

**XXIV ENCONTRO NACIONAL DO  
CONPEDI - UFS**

**DIREITO E NOVAS TECNOLOGIAS**

**JOSÉ RENATO GAZIERO CELLA**

**VALÉRIA RIBAS DO NASCIMENTO**

**AIRES JOSE ROVER**

Todos os direitos reservados e protegidos.

Nenhuma parte deste livro poderá ser reproduzida ou transmitida sejam quais forem os meios empregados sem prévia autorização dos editores.

#### **Diretoria – Conpedi**

**Presidente** - Prof. Dr. Raymundo Juliano Feitosa – UFRN

**Vice-presidente Sul** - Prof. Dr. José Alcebíades de Oliveira Junior - UFRGS

**Vice-presidente Sudeste** - Prof. Dr. João Marcelo de Lima Assafim - UCAM

**Vice-presidente Nordeste** - Profa. Dra. Gina Vidal Marcílio Pompeu - UNIFOR

**Vice-presidente Norte/Centro** - Profa. Dra. Julia Maurmann Ximenes - IDP

**Secretário Executivo** - Prof. Dr. Orides Mezzaroba - UFSC

**Secretário Adjunto** - Prof. Dr. Felipe Chiarello de Souza Pinto – Mackenzie

#### **Conselho Fiscal**

Prof. Dr. José Querino Tavares Neto - UFG /PUC PR

Prof. Dr. Roberto Correia da Silva Gomes Caldas - PUC SP

Profa. Dra. Samyra Haydêe Dal Farra Napolini Sanches - UNINOVE

Prof. Dr. Lucas Gonçalves da Silva - UFS (suplente)

Prof. Dr. Paulo Roberto Lyrio Pimenta - UFBA (suplente)

**Representante Discente** - Mestrando Caio Augusto Souza Lara - UFMG (titular)

#### **Secretarias**

**Diretor de Informática** - Prof. Dr. Aires José Rover – UFSC

**Diretor de Relações com a Graduação** - Prof. Dr. Alexandre Walmott Borgs – UFU

**Diretor de Relações Internacionais** - Prof. Dr. Antonio Carlos Diniz Murta - FUMEC

**Diretora de Apoio Institucional** - Profa. Dra. Clerilei Aparecida Bier - UDESC

**Diretor de Educação Jurídica** - Prof. Dr. Eid Badr - UEA / ESBAM / OAB-AM

**Diretoras de Eventos** - Profa. Dra. Valesca Raizer Borges Moschen – UFES e Profa. Dra. Viviane Coêlho de Séllos Knoerr - UNICURITIBA

**Diretor de Apoio Interinstitucional** - Prof. Dr. Vladimir Oliveira da Silveira – UNINOVE

---

D598

Direito e novas tecnologias [Recurso eletrônico on-line] organização CONPEDI/UFS;

Coordenadores: José Renato Gaziero Cella, Aires Jose Rover, Valéria Ribas Do Nascimento – Florianópolis: CONPEDI, 2015.

Inclui bibliografia

ISBN: 978-85-5505-054-1

Modo de acesso: [www.conpedi.org.br](http://www.conpedi.org.br) em publicações

Tema: DIREITO, CONSTITUIÇÃO E CIDADANIA: contribuições para os objetivos de desenvolvimento do Milênio.

1. Direito – Estudo e ensino (Pós-graduação) – Brasil – Encontros. 2. Tecnologia. I. Encontro Nacional do CONPEDI/UFS (24. : 2015 : Aracaju, SE).

CDU: 34

---



# XXIV ENCONTRO NACIONAL DO CONPEDI - UFS

## DIREITO E NOVAS TECNOLOGIAS

---

### **Apresentação**

#### APRESENTAÇÃO

No XXIV Encontro Nacional do CONPEDI, realizado na Universidade Federal de Sergipe - UFS, em Aracaju, de 03 a 06 de junho de 2015, o grupo de trabalho Direito e Novas Tecnologias novamente esteve presente com destaque pela qualidade dos trabalhos apresentados e pelo numeroso público, composto por pesquisadores-expositores e interessados. Esse fato demonstra a inquietude que o tema desperta na seara jurídica, em especial nos programas de pós-graduação em Direito que procuram empreender um diálogo que suscita a interdisciplinaridade na pesquisa e se propõe a enfrentar os desafios que as novas tecnologias impõem ao Direito.

Foram apresentados 22 artigos que foram objeto de um intenso debate e agora fazem parte desta coletânea. Numa tentativa de organizar quantitativa e qualitativamente os artigos e seus temas, segue uma métrica:

Cinco artigos trataram da Internet, em diversos âmbitos.

Quatro artigos discutiram a proteção da privacidade e dos dados pessoais e corporais.

Quatro artigos foram sobre responsabilidade civil e capacidade na internet.

Dois artigos versaram sobre aspectos regulatórios das nanotecnologias.

Dois artigos sobre marco civil da internet.

Dois artigos trataram do processo eletrônico, com enfoque de questões como inclusão, acesso à justiça e nova cultura.

Dois artigos discutiram redes sociais em temas como a violação de direitos e bloqueio de conteúdos ilícitos.

Dois artigos foram sobre o mercado de trabalho, tratando do pleno emprego e do analfabetismo digital.

Dois artigos versaram sobre a democracia eletrônica, envolvendo temas como o voto eletrônico e a democracia direta.

Um artigo sobre inovação e regulação tecnocientífica.

Um artigo sobre o direito de autor e plágio em software.

Um artigo sobre a tutela da honra no âmbito da internet.

Um artigo sobre rádio/tv na sociedade da informação.

Nota-se nessa classificação que o tema tecnológico mais tratado é a internet, mas se discute também redes sociais, nanotecnologias, urnas eletrônicas, software e tv/rádio. Dos temas jurídicos a privacidade e a responsabilidade civil são numericamente majoritários. Processo eletrônico, democracia digital e mercado de trabalho estão em seguida. Com únicos artigos seguem temas diversos, mas em pouco número considerando o total de artigos. Observa-se, portanto, algumas temáticas se tornando focais nessa edição e mantendo o interesse que vem das edições anteriores dessa coletânea.

Enfim, os artigos que ora são apresentados ao público têm a finalidade de fomentar a pesquisa e fortalecer o diálogo interdisciplinar em torno do tema direito e novas tecnologias. Trazem consigo, ainda, a expectativa de contribuir para os avanços do estudo desse tema no âmbito da pós-graduação em Direito brasileira, apresentando respostas para uma realidade que se mostra em constante transformação.

Os Coordenadores

Prof. Dr. Aires José Rover

Prof. Dr. José Renato Gaziero Cella

Profa. Dra. Valéria Ribas do Nascimento

# AS NOVAS TECNOLOGIAS À FRENTE DA DETECÇÃO DE PLÁGIO EM SOFTWARE E DO DIREITO DE AUTOR

## NEW TECHNOLOGIES APPLIED TO PLAGIARISM DETECTION ON SOFTWARE AND COPYRIGHT

Cinthia O. A. Freitas  
Claudia Maria Barbosa

### Resumo

O artigo baseia-se nas leis de proteção ao software e ao direito de autor. Ele apresenta e discute o plágio em programas de computador, motivado pela questão de que o plágio deste tipo de bem intelectual não pode ser tratado tal qual o plágio em textos literários no que diz respeito à detecção e verificação. Não se pode estar restrito apenas à comparação direta de funcionalidades e telas. Entende-se que somente a análise técnico-científica dos códigos-fonte dos programas de computador é que permite a verificação de plágio. Propõe-se o uso de ferramentas de Mineração de Textos aplicadas aos códigos-fonte dos sistemas de modo a construir análises consistentes e objetivas. Conclui-se que as leis de proteção de softwares e copyright levam em conta tais técnicas e por isso estão aptas a proteger de forma adequada o software contra o plágio.

**Palavras-chave:** Novas tecnologias, Direito de autor, Propriedade intelectual, Plágio, Programa de computador

### Abstract/Resumen/Résumé

The paper is based on the laws of protection of Software and Copyright. It presents and discusses plagiarism on software, motivated by the question that plagiarism, as an intellectual common good cannot be treated such as plagiarism in literary texts, with respect to the detection and verification. Analysis of plagiarism cannot be restricted to the direct comparison of features and screens. Only the technical-scientific analysis of the source code of software allows the professionals detecting and verifying the plagiarism. It is proposed the use of Text Mining techniques applied directly to source code of the systems to build consistent and objective analyses. The laws of software and copyright consider those technics, so they are able to protect the software from plagiarism.

**Keywords/Palabras-claves/Mots-clés:** New technologies, Authorship law, Copyright law, Intellectual property, Plagiarism, Software

## 1. Introdução

O Direito reconhece e distingue o patrimônio material e imaterial, e os protege de formas distintas. O patrimônio material é, em regra, protegido pelo conceito de propriedade, enquanto o patrimônio imaterial possui um substrato misto, que o torna a um só tempo objeto de apropriação, mas também de valoração moral, cuja proteção ainda é inadequada. Ascensão corrobora no sentido de lembrar a todos sobre “... o incremento da abstração nas sociedades contemporâneas” (ASCENSÃO, 2006, p.03). E, portanto, “os limites do Direito de Autor não são tomados como exceções, mas como via da satisfação simultânea de interesses individuais e da comunidade” (ASCENSÃO, 2006, p.05).

Em ambos os casos a proteção jurídica existente, pressupõe o reconhecimento, a apropriação, a titularidade e a repartição dos direitos decorrentes desse patrimônio, o que faz com que o Direito, atualmente, seja um limitador ao reconhecimento e à adequada proteção de determinados direitos, como os decorrentes do conhecimento tradicional e aqueles que são expressões culturais de um povo ou comunidade que, por natureza, não tem titularidade individualizada e/ou definida e não devem ser passíveis de apropriação (ZANIRATO e RIBEIRO, 2007, p.39-55).

Embora limitada e por vezes inadequada, os demais bens que compõem o patrimônio imaterial são protegidos juridicamente no direito interno por meio do conceito de propriedade imaterial, cujo reconhecimento legal gera os chamados direitos imateriais, gênero de que são espécies: a propriedade intelectual e os direitos de personalidade (VIDE e DRUMMOND, 2005, p.8-17).

A propriedade intelectual, por sua vez, divide-se entre os direitos autorais e conexos, e a propriedade industrial, regulada no Brasil pelo Instituto Nacional da Propriedade Industrial (INPI). Porém, nem toda produção intelectual é merecedora de proteção, seja por Direito de Autor ou por meio de Patente, o qual gera os denominados direitos intelectuais (BARBOSA, 2010, p.1883).

A doutrina divide-se em torno da natureza jurídica do Direito de Autor, visto que para alguns ele possui um substrato patrimonial, de maneira que pode ser enquadrado como espécie de direito real, mas para outros autores sua base são os direitos da personalidade, uma vez que os direitos de propriedade intelectual têm sempre um substrato moral e, nesse sentido, não são adequadamente protegidos pelos direitos reais. No Brasil os Direitos de Autor são

regulados pela Constituição Federal, pelo Direito Civil e leis esparsas, protegendo o autor e sua obra seja esta artística ou literária.

O respeito à propriedade intelectual fez surgir mecanismos de proteção, cujo desrespeito enseja penalidade, tanto na esfera cível quanto na criminal. Uma das formas de desrespeito ao direito de propriedade imaterial é a contrafação, expressão genérica que enseja em seu interior qualquer utilização não autorizada de obra intelectual. Há vários dispositivos que tratam dos Direitos de Autor, como a própria Lei de Direitos Autorais - Lei No. 9.610/98 (BRASIL, 1998) a Lei dos Cultivares (arts. 4º ao 27º da Lei No. 9.456/97) e o Código Penal (§§ 1º, 2º e 3º, do art. 184), e, tratando especificamente, da proteção da propriedade intelectual de programas de computador, há a Lei de Software - Lei No. 9.609/98 (BRASIL, 1998). O plágio é espécie do gênero contrafação.

De uma forma geral o plágio pode envolver textos literários ou não (manuscritos, teses e dissertações), músicas, vídeos, filmes, sites, marcas registradas visuais (do inglês, *trademark*) e programas de computador (do inglês, *software*); assim, são necessárias diferentes abordagens para detectar e constatar sua existência. Este artigo aborda o uso de novas tecnologias para detecção de plágio em programas de computador (*software*), descrevendo como as técnicas e os métodos baseados em Aprendizagem de Máquina podem fornecer um conjunto robusto de métricas que indicam o grau de similaridade entre os objetos sob análise, ou seja, um objeto de referência e um objeto questionado.

É com base na literatura de Direito de Autor que se fundam os princípios da detecção e caracterização de plágio de um modo geral, visto que tal questionamento pode integrar um processo judicial, seja na esfera Cível, Criminal, Administrativa, Trabalhista e até mesmo Interprofissional.

O artigo trata de plágio em programas de computador motivado pela questão de que o plágio deste tipo de objeto de interesse não pode ser tratado tal qual o plágio em textos no que diz respeito à sua detecção e verificação (COSMA e JOY, 2006, p. 09-10). A prática mostra que os profissionais muitas vezes não estão preparados para este tipo de problema, elaborando suas avaliações com base em elementos superficiais e fundamentadas somente na análise de semelhanças entre telas e funcionalidades.

A Lei de Software (BRASIL, 1998) no art. 2º estabelece que o programa de computador tem paternidade reivindicada pelo autor, o qual pode ainda opor-se a alterações não-autorizadas. E, ainda, que o regime de proteção à propriedade intelectual de programa de computador é o conferido às obras literárias. Eis aqui a confusão. É obra literária para fins de

enquadramento de proteção e legislação, mas quando o tema é plágio não se pode analisar e demonstrar o plágio tal qual se realiza em obras literárias.

É necessário que sejam aplicados outros métodos de análise, em particular os que consideram um conjunto robusto de características. Não obstante, as verificações sobre tais objetos necessitem de análises mais específicas, no caso em questão da área de Aprendizagem de Máquina. Ou ainda, aplicando técnicas de Mineração de Textos ao código-fonte do *software*.

Não se pode estar restrito apenas à comparação direta, que além de trabalhosa e tediosa, se revela impraticável no caso de grandes volumes de dados. Além disto, as avaliações desejadas devem ser sofisticadas de modo a permitir a descoberta de mudanças simples que busquem mascarar as similaridades entre objetos, por exemplo, a troca de letras entre duas marcas distintas (*Gucci<sup>TM</sup>* e *Guess<sup>TM</sup>*)<sup>1</sup> ou a simples troca do nome de variáveis ou da ordem dos procedimentos em um programa de computador. Neste contexto, o uso de ferramentas automáticas de Mineração de Texto relacionado à área de Aprendizagem de Máquina (do inglês, *Machine Learning*) se revela como uma alternativa promissora.

Considerando-se o exposto, o artigo é resultado de método dedutivo de pesquisa e se propõe a apresentar e discutir o Direito de Autor relacionado à detecção de plágio, permitindo uma visão não primariamente focada no plágio acadêmico de textos, mas visando análises consistentes e objetivas em objetos inerentes ao mundo digital. Objetos agora criados pelas novas tecnologias de informação e comunicação (TIC) e que necessitam, portanto, de métodos e técnicas também computacionais para detectar e demonstrar o plágio, visto que os métodos tradicionais ou comparativos não se apresentam hábeis para esta tarefa não trivial.

## 2. Plágio

Apresenta-se a conceituação de plágio, que não obstante aspectos históricos e culturais, visando o entendimento básico da conduta ao plagiar e, ainda, se enfrenta uma questão importante, a saber: O plágio pode ser verificado sob critérios objetivos Ou é uma avaliação subjetiva que depende do avaliador e de sua experiência? Moraes (2007, p.96) afirma que não existem critérios objetivos para avaliar o plágio e que sua avaliação é casuística, ou seja, caso a caso.

---

<sup>1</sup> Disponível em <<http://www.huffingtonpost.com/2012/05/21/gucci-lawsuit-gucci-wins- n 1534491.html>> Acesso em 20 mar. 2015.



## 2.1. Conceituação

Entende-se que plágio, de acordo com Dicionário da Língua Portuguesa (BUENO, 2007, p. 601) é: "s.m. Cópia; imitação; apropriação (de trabalho artístico alheio)". Para o verbo plagiar encontra-se que é "Assinar ou fazer passar por seu (trabalho artístico alheio); imitar servilmente; cometer plágio".

Para Moraes (2007, p.95) a conceituação de plágio abrange "imitação fraudulenta de uma obra, protegida pela lei autoral, ocorrendo verdadeiro atentado aos direitos morais do autor: tanto à paternidade quanto à integridade de sua criação". Observa-se não somente o agir, mas a conduta que afeta tanto paternidade (quem é o autor?) quanto integridade (qualidade de íntegro, correto, sólido e certo).

Collins e Amodeo (2005, p. 528) definem plágio como "o ato de representar trabalho criativo ou acadêmico de outra pessoa como um seu próprio na totalidade ou em parte"<sup>2</sup>. Trabalho mais recente, como o de Frechtling e Boo (2011, p.01), postula que o surgimento da Internet e por consequência de capacidades possibilitadas e mediadas por meio dos computadores têm aumentado a questão do plágio. Neste sentido, a questão do plágio tem se tornado cada vez mais evidente, seja no ambiente acadêmico, empresarial e industrial. As consequências do plágio são complexas, não passando somente pela questão da não citação de fonte ou autor, mas afetando muitas vezes os aspectos sociais e econômicos da sociedade de informação e tecnológica contemporânea.

Para *US Committee on Science, Engineering, and Public Policy*<sup>3</sup> a conceituação de plágio apresentada por Eysenbach (2000, p.01) é "usar as idéias ou palavras de outra pessoa sem dar crédito apropriado."<sup>4</sup>. Esta definição foca em plágio acadêmico, explicitando o uso de palavras de outrem sem mencionar os autores de origem. O presente artigo trata de plágio em objetos que não se apresentam por meio de palavras utilizadas para elaborar textos literários ou acadêmicos, mas tem por base a construção de programas de computador por meio da utilização de linguagens de programação.

A tipologia para o plágio é diversa e depende do ponto de vista sob o qual se "olha" para o plágio, ou seja, pode-se estabelecer uma classificação a partir da origem do plágio, do

---

<sup>2</sup> Texto original: "the act of representing someone else's creative or academic work as one's own whether in full or in part".

<sup>3</sup> Disponível em <<http://sites.nationalacademies.org/pga/cosepup/index.htm>> Acesso em 20 mar. 2015.

<sup>4</sup> Texto original: "using the ideas or words of another person without giving appropriate credit".

modo, da gravidade e, ainda, do objetivo final do ato de plagiar. Não se pretende tratar todas estas classificações, porém duas se destacam. A primeira classificação tem por base o modo como o texto propriamente dito é utilizado e foi estabelecida por Jones (2001, p.7-14). A segunda tem por foco a gravidade da conduta e foi estabelecida pela *University of Birmingham* (2014, p.6-7).

A classificação estabelecida por Jones (2001, p.7-14) inclui a definição de 4 (quatro) categorias de plágio: 1) trabalho colaborativo não reconhecido e/ou não autorizado, 2) tentativa de apresentar ou passar como seu próprio trabalho, um trabalho integral ou parcial de obra pertencente a outra pessoa, grupo ou instituição, 3) uso de qualquer quantidade de texto que tenha sido indevidamente parafraseado, incluindo-se o "*mosaic plagiarism*" ou "*cut-and-paste plagiarism*", 4) uso de parte de um texto, o qual é adequadamente parafraseado, mas que é ou não citado ou que é citado inadequadamente.

Interessa ao trabalho a classificação da *University of Birmingham* (2014, p.6-7), sendo que foram estabelecidos 3 (três) níveis, a saber: severo, moderado e fraco. O nível severo é considerado quando o “estudante se utiliza de material de outrem como se fosse seu”, sem que ao estudante não se possa atribuir a autoria (University of Birmingham, 2014, p.07). Neste caso, está incluída, por exemplo, a geração fraudulenta de dados, sendo que o estudante alega ter realizado o trabalho para produzir os dados apresentados (resultados experimentais), quando na verdade os dados não foram produzidos pelo estudante. O nível moderado se caracteriza quando o “estudante não segue as orientações sobre o que é considerado como trabalho do próprio estudante” (University of Birmingham, 2014, p.07), ou seja, o estudante tem conhecimento de um conjunto de procedimentos e orientações sobre o que pode ser considerado como trabalho de sua autoria mas não segue as recomendações. No nível fraco são enquadrados os casos de plágio nos quais “os estudantes que ainda podem não estar familiarizado com os requisitos de avaliação ao nível de uma Universidade” (University of Birmingham, 2014, p.07).

Observa-se que as classificações apresentadas estão fundamentadas no plágio acadêmico e, portanto, cabe analisar o objeto de interesse deste artigo, ou seja, os programas de computador.

## 2.2. Programa de Computador

A Lei No. 9.609 de fevereiro de 1998 (BRASIL, 1998), conhecida como Lei de Software, a qual substitui a Lei No. 7646 (de 18 de dezembro de 1987), define programa de computador:

“Art. 1º. Programa de Computador é a expressão de um conjunto organizado de instruções em linguagem natural ou codificada, contida em suporte físico de qualquer natureza, de emprego necessário em máquinas automáticas de tratamento da informação, dispositivos, instrumentos ou equipamentos periféricos, baseados em técnica digital ou análoga, para fazê-los funcionar de modo e para fins determinados”.

Laudon e Laudon (1999, p. 98) apresentam que “programa é uma série de declarações ou instruções para o computador” e, ainda, classificam os programas de computador como: *software* de sistemas e *software* aplicativo. Os autores mostram que “software de sistemas consiste em programas genéricos que gerenciam recursos do computador como a unidade central de processamento (CPU), impressoras, terminais, dispositivos de comunicação e outros equipamentos periféricos” (LAUDON e LAUDON, 1999, p. 98). Este tipo de software é conhecido muitas vezes como sistema operacional. O software aplicativo “consiste em programas elaborados para fazer o computador solucionar um determinado problema” (LAUDON e LAUDON, 1999, p. 98). O *software* de sistema é a base para o funcionamento dos programas de aplicação ou aplicativos.

Von Staa (1983, p.03) explica que programas “são, usualmente, componentes de sistemas automatizados”. Retoma-se, portanto, o entendimento da palavra informática que é o resultado da união de “informação” com “automática”. Von Staa (1983, p.03) reforça que os programas “são compostos por documentação, dados, códigos e procedimentos, e são construídos com o objetivo de instruir máquinas e pessoas no sentido da realização de um conjunto bem definido de tarefas de processamento de dados”. Resumidamente, os programas permitem o processamento de dados e a “obtenção de resultados confiáveis, úteis e oportunos” (VON STAA, 1983, p.03).

Para o desenvolvimento dos programas de computador, lança-se mão das ferramentas e linguagens de programação, as quais permitem redigir ou escrever os programas de computador utilizando-se de uma padronização compreensível pelo computador (Von Staa, 1983, p.90). Cada linguagem de programação é voltada a resolução de determinados tipos de problemas e, portanto, apresentam características de implementação, pontos fortes e

limitações que permitem aos desenvolvedores de *software* utilizar uma ou outra linguagem. Alguns exemplos de linguagem de programação podem ser citados: Assembly, “C”, C++, C Sharp, Java, BASIC, FORTRAN, Pascal, COBOL, PHP, MySQL, Python, entre muitas outras. As linguagens de programação sofrem evolução ao longo do tempo e foram sendo criadas e adaptadas para suportar a evolução que também ocorreu nos sistemas operacionais, visto que na década de 80 o desenvolvimento era realizado com base no sistema operacional para discos da Microsoft, (MS-DOS, *MicroSoft Disk Operating System*) e, atualmente, os sistemas são desenvolvidos para o sistema operacional Windows, LINUX, UNIX ou Mac OS e quando o assunto é mobilidade trata-se com os sistemas operacionais Android (Google) e iOS (Apple).

A linguagem de programação permite que o conjunto de instruções e procedimentos seja criado e assim surge o que se denomina de código-fonte. É o programa na representação que o humano pode ler, analisar, editar, alterar, armazenar, copiar e até mesmo apagar. Cada linguagem de programação, por sua vez, possui um conjunto finito de palavras reservadas que são, como o próprio nome, reservadas para uso da gramática da linguagem. Como exemplo, pode-se citar a palavra “for”, utilizada somente para construção de estruturas de repetição em diversas linguagens de programação (C e Java). Sabe-se, por exemplo, que Java possui 52 palavras reservadas enquanto Pascal possui 50. A Figura 01 exemplifica trechos de programas de computador em duas linguagens de programação distintas: C++ e Pascal.

<pre> void estado_civil()/*processamento*/ {     switch (estado)     {         case 's' :             cont1++;/*conta solteiros*/             break;         case 'c':             cont2++;/*conta casados*/         default:             cont3&lt;&lt;"\n";     } } </pre>	<pre> programa SOMA; var A, B, R, I : integer; begin     for I := 1 to 5 do         begin             write('Entre um valor para A: ');             readln(A);             write('Entre um valor para B: ');             readln(B);             writeln;             R := A + B;             writeln('O resultado corresponde a: ',R);             writeln;         end;     end. </pre>
a) Linguagem C++	b) Linguagem Pascal

Figura 01: Exemplos de programas de computador nas linguagens de programação: a) C++ e b) Pascal.

Tomando-se como base estes conceitos e premissas, pode-se discutir o foco do artigo que é o plágio em programas de computador e, então, apresentar como as novas tecnologias podem auxiliar na verificação deste tipo de problema.

### **2.3. Plágio em Programa de Computador**

Como já mencionado, o plágio em *software* não pode ser tratado tal qual o plágio em trabalhos literários. Inicialmente, tem-se por premissa metodológica que o plágio em programas de computador necessita ser avaliado no código-fonte não podendo ser verificado por meio de comparações diretas entre telas e funcionalidades.

Cosma e Joy (2006, p. 25) definem que o plágio em *software* pode ocorrer de 3 (três) maneiras distintas por meio da reutilização do código-fonte, a saber: 1) o desenvolvedor reutiliza código-fonte criado por outra pessoa por meio da obtenção deste código (os meios e maneiras de obtenção não fazem parte do escopo deste trabalho, mas sabe-se por experiência profissional que a maneira mais corriqueira de obtenção ocorre quando um programador migra de uma empresa a outra, levando consigo o código-fonte e fazendo uso nesta nova empresa); 2) o desenvolvedor reutiliza o código-fonte, com ou sem a permissão do autor original e não faz a devida referência ao autor seja intencionalmente ou não e 3) o desenvolvedor pega o código-fonte e o apresenta como sendo seu próprio autor. Utilizou-se o termo desenvolvedor de *software* de maneira genérica, podendo-se substituí-lo no meio acadêmico por estudante ou aluno, como tratado pelos autores Cosma e Joy (2006, p. 25).

Também para programas de computador há que se verificar o auto-plágio (*self-plagiarism*) o qual se caracteriza quando o autor reutiliza parte de um código-fonte que ele produziu para uma empresa e depois faz uso para o desenvolvimento de outro programa de computador em outra empresa sem o correto referenciamento. No meio acadêmico, o auto-plágio se caracteriza quando o estudante reutiliza parte de um código-fonte com o qual já obteve grau (avaliação) sem fazer a devida referência.

Considerando-se o exposto, Cosma e Joy (2006, p.25) indicam que a reutilização e, portanto, o plágio em *software* pode ocorrer das seguintes formas:

- Reprodução/Cópia do código-fonte (linguagem original em que o programador codifica um programa) sem que sejam realizadas alterações e que sejam fornecidos conhecimentos adequados;
- Adaptações mínimas ou moderadas do código-fonte autorizadas por outra pessoa sem os conhecimentos adequados;

- Conversão de todo ou parte do código-fonte para uma linguagem de programação similar;
- Uso de software de geração de código para criação de programas (código-fonte) sem que se torne explícito este fato;
- Pagamento a mais alguém para desenvolver o *software*, sendo que a atribuição de desenvolvimento é da pessoa indicada/ contratada para tal;
- Recebimento da colaboração de outras pessoas, quando esta colaboração não é permitida.

Cabe, portanto, verificar e demonstrar a ocorrência de plágio, pois o mesmo não pode ser mera especulação ou julgamento de valor. Pode-se dizer que verificar e demonstrar o plágio não são tarefas triviais e que os profissionais necessitam avaliar no código-fonte diferentes aspectos, podendo-se citar tal qual apresentado no trabalho de Roy e Cordy (2007, p.43-64):

- Cadeias (*strings*) – pode-se buscar por jogos textuais exatos, porém pode-se incorrer em erro se determinadas palavras forem trocas ou substituídas;
- *Tokens* - buscar por *tokens*, tais como cadeias de símbolos. Neste caso, é possível se incorrer em erro se primeiramente tiver sido aplicado ao código-fonte um conversor de *tokens*. Este tipo de procedimento descarta espaços em branco, comentários e identificador de nomes, tornando o sistema mais robusto para substituições de texto simples. Sistemas automáticos de detecção de plágio trabalham neste nível de comparação, aplicando algoritmos diferentes para medir a similaridade entre as seqüências de *tokens*;
- Análise de Estruturas em Árvores (*Parse Trees*) - construir e comparar as estruturas (árvores) que descrevem o encadeamento das partes que compõem o *software* (rotinas, funções, procedimentos, classes, etc). Este tipo de análise permite que semelhanças em um nível mais alto de abstração sejam detectadas. Um exemplo é a determinação das árvores de sufixos utilizadas na formação dos nomes das variáveis;
- Gráficos de Dependência do Programa (PDG - *Program Dependency Graphs*) - um PDG capta o fluxo real de controle em um programa, e permite em um nível mais alto de abstração localizar equivalências, as quais podem vir a gerar um grau de complexidade mais elevado ou até mesmo aumentar o tempo de processamento;
- Métricas – considerar análises que permitem a obtenção de métricas sobre os segmentos do código-fonte de acordo com certos critérios, por exemplo: o número de

estruturas de repetição (*loops*) e estruturas condicionais, ou o número de variáveis utilizadas. Métricas são simples de calcular e podem permitir uma análise rápida, mas podem também conduzir a falsos positivos (erros): dois fragmentos com os mesmos resultados sobre um conjunto de métricas podem realizar tarefas completamente distintas. Assim, necessita-se não somente avaliar a quantidade, mas o contexto dos resultados obtidos.

Considerando o exposto, observa-se que a detecção e constatação de plágio em *software* não são tarefas simples, exigindo dos profissionais a realização de análises em diferentes níveis de abstração, bem como de diferentes elementos caracterizadores do *software*. Portanto, a análise do código-fonte – linguagem na qual o desenvolvedor codificou o programa de computador – é imprescindível, visto que todos os elementos descritos anteriormente só podem ser avaliados a partir do código-fonte.

### **3. Análise das Leis 9.609/98 (Lei de Software) e 9.610/98 (Lei de Direitos Autorais)**

A Lei N° 9.609/98 dispõe sobre a proteção da propriedade intelectual de programa de computador, sua comercialização no País, e dá outras providências apresentado no art. 6° que (BRASIL, 1998):

Art. 6° Não constituem ofensa aos direitos do titular de programa de computador:

(...)

III - a ocorrência de semelhança de programa a outro, preexistente, quando se der por força das características funcionais de sua aplicação, da observância de preceitos normativos e técnicos, ou de limitação de forma alternativa para a sua expressão;

IV - a integração de um programa, mantendo-se suas características essenciais, a um sistema aplicativo ou operacional, tecnicamente indispensável às necessidades do usuário, desde que para o uso exclusivo de quem a promoveu.

Além disto, no art. 3° da mesma Lei, tem-se que “Art. 3° Os programas de computador poderão, a critério do titular, ser registrados em órgão ou entidade a ser designado por ato do Poder Executivo, por iniciativa do Ministério responsável pela política de ciência e tecnologia”.

Como complemento, tem-se que a Lei N° 9.610/98, a qual regula os direitos autorais, entendendo-se sob esta denominação os direitos de autor e os que lhes são conexos, por meio

do artigo 7º considera programa de computador como obras intelectuais protegidas (BRASIL, 1998). Esta Lei substituiu a Lei Nº 5.988/73.

Percebe-se, portanto, que o registro de *software* junto ao INPI (Instituto Nacional da Propriedade Industrial) é somente declaratório e não obrigatório. Assim, considera-se que os direitos autorais de um programa de computador, são estabelecidos a partir da criação originária de um sistema computacional propriamente dito. Portanto, em casos de questionamentos ou suspeita de plágio, de acordo com o ordenamento jurídico considera-se que o *software* é obra intelectual incorpórea protegida pelo Direito Autoral (art. 7º, inciso XII, da Lei 9.610/98) que especifica:

Art. 7º São obras intelectuais protegidas as criações do espírito, expressas por qualquer meio ou fixadas em qualquer suporte, tangível ou intangível, conhecido ou que se invente no futuro, tais como:

I - os textos de obras literárias, artísticas ou científicas;

II - as conferências, alocuções, sermões e outras obras da mesma natureza;

III - as obras dramáticas e dramático-musicais;

...

XII - os programas de computador;

Está ainda protegido pela Lei de Software Lei - Nº 9.609/98 (BRASIL, 1998), a qual conceitua o programa de computador e, ainda, confirma que o *software* está sob a égide dos direitos autorais no art. 2º, a saber:

Art. 2º O regime de proteção à propriedade intelectual de programa de computador é o conferido às obras literárias pela legislação de direitos autorais e conexos vigentes no País, observado o disposto nesta Lei.

Neste contexto, o artigo 6º da Lei Nº 9.609/98 é de suma importância, quando se necessita constatar o plágio em *software*, visto que não se pode somente “olhar” as telas e as funcionalidades dos sistemas computacionais analisados. No entanto, destaca-se que a principal preocupação desta lei é a reprodução não autorizada de sistemas computacionais, por exemplo, o uso de cópias ilegais de editores de texto, sistemas operacionais, jogos, entre outros. Observa-se ainda que a palavra plágio não é utilizada uma única vez no texto da Lei.

E ainda, uma vez que o *software* é bem intelectual, cabe, portanto, levar em consideração os requisitos básicos do *software* para o Direito Autoral, de acordo com Wachowicz (2004, p.142-144): Função utilitária do *software*, Originalidade relativa,



Expressão de idealização do *software* e Fixação do programa para registro no INPI. Dentre estes requisitos, cabe destacar o que se refere à originalidade relativa, a qual de acordo com Wachowicz (2004, p.143) “deve ser verificada pelo conjunto de expressões que compõem o programa de computador, de modo a não se confundir com outro preexistente”. Neste contexto entende-se por “conjunto de expressões que compõem o programa de computador” as linhas do código-fonte representadas por meio de linguagem de programação.

Para Ascensão, o programa de computador é expressão vinculada e não livre como a obra literária, visto que “nenhuma variação é possível, porque se variar há erro – já não exprime fielmente aquela realidade” (ASCENSÃO, 2006, p.10). E, assim, “a proteção do programa é a superestrutura encontrada para o que é afinal a proteção dum processo através do Direito de Autor” (ASCENSÃO, 2006, p.11). Entende-se que o código-fonte exprime a fórmula (matemática, estatística, probabilística, entre outras) que representa a realidade substanciada em causa e, deste modo, faz referência ao processo ou modelo que se encontra implementado por meio de linguagem de programação e é executável em computador ou outro aparato eletrônico (ASCENSÃO, 2006, p.10-11).

Fecha-se o ciclo de análise e discussão sobre o plágio e sua verificação, visto que somente por meio de análises detalhadas pode-se constatar se o programa de computador incorre ou não em plágio. Sabe-se que sistemas computacionais possuem milhares de linhas nos códigos-fonte e que a comparação direta ou manual entre programas é inviável e inexecutável. Entram em cena as novas tecnologias por meio das técnicas de Mineração de Textos.

#### **4. Novas Tecnologias para Detecção de Plágio**

O uso de novas tecnologias para detecção de plágio em programas de computador (*software*) permite a aplicação de métodos e técnicas baseados em Aprendizagem de Máquina de forma a extrair um conjunto robusto de métricas que indicam o grau de similaridade entre os objetos sob análise. Cabe ao estudo a aplicação de ferramentas automáticas de Mineração de Texto relacionada à área de Aprendizagem de Máquina (do inglês, *Machine Learning*).

Entende-se por Aprendizagem de Máquina o desenvolvimento de técnicas computacionais sobre o aprendizado, bem como, a construção de sistemas capazes de adquirir conhecimento de forma automática (BISHOP, 2006, p. 02). De uma maneira mais específica, pode-se questionar: as máquinas podem aprender? Sabe-se que a aprendizagem é uma habilidade essencialmente humana. Mas os computadores podem ser programados para

aprender um determinado comportamento ou padrão de modo automático a partir de exemplos ou observações.

A Aprendizagem de Máquina não é somente aplicada em problemas relacionados ao plágio. Muitas aplicações já se encontram desenvolvidas, podendo-se citar, por exemplo, o reconhecimento de face, impressão digital, voz e retina e, ainda, sistemas para previsão do comportamento de consumidores com base em dados de comércio eletrônico.

Deve-se lembrar que a Aprendizagem de Máquina propõe ferramentas poderosas para aquisição automática de conhecimento, entretanto, não existe um único algoritmo que apresente melhor desempenho para todos os problemas. Novamente, recorre-se à casuística.

Outra área de interesse e estudo é a de Reconhecimento de Padrões, a qual busca construir métodos automáticos para detecção de regularidades em dados por meio da utilização de algoritmos de computador e, então, tomar decisões com base nessas regularidades, tais como classificar os dados em diferentes categorias (BISHOP, 2006, p.01). No caso em questão, o interesse recai sobre duas classes originalmente: plágio e não-plágio.

Entende-se por Mineração de Textos (*Text Mining*) o processo de obtenção de informação de qualidade a partir de textos em linguagem natural (FELDMAN e SANGER, 2007, p.01). Uma derivação desta área é a aplicação de tais procedimentos para a análise de linguagens artificiais, tais quais as linguagens de programação em que os programas de computador são desenvolvidos.

Os programas de computador processam dados que são “fatos brutos que podem ser modelados para serem convertidos em informações” (LAUDON e LAUDON, 1999, p.377). Os dados por sua vez podem ser: estruturados, semiestruturados e não estruturados (GHEZZI e JAZAYERI, 1985, p.77-137). Entende-se por dados estruturados aqueles que são armazenados e mantidos em Sistemas Gerenciadores de Banco de Dados (SGBD), estando organizados em blocos semânticos, ou seja, apresentam relacionamento entre dados ou tabelas de dados. Os dados semiestruturados são os dados que não são armazenados em SGBDs, visto que apresentam organização heterogênea. E, os dados não estruturados são os que não possuem uma estrutura definida, por exemplo, imagens e vídeos.

A Mineração de Textos é inspirada na área de Mineração de Dados, que consiste em se extrair informações úteis, implícitas e previamente desconhecidas de bancos de dados estruturados. Por outro lado, a Mineração de Textos extrai informações de dados textuais não estruturados ou semiestruturados (BERRY e CASTELLANOS, 2007, p.04) (FELDMAN e SANGER, 2007, p.01). Aplicável, portanto, aos programas de computador.

As técnicas de Aprendizagem de Máquina resultam em análises estatísticas ou probabilísticas e, portanto, associam um grau de certeza (100%) ou de incerteza (100% - *score*) ao resultado obtido. Ressalta-se que toda análise estatística ou probabilística está sujeita a dois tipos de erros: Erro Tipo I e Erro Tipo II (KAZMIER, 1982, p.156). O Erro Tipo I ocorre quando se rejeita a hipótese de pesquisa quando de fato ela é verdadeira. E o Erro Tipo II é o revés, ou seja, se aceita a hipótese de pesquisa quando ela é falsa.

Necessita-se aplicar métodos não-paramétricos para os quais as leis de probabilidade, em cada classe de objeto a ser analisada, são desconhecidas. Os programas de computador são, portanto, dados textuais não estruturados que podem ser minerados por métodos não-paramétricos, os quais estão sujeitos a erros. Qualquer sistema de análise humana ou automática é sujeita a erros. Cabe aos profissionais analisarem os limiares de aceitação ou de rejeição da hipótese a ser comprovada: É plágio? Qual o grau de certeza? Qual a taxa de erro?

Estes questionamentos permeiam toda e qualquer análise que envolva plágio, seja em textos literários ou programas de computador no âmbito acadêmico ou profissional. Para tal, apresenta-se a seguir uma proposta de um método de verificação de plágio em programas de computador, tomando como base métricas robustas.

#### **4.1 Proposta de um Método para Detecção de Plágio em Programas de Computador**

A análise da existência de plágio entre programas ou sistemas computacionais pode ser realizada por meio de diferentes técnicas, a saber:

- De funcionalidade/telas: Como já mencionado anteriormente, Wachowicz explica que para que um *software* seja considerado bem intelectual perante o Direito Autoral deve atender diversos requisitos, dentre eles a originalidade relativa (2004, p. 143). Neste sentido, a simples análise das funcionalidades ou telas dos programas e sistemas computacionais não pode ser indicativo de plágio, também de acordo com a Lei N° 9.609/98, art. 6°;
- De estrutura: a análise da estrutura ou do encadeamento das linhas de código-fonte não permite apontar os indicativos de plágio se realizada isoladamente; além de exigir dos profissionais o conhecimento das estruturas que formam o código-fonte. Este tipo de análise pode falhar, pois se fundamenta no princípio que todas as estruturas estarão organizadas seguindo uma mesma sequência. Isto pode não ser verificado, pois trocas deliberadas na ordem das rotinas e/ou sua estrutura levará o profissional a concluir

pela não existência de plágio, mesmo quando este foi mascarado pela troca na ordem das rotinas e/ou da estrutura do sistema;

- De tecnologia: análises com base nas tecnologias empregadas tanto para desenvolvimento do *software* (tecnologias funcionais puras *versus* tecnologias orientadas a objeto) quanto para implantação de formulários, telas, entradas de dados, armazenamento de dados (banco de dados), etc., exigem profundos conhecimentos técnicos dos sistemas a analisar. Estas comparações não são triviais e, portanto, devem ser realizadas por pessoal técnico experiente;
- Mineração de Textos: neste caso o objetivo é comparar códigos-fonte para identificar e apontar indícios de plágio. As técnicas de mineração de texto apresentam as seguintes vantagens:
  - são realizadas diretamente sobre os códigos-fonte dos programas;
  - permitem a obtenção de métricas objetivas, identificando distribuições de termos como as palavras reservadas da linguagem, assim como dos nomes de variáveis, expressões, entre outros;
  - permitem a identificação de substituições simples do nomes de variáveis, por meio da análise da frequência dos termos no código-fonte como um todo.

Neste artigo propõe-se um conjunto de métricas objetivas a serem empregadas na detecção do plágio, fundamentado nas técnicas de Mineração de Textos. Inicialmente, os profissionais devem utilizar a organização estrutural do *software*. Deve-se verificar, em ambos os sistemas, por inspeção se a organização dos módulos indicada corresponde efetivamente às funcionalidades. Além disto, deve-se verificar se foram empregadas as mesmas tecnologias e linguagens de programação, a fim de confirmar a possibilidade de existência do plágio.

É possível aplicar os procedimentos usuais empregados na análise de textos em linguagem natural ao problema da verificação de plágio em programas de computador (Baeza-Yates e Ribeiro Neto, 1999, p.163-170). Em geral a análise de textos é feita de acordo com uma sequência de etapas:

- Obtenção dos elementos estruturais do texto (títulos, seções, etc), se houver, e do texto “puro” (*plain text*) formado pelas palavras que compõem o texto;
- Padronização de todos os elementos para a mesmo tamanho ou “caixa” (CAPS) – alta ou baixa;

- Remoção de *stopwords*: entende-se por *stopwords* palavras que possuem alta frequência no texto, mas baixo conteúdo semântico, por exemplo, artigos, preposições, conjunções, entre outros;
- Obtenção dos radicais das palavras (*stemming*), de forma a identificar elementos de semântica similar como, por exemplo, flexões verbais de um mesmo verbo ou nomes derivados de variáveis (por exemplo, cadastro e cadastramento, operação, operador e operacional);
- Os elementos textuais gerados a partir das etapas anteriores constituem a “essência semântica” do texto, e seus elementos – denominados termos – formam a base para o modelo vetorial empregado com frequência na área de Recuperação de Informações (Baeza-Yates e Ribeiro Neto, 1999, p.169). Neste modelo cada termo corresponde a uma dimensão em um espaço  $N$ -dimensional – sendo que  $N$  é o número de termos distintos na coleção. Um documento é representado por um vetor cujas coordenadas são valores numéricos associados os termos; em geral utiliza-se a frequência de ocorrência do termo (*tf: term frequency*) como valor que representa a probabilidade de ocorrência do termo no texto (SALTON e BUCKLEY, 1988, p.516). Outras métricas mais sofisticadas, que incluem a contagem de documentos que contém o termo podem também ser empregadas (como a *tf-idf (term frequency–inverse document frequency)*, que considera o valor de *tf* e um elemento proporcional ao número de documentos que contém o termo na coleção, ou seja o *idf* (SALTON e BUCKLEY, 1988, p.516). Essas métricas refletem a importância de uma palavra para um texto dentro de um conjunto de textos de maneira que quanto maior a frequência de uma palavra em um texto, mais representativa é a palavra para o conteúdo e quanto mais textos contiverem uma mesma palavra, menos discriminante é o texto. A representatividade e discriminância são medidas de similaridade e, portanto, indicativos ou não de plágio, dependendo dos resultados obtidos;
- No caso da existência de documentos de referência e de documentos recuperados (ou, a serem inspecionados) é possível a definição de métricas que avaliam a qualidade do processo de recuperação. As técnicas mais empregadas são a precisão (*precision*), o recobrimento (*recall*), além de sua média harmônica denominada Medida-F (*F-measure*) (MONARD e BARANAUSKAS, 2003, p.106) (ROY e CORDY, 2007, p. 65). Estas métricas permitem estabelecer um conjunto único de valores que podem ser empregados na comparação entre um elemento textual de referência ( $S_R$ ) e um

elemento questionado ( $S_Q$ ). As métricas são calculadas pelas fórmulas indicadas às Equações 3 a 5, sendo que os resultados variam entre 0 e 1, representando que quanto mais próximos de 1 mais similares são os sistemas comparados. No caso de elementos textos idênticos, tem-se:  $P=R=Medida-F = 1$  e, portanto, similaridade máxima. Para valor igual a zero tem-se que não há similaridade ou mesmo palavras em comum. No caso de se comparar dois conjuntos com o mesmo número de elementos, isto é, quando  $\|S_R\| = \|S_Q\|$ , tem-se que  $P = R = medida-F$ .

$$\text{Precisão } (P) = \|S_R \cap S_Q\| / \|S_Q\| \quad (3)$$

$$\text{Recobrimento } (R) = \|S_R \cap S_Q\| / \|S_R\| \quad (4)$$

$$\text{Medida-F} = 2 P.R / (P+R) \quad (5)$$

Conclui-se que procedimentos que permitam o uso das métricas  $P$ ,  $R$  e  $Medida-F$  para verificar a existência de plágio possibilitam aos profissionais uma avaliação mais objetiva, sem que haja necessidade de embasar suas análises, laudos e pareceres em comparações de funcionalidades e telas.

## 5. Conclusão

Este artigo apresentou e discutiu o plágio em programas de computador frente ao Direito de Autor e, ainda, apresentou uma proposta para detecção de plágio por meio da aplicação de técnicas de Mineração de Textos ao código-fonte. Considera-se como pano de fundo a Lei de Software que estabelece a não usabilidade para fins de ofensa aos direitos do autor do programa de computador: características funcionais da aplicação do programa de computador, observância de preceitos normativos e técnicos, ou de limitação de forma alternativa para a sua expressão.

A proposta permite a identificação dos dois tipos mais comuns de modificações que podem ser feitas em um código-fonte para “mascarar” o plágio, quais sejam: a alteração na ordem (permutação) de trechos do código-fonte e a substituição de nomes de variáveis e de procedimentos por meio de comandos “encontrar e substituir”, usuais nos editores de texto e ferramentas de desenvolvimento de *software*.

A proposta apresentada tem a vantagem adicional de prover métricas objetivas para avaliar a similaridade entre programas de computador (precisão, recobrimento e medida-F), emprestadas para recuperação de informação, mas que aqui tem o papel de indicar a

similaridade existente entre os códigos-fonte sob análise. Deste modo, os profissionais poderão elaborar análises e laudos consistentes e não meramente embasados em julgamentos de valor ou simples análises e comparações entre telas e funcionalidades. Com base em análises fundamentadas em métricas robustas, pode-se garantir a análise da originalidade relativa e proteger adequadamente o programa de computador.

Conclui-se que o programa de computador, caracterizado como bem intelectual e sob a proteção do Direito de Autor, está sujeito a interesses que podem levar ao plágio de parte ou da totalidade do código-fonte, independente das funcionalidades e telas serem ou não semelhantes. O artigo mostra que sob a ótica dos aspectos técnicos, o plágio pode ser detectado pela aplicação de técnicas de Mineração de Textos aos códigos-fonte (de referência -  $S_R$  e questionado -  $S_Q$ ). Avalia-se também que a legislação existente sobre a proteção de software é suficiente e apta a atingir sua finalidade, na medida em que permite que características como as que foram apresentadas no artigo sejam consideradas, com o fito de avaliar e comprovar a existência de plágio em *software*, cuja proteção é crescentemente necessária na sociedade contemporânea.

## Referências

ASCENSÃO, José de Oliveira. O Direito Intelectual em Metamorfose. Revista de Direito Autoral. Ano II, N° IV, fev., 2006, p.03-24.

BAEZA-YATES, Ricardo; RIBEIRO NETO, Berthier. Modern Information Retrieval. Addison-Wesley, 1999.

BARBOSA, Denis Borges. Tratado da Propriedade Intelectual. Tomo III. Editora Lumen Juris: Rio de Janeiro, 2010.

BERRY, Michal W.; CASTELLANOS, Malu. Survey of Text Mining I: Clustering, Classification, and Retrieval (No. 1), Springer Verlag, 2004.

BISHOP, Christopher M.. Pattern Recognition and Machine Learning. Springer, 2006.

BOO, Soyong; FRECHTLING, Douglas C.. On the ethics of management research: an exploratory investigation. Journal of Business Ethics, 106, 2012, 149-160.

BRASIL, Lei No. 9.609/1998. Lei de Software. 1998. Disponível em <[http://www.planalto.gov.br/ccivil\\_03/leis/19609.htm](http://www.planalto.gov.br/ccivil_03/leis/19609.htm)> Acesso em 20 mar. 2015.

BRASIL, Lei No. 9.610/1998. Lei de Direitos Autorais. 1998. Disponível em <[http://www.planalto.gov.br/ccivil\\_03/leis/19610.htm](http://www.planalto.gov.br/ccivil_03/leis/19610.htm)> Acesso em 20 mar. 2015.

COLLINS, Mary Elizabeth; AMODEO, Maryann. Responding to plagiarism in schools of social work: Considerations and recommendations. *Journal of Social Work Education*, 43, 2005, p.527-543.

COSMA, Georgina; JOY, Mike. Source-code Plagiarism: a UK Academic Perspective. Proc. of the 7th Annual Conference of the HEA Network for Information and Computer Sciences, 2006.

EYSENBACH, Gunther. Report of a case of cyberplagiarism - and reflections on detecting and preventing academic misconduct using the Internet. *Journal of Medical Internet Research -JMIR*. Nº 1, Vol. 2, 2000. Disponível em <<http://www.jmir.org/2000/1/e4>>. Acesso em 20 mar. 2015.

FELDMAN, Ronen; SANGER, James. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, 2007.

GHEZZI, Carlo; JAZAYERI, Mehdi. *Conceitos de Linguagens de Programação*. Rio de Janeiro:Campus, 1985.

KAZMIER, Leonard J. *Estatística Aplicada à Economia e Administração*. Trad. Carlos Augusto Crusius. São Paulo:Pearson Makron Books, 1982.

LAUDON, Kenneth C.; LAUDON, Jane Price. *Sistemas de Informação*. Rio de Janeiro:LTC, 1999.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Conceitos sobre Aprendizagem de Máquina. In: REZENDE, S. O. (Ed).*Sistemas Inteligentes: Fundamentos e Aplicações*. São Carlos:Manole, Cap. 4, Vol. 1, 2003, p.89-114.

MORAES, Rodrigo. O Plágio na Pesquisa Acadêmica: a ploriferação da desonestidade intelectual. *Revista Diálogos Possíveis*, Nº2, Vol.6, 2007, p. 91-109. Disponível em <<http://www.faculdadesocial.edu.br/dialogospossiveis/artigos/4/06.pdf>> Acesso em 20 mar. 2015.

ROY, Chanchal Kumar; CORDY, James R. *A Survey on Software Clone Detection Research*. School of Computing, Queen's University, Canadá, 2007.

SALTON, Gerard; BUCKLEY, Christopher. *Term-weighting approaches in automatic text retrieval*. Information Processing and Management – Cornell University, Ithaca, 1988.

UNIVERSITY OF BIRMINGHAM. *Code of Practice on Plagiarism*. 2014. Disponível em <<http://www.birmingham.ac.uk/Documents/university/legal/plagiarism.pdf>> Acesso em 20 mar. 2015.

VIDE, Carlos Rogel; DRUMMOND, Victor. *Manual de Direito Autoral*. Editora Lumen Juris: Rio de Janeiro, 2005.

WACHOWICZ, Marcos. *Propriedade Intelectual do Software & Revolução da tecnologia da Informação*. Curitiba: Juruá, 2004.



ZANIRATO, Silvia Helena e RIBEIRO, Wagner Costa. Conhecimento tradicionais e propriedade intelectual nos organismos multilaterais. Revista Ambiente & Sociedade, Vol. X, Nº 01, jan.-jun., 2007, p. 39-55. Disponível em <<http://www.scielo.br/pdf/asoc/v10n1/v10n1a04.pdf>> Acesso em 20. mar. 2015.