

Computational methodology Statistical of automated detection of copyright plagiarism - A proposal of the results interpretation of the program PlagiarismCombat

Maximiliano Zambonato Pezzin, MSc
<maximilianopezzin@gmail.com>
Universidade do Contestado - UnC Concórdia - 10/2010

Abstract

This document covers a polemic theme of difficult definition. The complexity and the different variables involved in approaching what literary is plagiarism, end up generating a vacuum precisely on the definition and affirmative. After all, was a text plagiarized or not? This document must be seen as a proposal of definition for plagiarism, not exactly in numbers, but based on the methodology of PlagiarismCombat program. The variables and techniques used by the program are the basis of the author's proposal and should be seen as a proposal, since there are no (it is notable that there are not) legal or juridical aspect involved to the sniffer processing, becoming the reader's responsibility the final definition of the plagiarism existence.

1. Introduction

The brazilian law number 9610 describes in its first article "This Act regulates the copyright, being understood under this denomination the author's right and those connected to them." However, despite being complete when it comes to defining literary works, the contextualization and definitions presented are, in certain way, dodges. It can be said that the law is absolutely vague regarding the formal definition of something that is so dubious: after all, what is plagiarism? How to define what was copied? How to speed up the process of identifying plagiarism?

2. The copyright in Brazil and in the World

Some organizations of a large representation, such as the WIPO (World Intellectual Property Organization) show quite clearly the need to combat plagiarism and conduct the defense of copyright. Dr Uchtenhagen in his compendium, focused on cultural and musical works, comments: "The document Provides short Explanations, on the different steps, and conditions on the various parameters indispensable for the creation of a collective management organization for musical works. It includes an

interesting plan Indicating work in a team frame the different stages Necessary for Achieving such goal. "

These words, originally directed to the musical copyright, are perfect in this context, once a musical work and a literary work are, in principle, equivalent towards the creation and property intelectual. Thus, one can extend the idea of copyright for the various areas of human creation. These studies are the philosophical bases worked and defended by WIPO, an organization maintained by ONU that has as a goal to ensure the intellectual rights in all spheres, models and instances.

The defense of copyright is also defended by many researchers from our tupiniquins land (related to Tupiniquins, ancient nation of Brazilian Indians), as MENDES, which relates the emerging economies with the need for research and development, noting that more than mere technology users, we fulfill our mission of generating knowledge and innovation with ideas .. " Over the last 20 years, emerging economies have generally benefitted from global trends in research and development (R&D) which have helped to boost national development and nurture broader international cooperation. An emphasis on innovation, its promotion and associated intellectual property (IP) aspects are key features of any policy that seeks to effectively promote economic growth and development. ".The partner WIPO vision is clear, stating that we should act creatively, promoting and defending the IP (Intellectual Propriety) in order to generate a culture that evolves with a sustainable growth.

The copyright is treated in different ways in the various academies of science, around the four corners of the world, and a simple analysis indicates that the respect for intellectual property is as greater as the development degree (not necessarily academic) of a population . Developed countries tend to provide greater assurance of IP.

Accordint to ROMANCINI, plagiarism is characterized as a false attribution of authorship, a misappropriation of the work of one author by another individual (the plagiarist). The copy of ideas or content of work from another person. It is interesting to note that the etymological origin of the word (from the Greek "plagiarism" to Latin "plagiu") carries meanings that illustrate the concept: "oblique", "sly", "cheater".

In France, plagiarism is considered a crime, being applied criminal and civil penalties, as it is clearly presented and explained by Poirier, on his website: "Pour ce type de procédure, il est Preferable s'adresser of the un avocat, ce qui est dans la procédure civile obligatoire, devant le Tribunal de grande instance. La contrefaçon peut donner lieu à des sanctions et penales the sanctions for Civil (paiement of dommages-intérêts)." So, as defined by the Ministère de la Culture, is a police matter, and it should be treated this way.

According to the Brazilian lawyer, Rodrigo Moraes, "... the first specific Law on Copyright took effect in 1710, in England, during the period of Queen Anne (Statute of

Anne), and aimed to protect literary works. It was named as Copyright Act. That was the origin of the term copyright, used today in the English language countries ... "

Surfing and reading the words of defenders of the 'creative writer', in a brief search on Croatian texts, it is possible to find the professor Pažur , who, with quite property tells us, in Croatian: "S pojavom masovnog tržišta za tiskanu i kasnije snimljenu i emitiranu riječ, autori postaju značajan ekonomski čimbenik u trgovini¹. Pravni balans prebacuje se s izdavača na autora čemu su posebno pridonijeli međunarodni zakonski akti počevši od Bernske konvencije iz 1886", which, with a brief translation "... With the advent of the mass market for printed and recorded and later the transmitted word, the authors must consider the economic factors significant to its production, being attributed a legal value to the author itself, which are attributed based on international juridic acts since the Berne Convention of 1886 ... "shows that it is not only a today's concern with the subject, intellectual property.

It must be emphasize and enhance some actions, such as proposed by ABNT, with the normalizing standard, regarding quotes (NBR 10.520: 2002), which are very welcome, as they tend to facilitate and define models of representation of the primary sources, which ends up exposing plagiarism and highlighting the lack of citation of the source.

Finally, we make it clear that the fight against plagiarism and the intellectual property protection occurs in all societies, being more visible and obvious where the concern for education and the intellectual and creative development are placed as a priority.

3. The methodology of the Sniffer

The Plagiarism Sniffer program - in portuguese Farejador de Plágio (<http://www.farejadordeplagio.com.br>) , gently described by many as FDP, is a tool created in 2006, which has thousands of users who have performed the analysis of around 520,000 documents. The principle of the Sniffer is quite simple. Based on reading a document and some basic user settings, it runs sequential scans of continuous sections of the document on search websites. At the end of the search, several data processing techniques are applied in order to list what is or is not copied from the internet.

3.1 Variables of the program

The program has the document to be analyzed as the input of data, the definition of the search websites used, and the query pattern, which can be: Quick, Normal, Detailed and Rigorous.

When it comes to the output variables we can set them as the analysed document, a table of links found (presented on ABC curve), and statistical data:

percentage of participation of the most used, number of suspicious areas, number of suspects excerpts and total number of occurrences in each browser used.

The output variables are generated by processing routines, and must only be analyzed and interpreted after the conclusion of all processing. The variables themselves are:

Searched Excerpts	Total number of searched excerpts, each excerpt is sent to each search website chosen to be used.
Similar Websites	Total number of results found in all search websites, in all searches performed.
Search Websites	The number of results of each search website is presented separately.
Suspicious Areas	A suspicious area is pointed out whenever that research sequences return the continuous presence of similarities. Generally a higher incidence of plagiarism is associated with areas of continuous plagiarism. Another factor that generates suspected areas is permuted plagiarism, where phrases are alternated. The way to analyse of the program facilitates and easily exposes these occurrences.
Suspicious Sites	All sites which match more than 4 times are shown as possible sources of plagiarism. If it matches 4 times, and is on a suspicious area, it will also be shown.
1 to 2 most used	This factor (no index) indicates the percentage of 2 sites that most happened in relation to the total number of matching records indexed in the search. (See 4.3.3).
1 to 5 most used	This factor (no index) indicates the percentage of 5 sites that most happened in relation to the total number of matching records indexed in the search. (See 4.3.3).
1 to 10 most used	This factor (no index) indicates the percentage of 10 sites that most happened in relation to the total number of matching records indexed in the search. (See 4.3.3).
Pages / minute	Scan Rate / searches per minute, which indicates the speed of each research.
Confirmations / minutes	Confirmation rate, high rates indicate a higher incidence at a specific excerpt. The rate is NOT linear, what means that if there are a lot of occurrence in an area, the rate will be higher.

The partial analysis of the data during processing may give a distorted view of the document, it is suggested, always wait for the end of the process.

3.2 Functions of Processing

The routine of processing the sniffer is based on a brute force processing model. All information is processed internally in the analysis computer and need a strong access to WEB. Each information found is managed and stored in data structures in the computer's memory, and at the end of the process, the data is organized and presented.

The management, storage and comparison performed in real-time, associated with the control of various objects of programation, the interaction with the various research websites, the data interpretation of websites and, finally, the application of ongoing heuristics analysis, as it is expected, consumes a considerable portion of the computer's processing.

The processing occurs in the classical form of an information system: input variables, processing and generation of output information. What differs the FDP is

precisely the dynamic character of the program's processing, with high interaction with internet, allows that dynamic data interfere directly in the results of the program.

3.3 Management of Variables Processing

The program, along the process, collects data and stores information in variables of processing. Considering a 100-page document, with search in NORMAL mode, will generate about 2000 research that, using 8 searchers, with 10 responses per searcher will reach a maximum number of 160,000 indexed results.

All these information must be stored in memory, storing information of location, excerpt, website, order and search website. To get an idea of the volume of information, and amount of memory used, each indexed outcome could use up to 200 bytes, which would entail the use of more than 3 Gbytes of data. This shows the large amount of data to be managed, and the difficulty in manipulating this data in a non-computational environment.

3.4 Heuristic of Search and data organization

Among the various approaches of solution implemented in the program, the processing with brute force is perhaps the main feature of the program. The case is that the analysis and search of plagiarism requires that the entire document must be read, and sporadic or random searches could fail by ignoring a stretch that is copied.

Considering the text as a sequence of words, the program will do sequential searches, such as the teacher would do, by copying a part of a sentence and inserting into a search website.

In the case of the program, the number of words of each search is defined by the user, when the user chooses settings between FAST, NORMAL, DETAILED and ACCURATE. For each configuration, a search way is set:

	Number of searches per cycle	Number of jumps per cycle
Quick	8 or 9 words	8 to 10 words
Normal	6 or 7 words	7 to 8 words
Detailed	5 or 6 words	5 to 7 words
Rigorous	4 to 5 words	4 to 6 words

In previous versions of the program, the user was allowed to set these values manually, however erroneous settings of the program ended up in processing failures. Thus, the models described above were defined, in order to avoid processing failures and errors. From experience, reports and tests, the pre-configured model is the one with the best results.

3.5 Inference engine and organization of sets

Considering that the analysis of indexed links only makes sense at the end of the verification through the document, the main processing routine will hold the appointment of plagiarism in the analyzed document itself.

However, some partial results can be presented and others suppressed, based on some basic assumptions related to historical information, stored in accordance with pre-processed models. This way several generic and repetitive websites are suppressed from the analysis, reducing the processing load, accelerating and facilitating the study of the final results.

Throughout the process, various control and management information are stored and managed in order to enable the treatment of the input variables and processing. These info combined with the processing of links treatments allow to do notes of any eventual typical coincidences of excerpts intentionally copied.

Some forms of copies can be seen as unintentional, once occur sporadically and randomly. These occurrences are ignored by the FDP, because they are considered hard to prove. They are considered unintentional because occur infrequently, making it difficult hold a claim of breach of copyright.

Computationally speaking, the generation of models is a set restricted to few unconnected elements with no connection that do not generate a model that fits the definition of the use of others' ideas, according to the FDP would not be plagiarism, by absolute definition.

3.6 Fuzzy logic in defining boundaries of response sets

Fuzzy Logic or Fuzzy Theory, is a form of data processing widely used in several areas of science, from medicine to mechanics, metallurgy, pharmacy and accounting.

Among the advantages of this technique, we can highlight the most "humane" treatment of variables, allowing to consider and give partial value of the variables analyzed, not deleting anything near to the income limit, and adding some uncertainty in the variables that are positive, in another words, the limits of the results may not be clear, but should be considered including data that is not within the set of response and sometimes eliminating data considered valid because they are considered as having a lower value, based on values of other sets of response.

At this point one may to ask: How can the program validate data that is outside the relevant rule and exclude other results considered relevant?

The case is that based on other sets, it is possible to define a degree of relevance to the data. One way to explain the model of fuzzy processing of FDP is using as example an election with two rounds. Two candidates detain vast majority, but the results of the minority can influence the final results, and even with some significant

values may change the result, to prevent a candidate obtains 50% needed. Even an amount of 0.6% would have a great importance for a candidate who has made 49.8%.

The fuzzy idea allowed to be evolve the form of analysis of the FDP, and greatly facilitated the work of analyzing the results generated by the program, while allowing constant evolution of the results, to make a cumulative analysis of various results.

4. The propose of interpretation

As has been said, there is no law that defines plagiarism. Even recommendations are unclear. The WIPO makes considerations and appointments on how to handle similar information from diverse sources, but is cautious and afraid to define plagiarism.

Based on the experience of the author, and the various analyzes and the studies conducted, which culminated in the computer modeling described in Chapter 3, what will be presented below is a proposed analysis. I make it clear at this point that I consider unwise to present my analysis as something definitive, complete or that will be used as an ideal model of analysis.

As mentioned initially, it is a mere proposal, which can and should be evolved based on mutual and cooperative experience of other researchers who have combat plagiarism as a flag, and the encouragement of research and innovation as a goal to all their pupils.

4.1 Types of plagiarism detected

Forms of plagiarism detected and proposed by FDP are classified into:

- a) Excerpts: Continuous or fragmented - the suspected areas.
- b) Sparse Sentences in long documents - sparse coincidences.
- c) Webites most used.
- d) Fragments of paragraphs, phrases or sentences
- e) Modified entences with a high degree of similarity
- f) Dispersed text and alternations of sentences order
- g) Use of information on high indexing websites
- h) Similarities in phonetic / syntactic errors in similar sentences

Again I emphasize that I present a proposal, formulated to allow the processing of a computational model of generic heuristic compilation, working with variables, techniques, models and routines presented in Chapter 3.

Based on these rules, it is necessary that the FDP generates indexes and factors that allow the analysis of the forms of defined plagiarism. These values will be the basis for an evaluator that can define the degree of similarity of the document analyzed with other documents found on internet.

4.2 Index and factors of the FDP

The comparison of results obtained by the FDP, after the entire document being scanned, brings the need to use a combined and compiled sparse matrix of terms, excerpts and links, which will allow the creation of a wide and general overview of context of the document, in another words, what do we have on this document that exists in other documents on the Internet?

4.3 What FDP considers as plagiarism

The proposed definition of plagiarism listed below have been set to allow the computational analysis of the document, and to enable the generation of sets of basic rules and analysis.

More than suggesting caution in the use of this proposal, it is requested to be used GOOD JUDGEMENT when analyzing the results, especially when the end result by stating that a paper has been plagiarized, and an academic work is incorrectly tagged as plagiarized.

Bellow it will be presented the eight cases computationally described as plagiarism by the FDP. An excerpt from a random brazilian article which allows to present and expose the eight situations. The excerpt is shown below, with numbered sentences as references to the topics of analysis 4.3.1 to 4.3.8

In order to analyze and clearly demonstrate, a sample text will be used, where appointments and demonstrations will be seen. In the example text, 10 searches occur, highlighted in blue. As in the program, underlined blue has been appointed as plagiarism and *{links}* after the underlined are the possible websites of origin.

Example Text:

[1]Sem hesitar, as obras musicais, fotografias e as audiovisuais, em face da subsistência (programas [2]de computador) e hardwares (máquinas) que www.ldphi.com/artigoluizcarlos.pdf permitem com facilidade seu armazenamento, cópias, distribuição e alterações com fins econômicos no [3]e-commerce por mediação da rede mundial de computadores, www.ldphi.com/artigoluizcarlos.pdf www.tecnologiasa.com.br/tecnologias.pdf por qualquer indivíduo, trazem grandes perdas pecuniárias aos [5]autores de softwares

No tablado científico, há probabilidade de negociação de banco de dados ou textos com resultados de experimentos e [6]pesquisas, em específico, pautados com princípios ativos www.ldphi.com/artigoluizcarlos.pdf usados no preparativo de medicamentos, muito preciosos para a indústria [7]farmacêutica.

Quanto aos programas de computador (softwares), por sua origem digital, isto é, compostos por binários [8]numéricos (0 e 1), desorelo comenta acerca da www.scribd.com/camarotti_souza.pdf facilidade de clonagem, rompimento de códigos de segurança, distribuição, [9]transmissão, armazenamento, etc. Habitualmente, nota-se www.scribd.com/artusi_2008.pdf na internet ofertas de programas de jogos, seja em CD-ROM's – entregues pelos correios – ou, [10]facilmente, posterior a confirmação de pagamento www.ldphi.com/artigoluizcarlos.pdf www.scribd.com/artusi_2008.pdf e fornecimento de uma senha, oferecida via eletrônica.

The various cases of plagiarism detected by the program are now exposed and explained, based on the sample text.

4.3.1 Excerpts: Continuous or fragmented - the suspected areas

Plagiarism in excerpts is the easiest to be detected as well as the most difficult to deny. In this case, the similarity is detected more than 4, 5 or 6 sections of the document analyzed with the documents on the Internet. It is the simplest form of detection, where, at the end of the analysis of the FDP, are verified in processing variables, how often each particular site was found riding a curve ABC and pointing, to each site, how many occurrences of the same, were highlighted.

These cases are highlighted in the analysis of sample text, such as quotes *{www.ldphi.com/artigoluizcarlos.pdf}* *{www.ldphi.com/artigoluizcarlos.pdf}*, in both cases, what we have are positive sequences of coincidences that end up generating suspicion areas, because are sequential excerpts with the same source or sources.

Every suspicious area is worrying, especially if it occurs outside of areas of service and generate more sequences than 3 or 4 links repeated, because it would indicate the copy of a large text area.

4.3.2 Sparse sentences in long documents - Sparse coincidences

This is complementary case shown in 4.3.1 where a source is identified, which is the case of *{www.tecnologiasa.com.br/tecnologias.pdf}* *{www.scribd.com/camarotti_souza.pdf}* where the text matches with websites, not continuously, but at least another 4 or 5 times throughout the document, copied probably excerpts were adjusted by generating discontinued copies.

The failure of this kind of plagiarism is the possibility that, in a given subject, the textual continuity occurs and may generate documents coincidents with other phrases, precisely because of their many similarities regarding the contents.

It is suggested to be careful with these data, especially when the number of occurrences identified in the ABC line is less than 8 times. This is the case of plagiarism nearly impossible to be find manually, and a great contribution regarding the search for the FDP. I consider as a case of indirect plagiarism, which supports and strengthens the direct plagiarism.

4.3.3 Most Used Websites

Based on the ABC curve presented by the program, one should consider that the websites most found (ABC top of the list) are the most likely to appear as plagiarism, just because appears more often.

How to interpret the factor of the most used:

	1 to 2 websites	1 to 5 websites	1 to 10 websites
Up to 10 pages	1% little 3% verify + 6% worrying	1% little 4% verify + 8% worrying	2% little 4% verify + 9% worrying
10 to 30 pages	1% little 3% verify + 5% worrying	1% little 4% verify + 7% worrying	2% little 4% verify + 8% worrying
30 to 60 pages	0% little 2% verify + 4% worrying	1% little 3% verify + 5% worrying	1% little 3% verify + 6% worrying
60 or + pages	0% little 1% verify + 3% worrying	0% little 2% verify + 4% worrying	1% little 3% verify + 5% worrying

I emphasize that these are values reference , and should be excluded the references properly cited and laws, which can often change the percentage. Caution and attention in the analysis of these data is suggested.

These numbers should be treated with caution, that is because in a technical, legislative and conceptual document, the indirect quotes should be excluded, to not generate false alerts of plagiarism, and end up creating false accusations of plagiarism.

This kind of analysis is not indicated for rapid analysis mode with less than 10 pages document, although work in cases of obvious plagiarism, with high percentages of coincidence.

4.4.4 Exchanges of fragments of paragraphs, phrases or sentences

Something quite common, is to copy several paragraphs from multiple sources and perform the exchange of sentences or sentence fragments. Although considerably complicate the connection and detection by traditional, manual methods, is easily detectable by the FDP, especially in its STRICT mode.

Obviously suspicious areas are not generated, but will appear as sparse coincidences, but usually contain the curve ABC program. By experience, it is possible to affirm the predisposition of the writer to actually attempt to trick the authorship, the intention was really make use of other people's ideas, because besides copy and not reference, if manipulated the paragraphs on purpose.

Some programs and websites are proposed including automating the generation of texts based on reference texts, obviously without mention the references. By innocence or lack of intelligence, some people do not consider the mix of phrases of documents as plagiarism, which certainly is a big mistake.

It is also interesting the clear way that the FDP presents the results, since there is a great tendency in these cases of exchange of fragments, repetition of authors

generating sets, which, although unconnected, show the lack of reference to the authors of the ideas.

4.4.5 Sentences modified with a high degree of similarity

This type of plagiarism is much more complex, both in finding it and in its proof, because it is not of the same words, but the same ideas, which, by definition of intellectual property, is what really matters.

The detail is that the level of processing, the subtlety of the adjectives and nouns alternation may go unnoticed, even because search engines have implemented this idea in their searches, so the FDP ends up inheriting the ability to pursue the idea, and not the literal text.

This kind of plagiarism, is considered by this author, as a direct plagiarism, but it should also be considered the fact that changing a few letters or words can be seen by some as plagiarism and by others as a new work.

One way to settle the doubt and allow to know whether or not the plagiarism exists, the use of phrase with a high degree of similarity and using the principle of intellectual property. The FDP searches and has the ability to point sources where the difference between the texts are small or subtle.

The author suggests extreme caution when affirming the plagiarism using only this argument, because even being clear, it can generate controversies regarding the understanding of what really is plagiarism, because it is highly subjective to define it on an idea that has been changed.

The interesting thing is using plagiarism by similarity as a complement to direct plagiarism, to reinforce the idea that a text had been copied. Would be adding strength to the idea of direct plagiarism found, because it would be an indication that there are problems in the development and creation of the document.

4.4.6 Dispersed text and alternations of sentences orders

Very similar to the item 4.4.4, it is quite common that it takes any text and spreads it, placing for example, one or two sentences per sheet over 20 or 30 pages. Thus, we can have a false impression of volume relatively quickly, because even the text getting disconnected, something hardly noticeable because it continues existing thematic concordance, which for some people is seen as normal.

Personally I consider it almost impossible to manually detect plagiarism from a source along several pages, when only one sentence per page is present, but again, for the FDP, this is not a problem, because it works with the indexing of pages and can easily 'grab' this trickery.

As there are no changes in the original documents, I consider this a direct plagiarism, and the repetition for several pages imply in an assertion of bad faith and interest in actually copy the ideas.

4.4.7 Use of information on high indexing sites

An unknown device that goes easily unnoticed is the use of documents that are not at the beginning of the responses of the main indexers, such as Google and Bing. This is because, as the result is not straightforward, and would correspond to page 30 or 40, would eventually go unnoticed, since it is a unknown document, which would generate a false impression that the document is authentic.

The FDP can get these documents, and gives to them the same value of documents on the top list of the search engines. For this, the user just need to access the Internet Explorer, and manually configure the searchers only ONCE so it will return more search pages. The result will be a slower search, however much more efficient, in results and in the precision when pointing plagiarism.

But attention, it is suggested the use of these stricter settings in case there is a suspicion that the work may have been defrauded in this sense, because the response time, depending on the speed of the internet, can be much higher.

Several reports from users point the use of this device and the results are very interesting, altering the final results in each case. If there is a suspicion, it is suggested to change the number of answers from the searchers, as well as the default value of 10 may be returned at any time, even while using the FDP.

4.4.8 Similarities in phonetic/ syntactic errors in similar sentences.

This way of identifying plagiarism is, at least, curious and unusual.

Considered a document that contains several orthographic, syntactic and semantic errors. The results of the FDP can easily check for this error in a document during the analysis, precisely because of the discrepancies that exist between this correct writing and the other one, which contains errors.

The result you get is exactly the pointing of the site based on several failures of writing. This type of plagiarism, besides easy to detect, ends up betraying the author, because the texts contained in other documents, have a high degree of discrepancy at the level of writing quality. This, I consider to be a direct plagiarism, of easy detection and hard to explain.

5. Conclusions

The development experience, maintenance and use of the program over the years, as well as the valuable and indescribable experience exchanged with some masters of the art of teaching, brought me and helped me to create my own vision of what is plagiarism. Something widely considered of little relevance or 'easily fixable', is seen by many, and for me, as an exclusive responsibility of the tutor.

In other words, the student, pupil or academic is there to learn. If this academic ends up committing the MISTAKE of copying, and teachers consider this normal, the irresponsible is not who is learning, because, and here I emphasize, the student is learning. The task of teaching and explaining the mistake is exclusively to the TEACHERS. It is up to the student to seek the easy way to solve the problem. And it is up to the teacher to explain that the easiest way is not always the best way.

Regarding the methodology of defining plagiarism presented: "it is a methodology based on computational techniques that, based on variables and rules of processing, searches and points coincidences, presenting them in order to expedite the analysis of the results and clearly define if a document is actually developed by a particular author, it is suggested that each result should be analyzed separately. "

References (after all, we must mention the authors, correct?)

LEI Nº 9.610, DE 19 DE FEVEREIRO DE 1998. disponível em https://www.planalto.gov.br/ccivil_03/Leis/L9610.htm#art115

WIPO World Intellectual Property Organization disponível em www.wipo.int em 2/11/2010

Uchtenhagen, Ulrich The Setting-up of New Copyright Societies disponível em http://www.wipo.int/freepublications/en/copyright/926/wipo_pub_926.pdf em 8/11/2010

Mendes, Ricardo Camargo Innovation Promotion in Brazil Wipo Magazine September 2010

Romancini, Richard. *A praga do Plágio Acadêmico*. Revista Científica da FAMEC, ano 6, nº 06/2007, ISSN-1677-4612.

Poirier, François Université Paris 13. Disponível em: <http://www.univ-paris13.fr/ANGLICISTES/POIRIER/Plagiat.htm> em 7/11/2010

Pažur, Ivana AUTORI ZNANSTVENIH RADOVA I AUTORSKO PRAVO Institut Ruđer Bošković, Croace