Ouriginal

Ouriginal Metrics

Verify the authorship of a document and prevent contract cheating

Ouriginal's Metrics takes stylometry to the next level. By using pre-defined quantitative parameters, the software analyzes different writing styles and patterns which can then be used to verify the authorship of a document and prevent contract cheating or ghost writing.

Why is this important?

In addition to safeguarding academic integrity, Ouriginal Metrics encourages and helps develop critical thinkers and creative writers. By providing users with the ability to check that the content submitted has actually been authored by the submitter, they can be deterred from taking any short cuts especially in the form of contract cheating, and instead develop their own views and thoughts.

How does Metrics work?

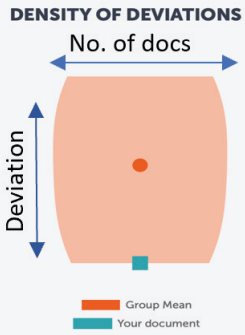
Linguists and researchers believe that writing styles of people in particular, students, in the same class, course or age-group, tend to be organized around a common base or denominator i.e., writing styles within groups have similarities. This assumption is referred to as the 'peer-group similarity hypothesis.' Using this hypothesis, Ouriginal Metrics analyzes each individual assignment for a specific group based on a set of parameters. The findings are then compared with that of the whole group. If a student's text stands out from that of the peer group across several parameters, it could be that the student is not the author of that content.

Style parameters explained

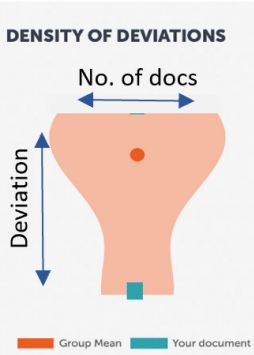
We evaluated more than 50 stylometry metrics out of over 350 and found that the nine described below were best suited for our analysis.

Sentence length - the average number of words per sentence in a text.
Reading level - the number of years of formal education needed to understand a text on the first reading.
Formality - how formal a text is.
Proportion of rare words - the percentage of words that appear only once in a text. This value is also known as 'Hapax Legomena'.
Speech diversity – speech diversity calculated by the 'Simpson's Index of Diversity'.
Punctuation Frequency - the average count of commas per 1,000 words in a document.
Character distribution - this feature evaluates how distant texts are from each other based on characters.
Lexical originality - the percentage of unique words in a document out of the total number of words in a set of texts.
Word length vectors - compares the count of word lengths between documents.

Example analysis



- Your document is represented by the blue square and the group mean by the orange dot.
- You may think your document is an outlier in this instance.
- But a result like this is no cause for concern.
- This is because almost all submitted documents – represented by the orange area - keep the same structure. The orange figure is almost formed as a square.



- Your document is represented by the blue sqare and the group mean by the orange dot.
- A result like this is a warning for two reasons:
- The green symbol if far away from the red.
- The orange figure does not form a square.
- Thus, implying that this document is very different from that of the rest of the group.

Languages supported

Ouriginal Metrics currently works for submissions in English, Swedish, German, Spanish, French, Portuguese, Dutch, Italian and Norwegian.
We are working to increase our language suite and will soon be adding more.