

# “Counter plagiarism detection software” and “Counter counter plagiarism detection” methods.

Yurii Palkovskii

SkyLine, Inc.

Ukraine, Mayakovskogo 15, 66 p.code: 10014

palkovskiy@yandex.ru

**Abstract:** This article describes the basic principles of counter plagiarism detection software, its methods and the countermeasures that can be applied in order to effectively oppose such software.

**Keywords:** plagiarism detection, plagiarism, counter plagiarism detection software, plagiarism detection methods.

## *1 Plagiarism detection as a complex problem in Higher Educational Institutions*

### **1.1 The necessity of plagiarism detection complex solutions in Ukrainian education**

Plagiarism detection and prevention becomes an important issue in Ukrainian higher educational institutions. There exists a number of reasons that deepen the academic dishonesty problem. The rapid spread of information technology and the Internet in particular has reached the maximum degree only during the recent years of 2004-2008. Multiplied by the insufficient experience of the teaching personnel in relation to the Internet technologies and plagiarism detection methods in particular, it resulted in lack of control over the students' written assignments quality. This latest trend caused the development and introduction of different plagiarism detection and prevention software solutions and their wide integration into the study process in different educational establishments all over the country.

## *2 “Counter plagiarism detection software” and “Counter counter plagiarism detection” methods*

### **2.1 “Counter plagiarism detection software”**

One of the most interesting aspects of the plagiarism detection on the territory of the former USSR is the fact that general students' IT competence is much higher than that of the teachers. This caused the number of cases of

special software being developed by students as a countermeasure to fight plagiarism detection services and software solutions. One of the examples to mention is: “AntiPlagiatKiller”. The exact title of the software in Russian is “антиплагиат киллер”. It is a special text post-processing application that uses effective algorithms that “shuffle” or otherwise scramble the target text using different methods to render the fingerprint search useless by “breaking up” the signatures that are generated for indexing and search.

### **2.2 Counter plagiarism detection algorithms**

Some principles used in: “AntiPlagiatKiller” are similar to the ones used in the “Random Plagiarist” that was developed for the PAN09 plagiarism detection competition to provide obfuscated plagiarism sections, some of these algorithms are generic and some are relevant to Slavonic languages only. For example – the so called “Cyrillic to English substitution”. The particular method implies a substitution of some or all of the particular character or characters from the native text alphabet for their Latin equivalent that has the same visual shape. This method breaks the fingerprint signature and remains visually undetectable if the investigation is done by a human reviewer. For example:

Russian “o” – Latin (English) “o”  
Ukrainian “x” – Latin (English) “x”

Example: **T**oast

The letter “o” in the word “Toast” is not English, but Russian.

This particular method can be implemented and is massively exploited in a particular software environment, for example Microsoft Word. It cannot be implemented in plain text and different document formats that does not support the required functionality.

This applies to a number of Languages that share the same graphical alphabet (or partially share): Russian, Ukrainian, Belarusian and some others. This method can be potentially applied to any relevant languages that share at least a pair of visually identical characters.

The other generic method that is widely used by the modification of “anti-plagiarism-killer” is even more original, it is called – “white link-character insertion”. The idea of this obfuscation method is simple and effective against the automated plagiarism detection systems and again visually hard to spot – a space between any words is filled with a single character or a pair of random characters that are later colored in white. This completely distorts the semantic structure of the sentence but remains visually undetectable. To illustrate this method the following example can be shown:

Original text: **I like apples**  
Scrambled text: **I like apples**

The scrambled text contains character “x” (and for the sake of the example it is colored in grey instead of white) as a “white link-character” that is inserted between the words. It has white as its background color that makes it absolutely invisible to visually detect.

The appearance of such software and its wide usage among students forces the developers of the plagiarism detection solutions to present a number of additional algorithms to detect such “shuffling” done to the text and add a special alert to the final originality report to inform the reviewer about the possible usage of the “counter plagiarism detection” algorithms.

The existence of such software points to the necessity of the development more sophisticated plagiarism detection systems. The first generation of such solutions uses the fingerprinting method to detect the cases of potential plagiarism on the basis of the fingerprint taken from the sentence – that is it uses a sentence to count a hash to be later used in the search. The “Counter plagiarism

detection software” is pretty effective in distorting the text to the degree when fingerprints cannot be effectively used any longer and thus the need to use another algorithm is brought to life.

### 2.3 Counter counter plagiarism detection algorithms

The implementation of each algorithm that will detect the cases of deliberate obfuscation is rather straightforward in case the obfuscation mechanism deals with the text meta characteristics – color, character set or encoding. Every implementation of such algorithm must be targeted against the specific obfuscation method with two possible output results – either neutralization of the counter plagiarism detection algorithm to remove the factors that prevent normal analysis or reporting the number and scope of the particular occurrences (or suspected occurrences) in the analyzed text.

References:

1. Antiplagiatikiller download page:  
[http://www.datalive.ru/2007/08/20/anti\\_plagiat\\_killer\\_borba\\_s\\_antiplagiatom\\_postudncheski.html](http://www.datalive.ru/2007/08/20/anti_plagiat_killer_borba_s_antiplagiatom_postudncheski.html)
2. Official reaction to Antiplagiatikiller:  
<http://www.antiplagiat.ru/forum/default.aspx?g=posts&t=223>
3. Internet board that contains the methods on how to check plagiarism detection systems:  
<http://www.antiplagiat.ru/forum/default.aspx?g=topics&f=5>