

iThenticate

v2

[iThenticate](#) is the leading provider of professional plagiarism detection and prevention technology used worldwide by scholarly publishers and research institutions to ensure the originality of written work before publication.

The Similarity Report

The *Similarity Report* provides an overall similarity breakdown for each submission to the *iThenticate* database. This breakdown determines the percentage of similarity between a submission and [content](#) existing in the database of the text comparison tool, *iThenticate*.

Interpreting the Similarity Report

iThenticate **does not** check for plagiarism in a submission. It checks a submission against the content database, and if there are instances where the submission's content is similar to, or matches against, one or more sources, it will be flagged for review. The database includes billions of web pages (both current and archived content), a repository of work that has been submitted to Turnitin in the past, a collection of documents, which comprises of thousands of periodicals, journals, publications, and *CrossRef* member content.

It is perfectly natural for a submission to match against sources in the database. If the submission has used quotes and has referenced correctly, there will be instances where there will be a match. The similarity score simply makes the user aware of any problem areas in the submission; *iThenticate* should be used as part of a larger process, in order to determine if the match is or is not acceptable.

The OVERALL SIMILARITY percentage is the *Similarity Score*. To calculate the *Similarity Score*, *iThenticate* scans your submitted document's text, and checks it against each of the repositories you've chosen. The system takes the number of matching words found within the document and divides it by the document's total word count to produce the *Similarity Score* percentage for the report.

The paper text is viewable with the instances of similarity highlighted. The sources are listed on the top of the screen, color-coded and listed from highest to lowest percentage of matching word area within the submission.

Similarity Reports and preprints

It is entirely possible (and acceptable) for an author to submit an article to a journal even though they've previously made the article available as a preprint. In this case, we expect a high degree of similarity between the preprint and author's submitted manuscript.

Therefore, if you find a high degree of similarity between a manuscript, you're checking in *iThenticate* and a preprint by the same author(s), this is likely to be because the manuscript is a match with its own preprint. However, if the manuscript and preprint do not have the same author(s), this may indicate a problem, and you should investigate further.

Report FAQ

How is the similarity Index calculated for each of the reports?

The similarity index does not change when viewing different report modes. To calculate the *Similarity Score*, our system first makes a digital fingerprint of the submitted document's text and searches it against each of the repositories selected to be searched against within the system. Then once the document is scanned against the selected content repositories our system takes the # of matching words our system found within the document and divides it by the document's total word count to produce the *Similarity Score* percentage for the report.

If exclusion options are applied to the document our system would recalculate the similarity index percentage and remove all matches that the exclusion option logic excluded from the report.

How is the Similarity Report constructed? What matches get hidden by other matches?

As much of the information surrounding the *Similarity Report* is proprietary, the following is a brief description of how the report is constructed. A document submission is converted into what we refer to as a digital fingerprint that our system then uses to search against our content databases using our proprietary algorithm. Our system then paints the document with highlights for each section of text that matches a source within our repositories. What our system deems to be the best matches to the sections of text are listed in the report sidebar. Although the best matches are listed there may be hundreds to thousands of other sources that match the document's text.

What if two sources have the exact same amount of matching text; which source would be displayed in the Similarity Report as a best match?

This entirely depends on which repository the document matched to. For example, if two internet sources were found to match the identical section of text, the most recently crawled internet source would be displayed as the best match. If an internet source and a publication source were found to match an identical section of text, the publication source would be displayed as the top match.

How to interpret the Similarity Report

iThenticate does not check for plagiarism - it checks for similarity. Where a section of the submission's content is similar or identical to one or more sources, it will be flagged for review. This doesn't automatically mean plagiarism, however - just similarity.

It's perfectly natural for a submission to match against some sources in the database. A high degree of overlap may indicate a well-researched document with many references to existing work, and as long as these sources are quoted and referenced correctly, this is perfectly acceptable. A high degree of overlap may also be present where an author has already shared their work on a preprint repository. If the author(s) are the same, this is not a problem.

It's important that you don't set a *Similarity Score* over which you automatically reject manuscripts - where there's a high degree of overlap it's important that your editors and reviewers decide if the match is acceptable or not, as part of their general review process.