

# INSTRUCTION FOR THE INTERPRETATION OF THE SIMILARITY REPORT OF THE SYSTEM PLAGIAT.PL

## Similarity Report:

- Facilitates the originality evaluation of the analyzed text;
- Determine in which percentage the document was compiled or copied;
- Indicates the number of fragments borrowed;

## 1. Similarity Coefficients

- Determine whether the document consists of fragments identical to those found in other texts;
- Indicates the percentage of the number of words found in other texts compared to the total number of words in the analyzed document;

## Interpretation of the Similarity Report

Title:	XXXXXXXXXXXXXXXXXXXX
Author:	XXXXXXXXXX
Coordinator:	XXXXXXXXXX
Report date:	2017-11-16 09:44:29
Similarity Coefficient 1: ?	<b>98,8%</b>
Similarity Coefficient 2: ?	<b>98,8%</b>
Similarity Coefficient 3: ?	<b>98,8%</b>
Similarity Coefficient 4: ?	<b>98,8%</b>
Similarity Coefficient 5: ?	<b>0,0%</b>
Phrase length for Similarity Coefficient 2: ?	<b>25</b>
Number of words:	81
Number of characters:	652
Skipped URL addresses:	

1. The value of Similarity Coefficient 1 determines which part of the documents contains phrase of 5 words or longer, found in the database of the university, other university affiliates, RefBooks database or Internet resources (excluding borrowings from legal acts found in the Legal Acts Database). The Similarity Coefficient 1 is primarily used to examine the linguistic independence of the author of the documents.

2. The value of Similarity Coefficient 2 determines which part of the documents contains phrase of 25 words or longer found in the databases mentioned above (excluding Legal Acts Database). Due to the length of the detected phrases, the Similarity Score 2 is a better tool for detecting unauthorized borrowings.

3. Similarity Coefficient 3 is calculated analogously to the Coefficient 1, with the addition of all 8 word or longer phrases found by the System in the Legal Acts Database.

4. Similarity Coefficient 4 is calculated analogously to the Coefficient 2, with the addition of all 25 word or longer phrases found by the System in the Legal Acts Database.

5. Similarity Coefficient 5 indicates the percentage of the analyzed document that is formed exclusively from phrases of 8 words or longer that are found in the Legal Acts Database.

## 2. Similarity Report Functions

- Drop-down list with the longest 10 fragments that have been identified as similar to fragments found in the source documents;
- Lists of documents containing fragments identified as identical to the analyzed text, documents which are found in the university's database, the databases of other universities, RefBooks database and the Internet;
- The possibility of highlighting the loans discovered;
- Determining the percentage of similarity in a specific fragment.
- Determine the number of fragments that come from a single source.

The screenshot shows a web interface for a similarity report. It features several sections with expandable/collapsible headers:

- Longest Fragments Identified as Similar**: A red header with a double right arrow.
- Documents Containing Similar Fragments: From RefBooks Database**: A red header with a double right arrow and an information icon.
- Documents Containing Similar Fragments: From the Database of Legal Acts**: A red header with a double left arrow. Below it, a note states: "Documents marked in bold font contain allegedly plagiarised fragments exceeding the Similarity Coefficient 2 limit in length".
- Documents Containing Similar Fragments: From Home Database**: A red header with a double right arrow.
- Documents Containing Similar Fragments: From the Internet**: A red header with a double left arrow. Below it, a note states: "Documents marked in bold font contain allegedly plagiarised fragments exceeding the Sim...".

The main content area displays a table with columns: "No. Legal Act Title", "Source", and "Number of identical words (number of fragments)".

Two callout boxes provide additional information:

- Callout 1** (pointing to a dropdown menu): "Drop down lists with text excerpts which were identified as similar".
- Callout 2** (pointing to a table cell): "Information about number of found words which occur in given number of text excerpts identified as similar".

Another callout box points to a table cell: "Possibility to mark by blue color text excerpts from given source. Helpful to find given text excerpt and check originality".

The table shows two entries:

No.	Legal Act Title	Source	Number of identical words (number of fragments)
1	Protokoły dodatkowe do Konwencji genewskich z 12 sierpnia 1949 r. o międzynarodowych konfliktach zbrojnych (Protokół I) oraz dotyczący międzynarodowych konfliktów zbrojnych (Protokół II). Genewa.1977.06.08 <i>REPORT_LEGAL_ACTS_DATABASE</i> <i>REPORT_WOLTERS_KLUWER_ACKNOWLEDGEMENT</i>	Wolters Kluwer Polska 1992 r. Nr 41, poz. 175	104 (6)
2	Międzynarodowa konwencja dotycząca praw i obowiązków mocarstw i osób neutralnych w razie wojny lądowej. Dz. U. z 1927 r. Nr 21, poz. 163 Haga.1907.10.18 <i>REPORT_LEGAL_ACTS_DATABASE</i> <i>REPORT_WOLTERS_KLUWER_ACKNOWLEDGEMENT</i>	Wolters Kluwer Polska	31 (2)

The interface also includes a "Clear markings" link and a magnifying glass icon for each row.

### 3. Similarity Report Content

Description of color that appear throughout the Similarity Report:

- **Green** – fragments detected on the Internet sources around the world;
- **Red** – fragments detected in databases: internal and external (eg. Other Universities);
- **Orange** – fragments detected in the RefBooks database;
- **Blue background** - fragments detected in the Legal Database;
- **Yellow background** - characters identified by the system as being from non-latin alphabets.

Content of Similarity Report with marked fragments which identified as similar

który stwierdził, że **dalszy atak wojsk rosyjskich na Gruzję będzie miał „znaczący i długoterminowy wpływ na stosunki rosyjsko-amerykańskie”**. Należy przy tym zauważyć, że choć **stanowisko USA w sprawie konfliktu Kaukaskiego** wypływało z różnych stron, było przy tym spójne, **klarowne i jednoznaczne. Opinia publiczna** otrzymała czytelny sygnał, że **stanowisko amerykańskie jest krytyczne, zaś administracja Georga W. Busha stanowczo domaga się wycofania wojsk**

### 4. Interpretation of the Similarity Report

To make a precise interpretation of the Similarity Report, the following steps must be taken:

- Determine the values for the similarity coefficients (it is considered suspicious if the first Similarity Coefficient exceeds 50% and if the second Similarity Coefficient exceeds 5%);
- Examine the list "Longest Fragments Identified as Similar" (fragments which have a number of words bigger than 200 are considered suspicious, therefore they require a rigorous verification; if this kind of fragment exists, it must be found by using the link "Mark fragment" and to verify whether it is cited or not);
- Scroll the lists "Documents Containing Similar Fragments", taking into account especially the documents containing fragments that exceed the Similarity Coefficient 2 (these are marked in bold). In the you have such documents, and especially if they are at the top of the list, you must use the link "Mark fragment" and verify whether they are short phrases scattered throughout the document (in this case we can consider them as

coincidental loans) or long text fragments that are adjacent and separated only by short phrases (this kind of situation will arouse suspicion);

➤ If any suspicions arise, the Similarity Report should be subjected to an accurate analysis, which is based, in addition to using the Similarity Report functions presented in paragraphs 1-3, on the exact analysis of the content of the document, taking into consideration the fragments that were discovered by the system in other texts.

## 5. Basic information about the interpretation of the Similarity Report

**Plagiat.pl is a tool for verifying the originality of checked documents.** Its purpose is to determine the exact proportion of the possible similarities of the verified text in comparison to the content of the databases and the Internet.

The system provides the above information, allowing an independent assessment on the legitimacy of borrowings found in the verified document. The purpose of the system is not to declare if the text was written independently or not, but to provide the needed materials in order to form an opinion regarding its originality. Therefore, the Similarity Report should always be examined by a competent person. In particular, a document must not be assessed based solely of the percentages of the Similarity Coefficients. It is necessary to check the documents content - if quotations are marked and if they come from documents listed under references.

**Plagiat.pl does not determine which document was created first** - the one analyzed by the system or the one found that is considered source. In case of doubts, the user can not establish, solely based on the Similarity Report, which of the documents is the original and which was copied. This conclusion can only result from a detailed analysis of both documents.

Due to the methods used to analyze similarities, the system will also detect certain phrases that are often used, such as: "As I have mentioned before" or "We are able to conclude that". However, the number of such phrases in the group of loans and their impact on the percentage similitude coefficient value should not exceed 50%.

The analysis conducted by our company show that when texts contain a large number of phrases taken from a professional field, similarity coefficient increases significantly. As a result, some documents may obtain relatively high percentages, although a detailed analysis of the Similarity Report will show that it does not contain unauthorized loans. For this reason we introduced the second similarity coefficient. Similarity Coefficient 2 defines a more accurate percentage of loans found in the analyzed documents. Its value is calculated as the Similarity Coefficient 1, but it identifies phrases that have a certain number of words, which is set by the university. Our recommendation for the universities we work with is to set the limit for Similarity Coefficient 2 up to 25 words. Also, we recommend a detailed analysis of reports which have Similarity Coefficient 2 more than 5%. Documents

that exceed permissible coefficients of similarity percentages cannot be automatically considered cases of plagiarism.

## 6. Glossary of useful terms

**Alert** - notification of Similarity Report showing the characters from non-latin alphabets appeared in the verified documents. Alerts exist to attract attention of the coordinating teacher on the use of characters, because they may be an attempt to forge the coefficients from the Similarity Report.

Documents containing alerts are highlighted in yellow in the list of documents and Similarity Reports of these documents are marked with an exclamation mark.

**Databases Exchange Program** - an option that Plagiat.pl offers to Universities so they can get access to the databases of other units participating in this program, thus increasing the number of sources for plagiarism analysis.

**Similarity Report** - a document generated by Plagiat.pl , which contains information on loans identified in the analyzed text.

**RefBooks** - a database created by our company which contains publications and texts from all fields of science and culture, which is the basis for the investigations made by the antiplagiarism system. It consists of many sets of texts provided by authors and publishers exclusively for antiplagiarism analysis.

**Legal Acts Database** - contains a multitude of texts in the legislative field, including Polish legislation, European legislation, magazines, jurisprudence and others. The Legal Database is updated daily with new acts that come into force.

## 7. Contact:

If you have any problems, please contact our Customer Service Department,

E-mail: [pomoc@antyplagiat.pl](mailto:pomoc@antyplagiat.pl)

Phone: (22) 100 11 11