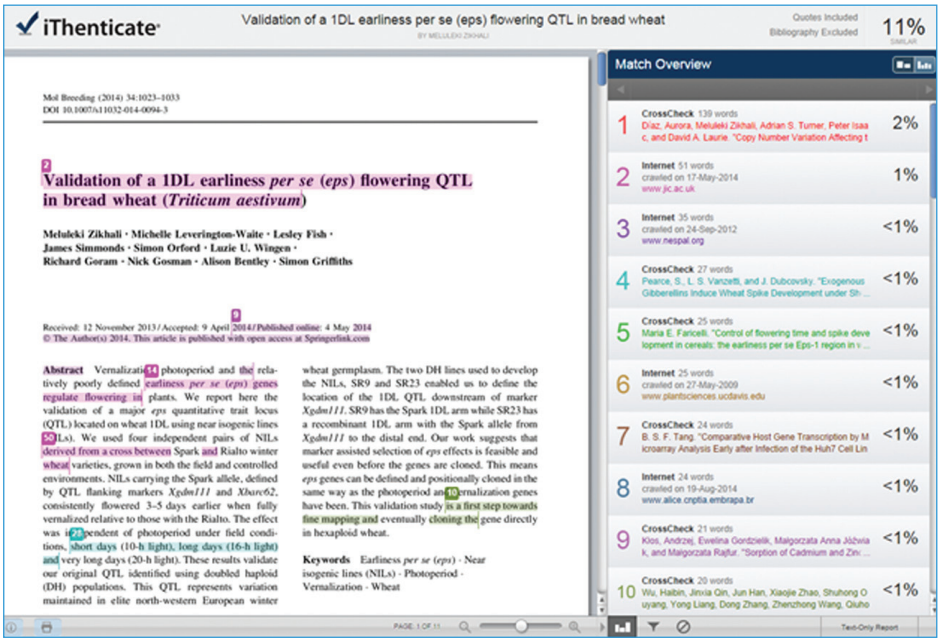
What are the document submission requirements?
Files cannot exceed 400 pages. Files cannot exceed 100MBs. Files cannot exceed 2MB of raw text.

What are the accepted file types?
MS Word, Word XML, WordPerfect, PostScript, PowerPoint (.PPT, .PPTX), PDF, HTML, RTF, HWP, OpenOffice (ODT) and plain text. ***Please note that LaTeX is not an accepted file type.***

What do you do if you have forgotten your password?
On the Log-in page click on “forgot password”.

On the next page enter your e-mail address and you will receive a one-time temporary password to this account. You will need to change it the first time you log in.

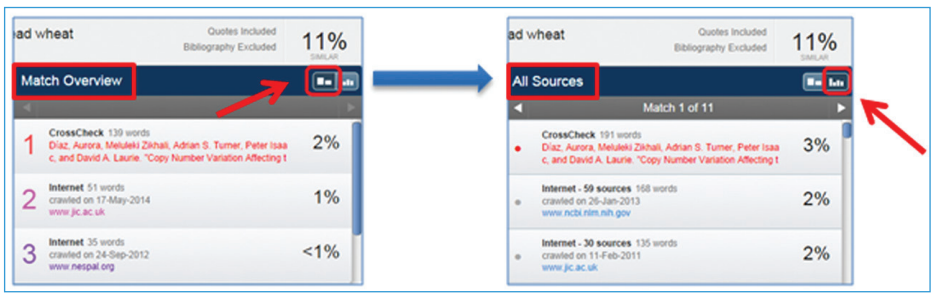
3 The Document Viewer Report



The document viewer has two views:

- **Match Overview:** This view gives a summary of all similarities found. The top source match is shown first.
- **All Sources:** This view shows the similarities with one source at a time and makes individual comparison of sources possible. As with the match overview, the top source match is shown first.

  Important note! Similarity with a particular source may be higher in 'all sources' then in 'match overview'. The match may be lower in 'match overview' as it is masked by similarities in other sources. The 'all sources' view includes all hidden matches.

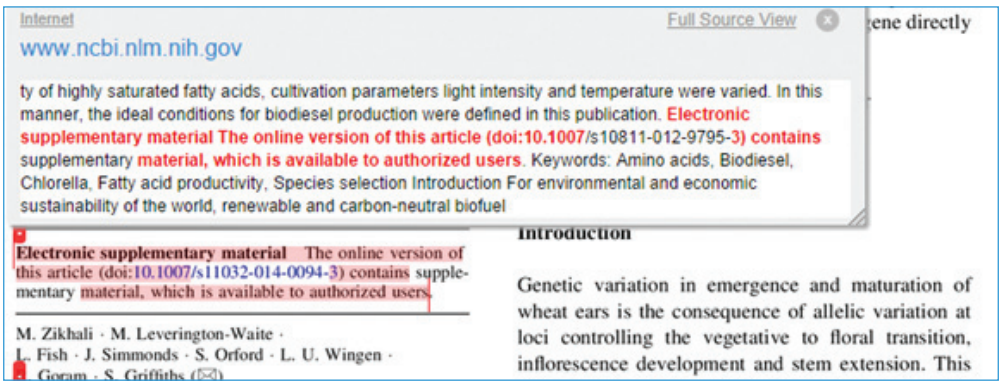


- The sources are clearly marked as “Internet”, “Publications” or “Crossref” source. The latter concerns research articles, books and conference proceedings available in the Crossref database. Publications are third party periodical, journal, and publication content including many major professional journals, periodicals, and business publications. Clicking on the bullet point under “Internet” will show a break-down of URLs crawled at multiple times.

  Important note! ‘Crossref posted content’ from publishers using Crossref Similarity Check is indexed. Posted content consists of preprints, eprints, working papers, reports, dissertations, and many other types of content that has been posted but not formally published. Content is indicated as “Crossref Posted Content”.

  Hovering over the percentage after the source will show an arrow. Clicking on the arrow will take you to the source. Sources may or may not be accessible - if the source is published by another publisher it may be behind a paywall.

- A side by side view will first show a preview of the matching text. Clicking on Full Source View will show the full source and the matches therein.

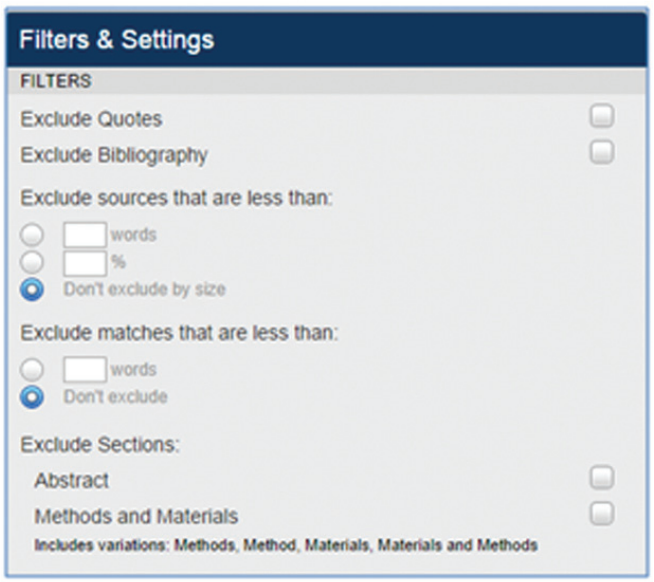


- A filter can be used to exclude quotes, exclude bibliography, exclude matches of X number of words (or percentages), exclude abstract and/or exclude Methods and Materials by clicking on the button **"View/edit filters and settings"**.

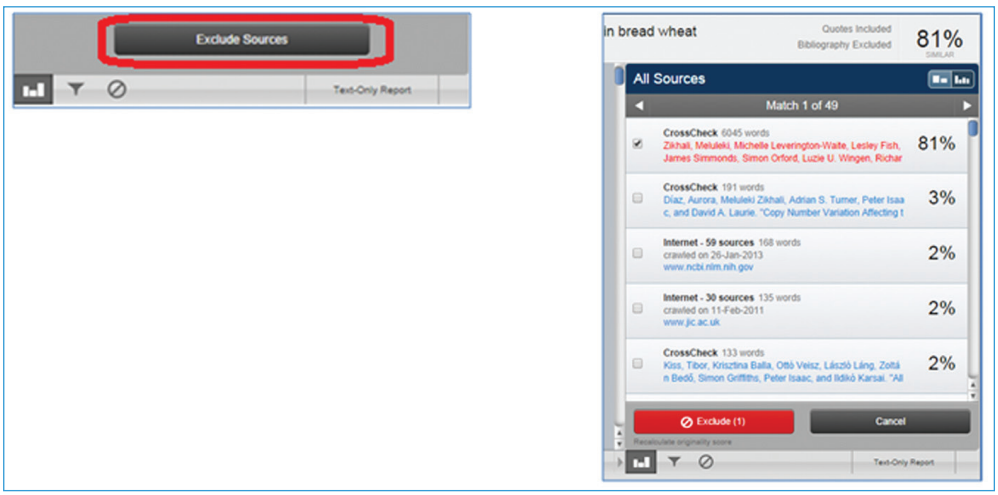
  Note: When the bibliography is excluded, the word count of the bibliography is excluded from the overall similarity percentage to allow for a more accurate similarity index.

Important! Please be aware that section exclusion may not work properly if documents contain:

- ▶ Watermarks
- ▶ Unevenly spaced line numbering
- ▶ Sub-headings that are indistinguishable from the Methods and Materials heading
- ▶ Abstract or Methods and Materials section appearing in a table
- ▶ Section headings and body text using the same font and font size



- Sources can be excluded in the All Sources viewing mode by clicking on the button **"Exclude Sources"**. By ticking the box in front of a source, the source is excluded.



- To view and/or edit excluded sources click on the icon in the bar. Tick the box in front of the source and click on Restore.

4 Limitations

iThenticate cannot compare:

- figures and tables
- text in different languages
- paraphrased texts
- content that is not in the database (see introduction)

We recommend checking the **"what's new"** link regularly to get the latest updates on improvements.

It should also be noted that there is a time lag between content being published and content being indexed in the CrossCheck database.

5 After peer review

The similarity percentage reports the total number of matching words found in the document divided by the total word count of the document. It does **not** indicate that the document contains plagiarism. The percentage therefore should not be interpreted as "this document contains x% overlap with other sources, nor should a percentage be used as a threshold or benchmark. E.g. *anything below 15 or 20% is fine*.

Each document should be viewed and evaluated separately. A percentage includes all matches, also matches to names, affiliations, information about funding, methods & methodology, and other usually innocent matches. These matches are added to the similarity index. The similarity index should thus be used only as an indication.

6 How to read the percentages

There is only a concern if the similarity check:

- Detects a substantial amount of text in the document that has been duplicated from one or more original sources without quotation marks, and/or
- The text from the original source is not cited in the document

A high similarity percentage should be further investigated for:

Text from Many Individual Sources

How many individual sources are linked to this document? The best way to check this is to use the Document Viewer Match overview mode. A document with single sources where each source shows similarity of only 1-5%, is generally showing no sign of potential plagiarism, provided that the majority of the document is not made up of borrowed strings of material.

High Percentage from a Single Source

Some documents with a low overall similarity percentage can have a high percentage from a single source, for example, the similarity index is 15%, but 12% is coming from one single source. Generally, the individual sources should not give figures higher than approximately 10%. If the figures are higher, then the level of similarity should be checked carefully to establish whether the copied material is posing a problem or not.

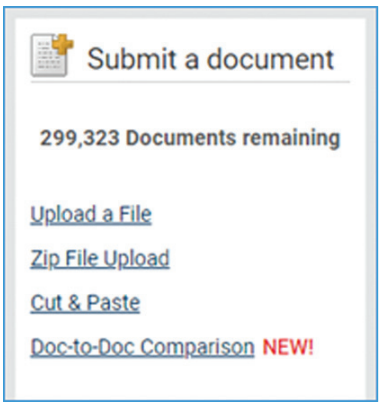
Tip! For large documents, for example book chapters, it makes sense to also take into account the number of words.

Recommended rule of thumb:

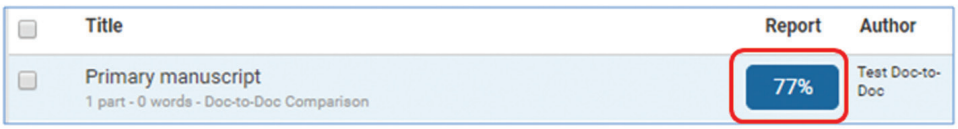
- Anything above 10% is screened quickly.
- Anything above 20% should be checked more carefully.
- Always look at the context of the match and keep Crossref Similarity Check (powered by iThenticate) limitations in mind.

7 Document to document comparison

When uploading a document there is the option to do a Doc-to-Doc comparison. This feature allows for a single, i.e. the primary, document to be compared with up to five documents.



Only the primary document will show in the results. Clicking on the percentage will show the comparison documents.



8 Cut & Paste

With the cut & paste option it is possible to check specific parts of a document. This may be helpful for documents that are not supported by the Crossref Similarity Check. See also section two 'How to get started'. It is only possible to cut & paste text. Images, tables will get lost when pasted. Clicking on the percentage will show the results.

9 Types of plagiarism

The following factors may be helpful in distinguishing types of plagiarism (identified by the Committee on Publication Ethics (COPE) (2011) and slightly modified):

- **Extent**
  - Does the similarity concern a few words or a few sentences, a whole paragraph, several paragraphs, large parts of a document or a whole document?
- **Originality of copied material**
  - Does the document contain technical language and/or standard phrases to describe materials, methods and/or procedures?
  - Is the similarity related to the author's own work?
  - Does the document contain uncited reuse of previously published material?
  - Is verbatim material enclosed in quotation marks?
- **Position / context / type of material**
  - Abstract: A large match in the abstract is a tip off that something could be amiss with the article.
  - Key words: If, next to the abstract, the keywords show a high level of matching words then this also tells you to look at the document more carefully.
  - Materials and Methods / Procedures: There are only so many ways in which this part of a document could be phrased. It should be checked whether the overlap concerns a standard procedure that cannot be rephrased. At the same time, taking whole paragraphs and simply finding and replacing one word is not acceptable.
  - Results and/or Discussion / Conclusions: A match in this area could be an indication of plagiarism or duplicate/redundant submission.
- **Referencing / attribution**
  - Is the similarity related to work that has been properly cited in the submission?

10 Potential pitfalls

It should be taken into account that scholarly works often summarize the work of other researchers and any boundaries between what is legitimate representation and what is copying of original material are therefore difficult to set.

A few pitfalls defined by Wager in a COPE discussion paper (2011) that may occur:

- **Review articles:** Review articles are expected to give a summary of existing literature. Authors should use their own words with exception of properly quoted and/or cited texts and the work should include a new interpretation.
- **Updating of systematic reviews:** "An updated review, naturally, contains large sections from the previous versions and this may appear to be plagiarism if the authors have changed (and iThenticate will not recognize acknowledgements to previous versions)." (Wager, 2011)
- **Is referencing enough or should quotation marks be used:** Generally, referencing is not enough. If it indeed concerns verbatim copying, no matter how small the text, quotation marks need to be used.
- **Copying out of flattery or 'homage' to an author:** It is never acceptable in international academic publications to copy the work of others without appropriate attribution and quotation marks.
- **'Patch writing' or 'microplagiarism':** Crossref Similarity Check will show many 'dislocated' matches, i.e., many individual sentences or parts of sentences coming from many different sources. 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 document has been lifted from elsewhere.
- **Self-plagiarism:** This may not seem a problem, but dependent on the amount of duplication, could lead to copyright problems with publishers. Further, every publication has the right to expect that new and novel information is being presented to further the science.

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

Important! Collect all the facts before making a decision.

11 Actions to take if a document appears to contain duplicated text

Every case is different and therefore decisions will vary per case. Ask yourself the following question: Does it concern an honest mistake or is there an intentional deviation from the scientific norm? Please note there are many grey areas between honest, questionable and fraudulent practices.

Whilst reviewing the case consider the following factors:

- Author seniority. Junior authors may be asked to paraphrase the copied text if it is believed that they are genuinely not aware that copying phrases is inappropriate. It is expected that a senior author should know better.
- Cultural background could be an indication for potential different behaviors concerning the amount of copying which could be seen as plagiarism.
- An author may have permission to duplicate already-published text.

11.1 Minor overlap in text:

If only a small section of a document is showing similarity (the copied text could be verbatim copying or close copying, i.e., not quite verbatim, but changed only slightly from the original), and it is clear that there is no intention to deceive, the document should be returned to the authors with a request to rewrite the plagiarized text in their own words. It is possible that an author may decide to withdraw the document, rewrite and submit again.

11.2 Substantial overlap in text:

If substantial sections of a document are showing high levels of similarity, there are two options for documents that have not been published yet:

- Reject the document (see guidelines and flowcharts below for further action);
- Send the document to a selected panel of reviewers for additional feedback. Based on their feedback the decision could be to:
  - Reject (see guidelines and flowcharts below for further action)
  - Ask authors to rewrite the duplicated text.

If the document is rejected, the findings should be presented to all authors (if all e-mail addresses are available) or to the corresponding author and s/he should be asked to respond on behalf of all authors. Please follow the flowcharts designed by COPE on what to do:

What to do in case of redundant (duplicate) submission?

What to do in case of plagiarism in a submitted paper?

For severe plagiarism cases (for example, plagiarism by the same group of authors affecting multiple Springer Nature journals or journals (or book publications) from other publishers, or other unethical practices), you are advised to inform your publishing contact before making a final decision.

There is a chance that you might come across a document already published that shows similarity with previously published material or that an (anonymous) whistleblower points you to a case of plagiarism or redundant publication. These concerns should be investigated, and if found to be valid, please follow the flowcharts designed by COPE on what to do for published documents.

What to do in case of redundant (duplicate) publication?

What to do in case of plagiarism in a published paper?

Note. The flowcharts are focusing on journal articles, but the principles also apply to books and proceedings.

Important note! If your investigation leads to the decision to correct the literature, either by an Erratum, Retraction Note, or Expression of Concern, please contact your publishing contact first.

11.3 Downloading a PDF version of a similarity report

A printer-friendly PDF can be downloaded by clicking on the printer button in the left-hand lower corner. The PDF depends on the view, e.g. the 'Match overview' or 'All sources' (see also section 3 The Document Viewer Report).

11.4 Sharing results with external parties

Sometimes it is necessary to share results with external parties, for example to solicit a response or explanation. **Full reports should not be shared.** Experience shows that sharing full reports can give a distorted view of the actual matches. The idea is to only share the most relevant match(es) via screenshots to avoid confusion.

12 Queries

If you have questions that are not addressed by this document, please send your queries to: **training@springernature.com**.

References

- Crossref Plagiarism Screening: Understanding the Similarity Score. **http://www.ithenticate.com/plagiarism-detection-blog/bid/63534/Crossref-Plagiarism-Screening-Understanding-the-Similarity-Score#.VBlv8JR_saA**
- What is Crossref? **http://www.iThenticate.com**
- **http://www.ithenticate.com/products/dv-beta-notes/**
- iThenticate User Manual. **https://www.ithenticate.com/hs-fs/hub/92785/file-1384442395-pdf/iTh_documentation/iThenticate_Manual.pdf?t=1462984332970**
- IEEE. Navigating the Crossref/iThenticate Site. **https://www.ieee.org/content/dam/ieee-org/ieee/web/org/pubs/IEEE%20CrossCheck%20Portal%20Guide.pdf**
- Committee on Publication Ethics. **https://publicationethics.org/resources/flowcharts**
- Wager, E. (2011) How should editors respond to plagiarism? COPE discussion paper **http://publicationethics.org/files/COPE_plagiarism_discussion_%20doc_26%20Apr%2011.pdf**