

From the Editor

## What is similarity check?

---

**Nygren P.** (2021). What is similarity check? *Silva Fennica* vol. 55 no. 5 article id 10682. 2 p.  
<https://doi.org/10.14214/sf.10682>

---

Silva Fennica, like all journals with a rigorous review process, submits all received manuscripts to the similarity check before peer-review. Earlier, similarity check was called plagiarism check. The old term is inadequate for a couple of reasons. First, plagiarism is a scientific misconduct and nobody should be blamed for a scientific misconduct without a proper investigation. Second, most similarities observed with published work are bona fide i.e., unintentional or justified. In this editorial, I briefly describe how similarity check is done in Silva Fennica.

When a manuscript is received in the manuscript submission and peer-review system either Managing Editor or Editor-in-Chief submits the manuscript to iThenticate software. iThenticate is a text comparison software. All it does it is to compare the new manuscript against a huge database of scholarly material. iThenticate identifies exactly matching strings of text in the manuscript and its database. The results are displayed on the new manuscript. A link to the similar text is provided for each text portion identified. It is important to note that iThenticate is not an artificial intelligence (AI) application. The software makes neither any recommendation on acceptance or rejection of the manuscript nor identify true plagiarism.

The results displayed include the overall similarity percentage of the manuscript with scholarly works in the database and the percentage of similarity with individual articles. In Silva Fennica, we do not apply any threshold for declining the manuscript but all results are inspected by the Managing Editor or the Editor-in-Chief. In the evaluation of the similarities, we pay attention to the points detailed below.

First, a high overall similarity percentage (> 20%) is a red flag to carefully inspect the document. Very low percentage,  $\leq 5\%$ , means often green light for going forward with the peer-review. Much more important than the overall percentage is an outstanding single source. If a manuscript has overall similarity of 20% divided across tens of scholarly works – nothing uncommon – there is probably no major copying problem. However, 5% similarity with a single source may mean hundreds of words in common and we have a good reason to carefully inspect the document. We often visit also the original article, with which the similarity appears.

Second, if a manuscript is identified to need a closer look, we check if the similarity is concentrated in a few paragraphs or if it is evenly distributed through the manuscript. Similar paragraphs are interpreted to indicate copying. One or two similar paragraphs may be hidden under a low similarity percentage. Thus, checking for potentially copied blocks of text vs. a few words here and there is an important criterion for our decision.

Third, we consider the part of the manuscript the similarity occurs. We feel that it is not necessary to reword the description of a methodology used by other scientist or the authors themselves in an earlier work just for avoiding flagging by a similarity check software. Equally, if same data is used in another article, it may not be reasonable to reword the description of data collection or the study site. Similarities in the results and, especially, in the discussion section are evaluated more critically. What is the real novelty if the results are described with the words of another study? Why the author has copied blocks in the discussion? Although the discussion section connects the results with existing knowledge, it must be written with own words. Copying from the work of other scientists raises the suspicion that the authors do not really understand the meaning of their work.

Fourth, we check if the similar source is referred to. Using the same methodology as other scientists is fully appropriate but the source must be credited. This does not mean crediting only the original, perhaps even classical, work in which the methodology was first presented. It is also important to refer to the article, which the authors use as a guidance for applying or developing their methodology. Checking the original work indicated by the software is good help for evaluating these cases. In principle, even small similarities without reference are not accepted while even similar paragraphs with adequate reference(s) may be accepted – depending on how they fulfil our other criteria for justified similarity.

Thus, the similarities indicated by the iThenticate software are inspected case by case using these four criteria. So far, we have not detected any case of suspected bold plagiarism. We have, however, declined a few manuscripts because of strong similarity. A review article manuscript that borrows verbatim text from tens or even hundreds of original works without any critical evaluation of them and synthesising discussion will be declined without peer-review. We feel that the reviewers do not need to use their precious time for such a manuscript as they would recommend declining it.

We have also declined a manuscript that was based on authors' own conference publication, a full article in the proceedings. They had included additional, new data, which was not in the proceedings. However, even after adding data, the conclusions were word-to-word the same as in the proceedings article. The overall similarity was exceptionally high. We would have considered the manuscript had the new data modified the discussion and conclusions. We have declined without peer-review also a manuscript with the stated aim to synthesise the results of a large research project published in several original articles. Unfortunately, the manuscript was written with copy-paste technique from the original articles showing > 10% similarity with several of them. The real synthesis was missing.

True AI applications for similarity check probably are not just around the corner. Thus, a human decision must be made after a text comparison software indicates the similarities. We do our best to ensure a fair evaluation of all cases flagged by the software. For the same reason, we look also the manuscripts not flagged. Usually, unflagged are soon reviewed. Flagged manuscripts require more work but we feel that the just treatment of all manuscripts requires the human work.

*Pekka Nygren*  
Managing Editor  
Journals of the Finnish Society of Forest Science