

**UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA**

FLÁVIA APARECIDA OLIVEIRA SANTOS

**CRIAÇÃO DA FERRAMENTA DE
DETECÇÃO DE PLÁGIO EM
AMBIENTE VIRTUAL DE
APRENDIZAGEM**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica como parte dos requisitos para obtenção do Título de Mestre em Ciências em Engenharia Elétrica.

Área de Automação e Sistemas Elétricos Industriais.

Orientadora: Dra. Lúcia R. Horta R. Franco.

**MAIO DE 2010
Itajubá - MG**

Uma homenagem ao meu grande tesouro, tão esperado e amado, meu filho Davi.

Agradecimentos

A DEUS por providenciar esta oportunidade me protegendo e iluminando.

A minha orientadora, Dra. Lúcia R. H. R. Franco, meus agradecimentos pela orientação na realização deste trabalho.

Aos membros e professores que fazem parte do NEaD e do CRTI, aos colaboradores M.Sc. José Renato C. Milanez, Isaac J. M. S. Rodrigues e ao estagiário João Francisco M. G. Domingues, pelo apoio técnico fornecido durante a realização desse trabalho.

A minha família por compreender e partilhar minhas angústias e aflições, em especial ao meu esposo por acreditar nos meus sonhos respeitando-os com carinho e admiração, principalmente por tudo que temos compartilhado juntos.

A todos aqueles que direta ou indiretamente colaboraram para que este projeto fosse concluído, de um modo geral a todos os amigos que me acompanham e aqueles que conquistei em Itajubá por terem, cada um do seu modo, me dado o incentivo necessário.

Finalmente, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES pelo auxílio financeiro para realização deste projeto.

Resumo

Este trabalho apresenta o software Araponga desenvolvido para dar suporte à investigação de plágio de atividades dissertativas em ambientes virtuais de aprendizagem, sua aplicação em várias situações de teste e uma análise de seus resultados. O ambiente virtual de aprendizagem escolhido para integração e os testes foi o TelEduc, instalado no servidor do Núcleo de Educação a Distância da Universidade Federal de Itajubá. Este desenvolvimento teve origem no estudo dos algoritmos dos softwares já disponíveis e respectivas deficiências. Dentre os encontrados, o mais próximo do objetivo almejado foi o software Sherlock, que foi aqui também implementado para avaliação das deficiências e busca de suas soluções na implementação do Araponga. O software Araponga enfatiza dois aspectos: o pré-processamento dos textos a serem analisados e a comparação entre frases dos textos aumentando a eficiência de detecção em situações de possíveis plágios antes não percebidas pelo Software Sherlock. Além de permitir uma análise intra-corporal, o Araponga permite uma análise extra-corporal no Google, de dentro do ambiente virtual de aprendizagem, o que auxilia mais o processo de investigação. O objetivo da pesquisa é desenvolver uma ferramenta para auxiliar o avaliador na detecção de plágio, disponibilizando um software livre que alcance a detecção de diversas ocorrências de plágio. Esta automatização favorece a confiabilidade dos resultados do processo de ensino-aprendizagem.

Palavras chave: Sherlock, TelEduc, plágio, detecção, educação, a distância.

Abstract

This paper presents the software Araponga developed to support the investigation dissertative plagiarism activities in virtual learning environments, its application in several test cases and an analysis of its results. The virtual learning environment chosen for integration and testing was TelEduc, installed on the server of the Center for Distance Education, Federal University of Itajubá. This development originated in the study of algorithms from already known softwares and their deficiencies. Among the matches, the closest to the desired objective was the Sherlock software, which was also implemented here for evaluation of weaknesses and finding solutions in the implementation of Araponga. The software Araponga focuses on two aspects: the preprocessing of the texts to be analyzed and a comparison of sentences of the texts by increasing the efficiency of detection of possible plagiarism in situations not previously perceived by the Sherlock Software. Besides allowing an intra-corporal the Araponga allows an extra-corporal analysis on Google, from within the virtual learning environment, which helps further the process of investigation. The objective proposed in this paper is to develop a tool to assist the evaluator in detecting plagiarism, providing a free software that achieves the detection of several instances of plagiarism. This automation helps the reliability of the results of the teaching-learning process.

Keywords: Sherlock, TelEduc, plagiarism, detection, education, distance

Índice

| | | |
|-------|--|----|
| 1 | Introdução | 1 |
| 1.1 | Aspectos Gerais - Evolução da Informática..... | 1 |
| 1.2 | Objetivo | 2 |
| 1.3 | Organização do Trabalho..... | 3 |
| 2 | Avaliação em Ambientes Virtuais de Aprendizagem..... | 4 |
| 3 | Detectando Plágio | 10 |
| 3.1 | Plágio | 10 |
| 3.2 | Plágio em documentos escritos..... | 15 |
| 3.3 | Detectando o plágio em AVAs | 19 |
| 3.4 | Softwares detectores de plágio | 21 |
| 3.4.1 | Software Comerciais..... | 21 |
| 3.4.2 | Software Livres..... | 22 |
| 4 | O funcionamento do Sherlock | 25 |
| 4.1 | Teste de eficácia do Sherlock* | 27 |
| 4.2 | Implementação do Sherlock*..... | 30 |
| 4.3 | Dados Experimentais e Resultados Obtidos do Sherlock*..... | 34 |
| 4.4 | Conclusão do Sherlock* | 36 |
| 5 | Proposta e Implementação (software)..... | 37 |
| 5.1 | Implementação..... | 37 |
| 5.3 | Detalhes da Implementação..... | 40 |
| 6 | Testes e Avaliação dos Resultados do Araponga | 47 |
| 6.1 | Teste com as Amostras de Umberto Eco | 47 |
| 6.2 | Teste com Plágios com diferenças em Voz Passiva e Ativa..... | 53 |
| 6.3 | Teste com Plágio com frases invertidas..... | 54 |
| 6.4 | Teste com plágio redigido sem “enter”..... | 54 |
| 6.5 | Teste com Plágios sem ponto final e com “enter” | 55 |
| 7 | Conclusão e Trabalhos Futuros..... | 58 |
| 8 | Referências..... | 62 |

Índice de Figuras

| | |
|--|----|
| Figura 1 - Resultado da avaliação do Sherlock* | 30 |
| Figura 2 - Diagrama de atividades da interface | 31 |
| Figura 3 - Interface do Sistema | 32 |
| Figura 4 - Tabela comparativa | 33 |
| Figura 5 - Resultado dos dados experimentais | 35 |
| Figura 6 - Estrutura básica do Sistema | 39 |
| Figura 7 - Diagrama de atividades do Algoritmo de Tratamento | 42 |
| Figura 8 - Texto associado na avaliação | 43 |
| Figura 9 - Tabela e as informações sobre a comparação | 44 |
| Figura 10 - Análise do Software Araponga | 45 |
| Figura 11 - Tela da Pesquisa Extra corporal | 46 |
| Figura 12 - Índice de similaridade entre trabalhos com o Software Araponga | 49 |
| Figura 13 - Índice de similaridade do Software Sherlock entre as amostras | 51 |
| Figura 14 - Gráfico Comparativo dos Softwares (amostra Eco)..... | 52 |
| Figura 15 - Gráfico Comparativo dos Softwares | 56 |

Índice de Equações

| | |
|-----------------|----|
| Equação 1 | 25 |
| Equação 2 | 25 |
| Equação 3 | 25 |
| Equação 4 | 26 |
| Equação 5 | 26 |
| Equação 6 | 43 |

Índice de Tabelas

| | |
|--|----|
| Tabela 1 – Tabela de Pré-Processamento | 41 |
| Tabela 2 – Tabela Voz Ativa e Passiva | 53 |
| Tabela 3 – Tabela do Texto sem Enter e com Enter | 55 |

Lista de Abreviaturas

API - Interface de Programação de Aplicativos

AVA - Ambientes Virtuais de Aprendizagem

BOSS - *Online Submission System*

CMC - Comunicação Mediada por Computador

EaD – Educação a Distância

IC - Instituto de Computação

MEC - Ministério da Educação

NEaD - Núcleo de Educação a Distância

NIED - Núcleo de Informática Aplicada à Educação

NTIC - Novas Tecnologias de Informação e Comunicação

PHP - *Hypertext Preprocessor*

TelEduc – Ambiente de suporte para ensino-aprendizagem

UAB - Universidade Aberta do Brasil

UNICAMP - Universidade Estadual de Campinas

UNIFEI - Universidade Federal de Itajubá

URL - Universal Resource Locator

1 Introdução

1.1 Aspectos Gerais - Evolução da Informática

Os diversos meios tecnológicos existentes propostos pelas (NTICs) - Novas Tecnologias de Informação e Comunicação, por exemplo: o uso da internet (e-mail, chats, fórum de discussão), estão, cada vez mais, sendo utilizados por todos os segmentos educacionais facilitando o processo ensino-aprendizagem. Vários educadores têm se preocupado com a introdução das diferentes tecnologias, como computador e softwares educacionais em sala de aula, procurando sistematizar o processo de aprendizagem, para que os professores possam atuar com segurança junto aos educandos.

Os recursos tecnológicos disponíveis diminuem as dificuldades existentes pela distância física entre alunos e professores, possibilitando o armazenamento, distribuição e acesso às informações, independentemente do local. A tecnologia da informática permite criar um ambiente virtual em que alunos e professores sintam-se próximos, contribuindo para o aprendizado interativo e colaborativo.

Para Moran (2000) os fatores que contribuíram para o sucesso e a difusão da modalidade educação a distância (EaD) - mediada pelo computador foram as facilidades proporcionadas pelo desenvolvimento tecnológico, que através da internet favoreceu a democratização do acesso à educação, flexibilidade e personalização da aprendizagem e incentivo da educação continuada. Por outro lado, potencializam as oportunidades de plágio ou fraudes já existentes na modalidade presencial, mas facilitadas pela

tecnologia. Isto trouxe à tona uma grande preocupação dos profissionais da educação em diagnosticar o plágio.

O Ministério da Educação (MEC), através do decreto nº 6.303 de 12 de dezembro de 2007, estabelece que “As atividades presenciais obrigatórias, compreendendo avaliação, estágios, defesa de trabalhos ou prática em laboratório, conforme o art. 1º, § 1º, serão realizados na sede da instituição ou nos pólos de apoio presencial, devidamente credenciados.”. Desta forma, as avaliações presenciais dos cursos à distância, a princípio efetivadas eletronicamente, dificultarão a incidência de plágio. De acordo com Hoffmann (2000) a melhor forma de se avaliar a aprendizagem do aluno é durante todo processo, mesmo que na maioria a avaliação seja feita virtualmente à distância e não somente em encontros presenciais.

Deste modo, o desenvolvimento de uma ferramenta para detectar plágio foi motivado pela consciência da desconfiança por parte da comunidade acadêmica no que diz respeito à execução de tarefas feitas à distância, das dificuldades encontradas por estes profissionais em identificar plágio de trabalhos realizados por alunos e do número de alunos cada vez maior na educação a distância, buscando ajudá-los neste processo.

1.2 Objetivo

Este trabalho tem como objetivo desenvolver um software de detecção de plágio eficiente para auxílio ao avaliador das atividades dissertativas postadas nos portfólios de um ambiente virtual de aprendizagem, a partir de análises dos softwares disponíveis. O ambiente escolhido para os testes foi o TelEduc, desenvolvido pelo Núcleo de Informática Aplicada à Educação (Nied) e pelo Instituto de Computação (IC) da Universidade Estadual de Campinas (UNICAMP), e utilizado pelo NEaD da Universidade Federal de Itajubá (UNIFEI).

A eficiência buscada neste trabalho se refere às questões de qualidade da detecção, tempo de identificação e custo, favorecendo a criação e o enriquecimento nas novas ferramentas educacionais e aumentando a confiabilidade dos resultados do processo de ensino-aprendizagem.

1.3 Organização do Trabalho

Este trabalho está organizado da seguinte forma: o capítulo 2 aborda como a avaliação pode ser realizada em ambientes virtuais de aprendizagem; o capítulo 3 introduz o conceito de plágio, classificação quanto ao tipo, classificação pela distribuição; o capítulo 4 mostra a funcionalidade e aplicação do algoritmo do software Sherlock; o capítulo 5 apresenta a proposta e os detalhes da implementação; o capítulo 6 exhibe os resultados obtidos através das simulações; e por fim concluindo, no capítulo 7 é discutida a avaliação dos resultados e expõe os trabalhos futuros.

2 Avaliação em Ambientes Virtuais de Aprendizagem

Este tópico descreve as modalidades da avaliação e a sua classificação.

A avaliação da aprendizagem é um tema polêmico, mesmo em se tratando de avaliação presencial. Quando o assunto é avaliação à distância, os problemas e polêmicas são ainda maiores. Sem levar em conta os aspectos de certificação e legitimidade, a avaliação torna-se muito mais complexa, por não se ter o *feedback* das interações face-a-face, que fornece alguns indícios da compreensão e interesse do aluno, possibilitando uma avaliação mais efetiva.

Segundo Perrenoud (1999), a avaliação da aprendizagem, é um processo mediador na construção do currículo e se encontra intimamente relacionada à gestão da aprendizagem dos alunos. Na avaliação da aprendizagem, o professor não deve permitir que os resultados das provas periódicas, geralmente de caráter classificatório, sejam supervalorizados em detrimento de suas observações diárias, de caráter diagnóstico. O professor, que trabalha numa dinâmica interativa, tem noção, ao longo de todo o ano, da participação e produtividade de cada aluno.

A avaliação é uma etapa do processo de ensino, cujo objetivo é garantir a aprendizagem, evidenciar posturas e escolhas metodológicas, bem como o resultado dos objetivos educacionais.

Na EaD, as preocupações com a avaliação são as mesmas da educação presencial. Embora ocorra de diferentes formas, considera as peculiaridades inerentes ao público-alvo, como a distância física entre os atores. Ao mesmo tempo, informa o

desempenho acadêmico, a atitude, o comportamento; avalia também os materiais, métodos e instrumento.

O ato de avaliar na EaD, permite preparar o cursista para o exercício de novas funções, delegando-lhe mais autonomia, o que na ótica de Hadji (2001), está na associação do desempenho pessoal do cursista com os instrumentos de avaliação e de auto-avaliação, com vistas à superação do modelo de avaliação classificatória, seletiva, autoritária e punitiva, como a praticada na pedagogia tradicional.

A avaliação em cursos a distância ou semi-presenciais com aulas virtuais pode ser realizada de três formas principais: presencial, virtual com aplicação de testes *online* e avaliação virtual ao longo do curso. Abaixo, segue uma breve descrição de cada uma dessas modalidades de avaliação:

- **Presencial:** a avaliação é feita por meio de uma prova, na presença do formador ou de outra pessoa responsável, para garantir a legitimidade da mesma.
- **Virtual com aplicação de testes *online*:** a avaliação é feita por meio de mecanismos de testes *online* a serem respondidos e enviados para o formador por meio de *e-mail* ou de formulários de envio.
- **Avaliação ao longo do curso (*contínua*):** a avaliação é feita de modo contínuo, baseada em componentes que forneçam subsídios para o formador avaliar seus aprendizes de modo processual, tais como as atividades realizadas, os comentários postados, as participações em grupos de discussão e em *chats*, as mensagens postadas no correio, etc.

Segundo Hoffmann (2000), pode-se delinear as formas de avaliação em três modalidades: diagnóstica, formativa e somativa. A diagnóstica estabelece um conhecimento prévio, definindo bases, detectando conceitos já construídos nas vivências, experiências, no cotidiano do aluno. Eles servirão de alicerce para as futuras

aprendizagens. De posse desse diagnóstico, adequar-se-ão as metodologias a serem utilizadas e serão traçados os procedimentos avaliativos. Nesse primeiro passo, é essencial o professor conhecer e se apropriar da história de cada um de seus alunos, pois esse diagnóstico fará com que as atividades sejam voltadas para a realidade dos mesmos, garantindo, assim, a aprendizagem e a qualidade do ensino. Para a autora a avaliação diagnóstica pretende averiguar a posição do aluno face a novas aprendizagens que lhe vão ser propostas e a aprendizagens anteriores que servem de base àquelas, no sentido de obviar as dificuldades futuras e, em certos casos, de resolver situações presentes.

Otsuka et al. (2003) ressaltam que as avaliações presenciais e por meio de testes online, em geral, são realizadas de forma somativa, apenas para verificar a aprendizagem dos pontos principais do conteúdo e determinar a promoção do aprendiz no final de um módulo ou curso.

Para Perrenoud (1999), a avaliação contínua pode ser considerada formativa quando tem como objetivo principal melhorar os processos de ensino-aprendizagem em andamento e não apenas verificar as aprendizagens adquiridas. O autor também auxilia esclarecendo que: *“Nem toda avaliação contínua pretende ser formativa. Muitas intervenções do professor, não têm como objetivo principal contribuir diretamente para a aprendizagem, porque sua tarefa não é somente ensinar, mas também manter a ordem, animar trocas, pôr para trabalhar, ao longo de todo o ano, em um espaço exíguo”*.

A Comunicação Mediada por Computador (CMC) abriu espaço para a exploração de novas abordagens de ensino-aprendizagem em Educação a Distância, distintas da abordagem tradicional baseada na transmissão de informações ao aprendiz.

Com os recursos introduzidos pela CMC, tornou-se possível a exploração de

abordagens baseadas na colaboração, na construção de conhecimentos e na avaliação formativa do aprendiz.

Nos cursos a distância ou nas aulas virtuais de cursos semi-presenciais, a avaliação formativa pode ser realizada por meio do acompanhamento contínuo das participações dos aprendizes nas atividades propostas no curso. Os ambientes de aprendizagem oferecem mecanismos de interação adequados para a realização de atividades em um curso a distância ou nas aulas virtuais de cursos semi-presenciais. Estes mecanismos devem ser orientados para uma mesma finalidade que é a aprendizagem. Campos (2003) cita três tipos de interações frequentes que ocorrem nos ambientes de aprendizagem:

- **Diálogo**

- O diálogo é realizado por meio de ferramentas de comunicação eletrônica (*chats*, fóruns de discussão, *news*, e-mail e outros instrumentos de interação baseados na linguagem);

- Os aprendizes participam da discussão do curso, fixando seus próprios objetivos, resolvendo problemas, procurando e descobrindo sentido para suas ações a fim de construir novas informações, e estabelecendo seus próprios critérios de "sucesso";

- Os aprendizes trocam informações com outros participantes.

- **Participação nas atividades**

- É estabelecido um ambiente que privilegia a resolução de problemas, onde os aprendizes trabalham em projetos e problemas geradores e compartilham soluções;

- O ambiente de aprendizagem é voltado para a formação dos aprendizes. É onde eles realizam exercícios com o intuito de melhorar conhecimentos e habilidades

específicas.

- **Apresentação de trabalhos**

- A interação é voltada para a apresentação de trabalhos e os alunos mostram o trabalho individual ou de grupo para os demais participantes do curso;

- A interação é voltada para a demonstração da habilidade dos aprendizes em resolver problemas e responder questões.

A partir da classificação das possíveis interações em um ambiente, pode-se pensar em uma avaliação formativa continuada baseada na interatividade. Para o sucesso desse tipo de avaliação em cursos a distância, é importante considerar os seguintes fatores:

- Na avaliação formativa, é importante que o professor conheça cada um de seus alunos. Não só o nome, mas também o jeito de ser, aprender e pensar, além dos seus gostos e expectativas;
- Após o desenvolvimento de uma ou mais atividades de aprendizagem, faz-se necessário verificar em que medida e por quantos alunos o objetivo desejado foi efetivamente alcançado;
- O professor deve observar como cada aprendiz procede em face aos problemas e como ele encontra suas próprias soluções;
- Os professores precisam constantemente intervir e auxiliar os alunos por meio de observações.
- É imprescindível a análise da participação dos aprendizes nas interações realizadas e a disponibilização dinâmica destas análises no decorrer do curso, para que os alunos tenham tempo de melhorar seu desempenho, a partir do conhecimento dos tópicos em que estão melhores e dos que precisam estudar mais.

A avaliação formativa exige um monitoramento contínuo da participação dos alunos e geração de *feedback* aos mesmos. Percebe-se então, que a quantidade de variáveis a serem consideradas para a eficácia de uma avaliação formativa gera uma sobrecarga de trabalho para o professor podendo comprometer e até inviabilizar o processo de avaliação formativa em um curso. Constitui-se, portanto, um desafio desenvolver tecnologia que dê suporte a este processo de avaliação, facilitando a tarefa do professor no desenvolvimento das atividades.

Nesta seção procurou-se dar uma visão geral de como a avaliação formativa pode ser realizada em ambientes virtuais de aprendizagem.

3 Detectando Plágio

Este tópico tem como objetivo apresentar o conceito de plágio e sua classificação.

3.1 Plágio

De acordo com Rosales et al. o plágio é o ato de assinar ou apresentar uma obra intelectual de qualquer natureza contendo partes de uma obra que pertença a outra sem que sejam dados os devidos créditos para o autor da obra consultada. Cada país trata o plágio com penalidades diferentes, mas independente do tratamento, a ética sempre deve prevalecer.

Vaz (2006) refere-se às idéias de Pierre Lévy, afirmando que as questões éticas estão envolvidas no contexto do *ciberespaço*¹. *A web* pode constituir-se em um instrumento privilegiado para construir uma civilização mais informada, mais consciente e ética. Para o filósofo a tecnologia que contribui para o benefício da comunicação, pode levar, de igual modo, ao aumento do isolamento e à alienação. O *ciberespaço* pode unir as pessoas, da mesma forma que as pode dividir, particularmente ou em grupos, separados por ideologias, políticas, posses, raças, etnias, diferenças de geração, religiões, valores éticos e morais.

O filósofo do *ciberespaço*, portanto, refere-se à ética do coletivo, cujo ponto de partida se configura em uma oportunidade para o exercício de um novo humanismo, como já referido anteriormente. Se alguém recebe outro em interação num espaço virtual, diz Lévy, é importante reconhecer que estes trazem contribuições que vão lhe acrescentar novos saberes e, em contrapartida, esperam reciprocidade.

¹ **Ciberespaço:** é um lugar onde o conhecimento é produzido e armazenado, onde pode ser articulado, onde pessoas e instituições encontram suas referências e informações. (Lévy, 2001).

Lévy diz que essa nova dimensão da comunicação humana deve “permitir-nos compartilhar nossos conhecimentos e apontá-los uns para os outros, o que é a condição elementar da inteligência coletiva²” (2000, p. 18). Para o estudioso, a transmissão, a educação, a integração e a reorganização do laço social precisam deixar de ser atividades separadas para realizar-se do todo da sociedade para si mesma, e potencialmente de qualquer ponto que seja de um social móvel a qualquer outro.

Segundo a autora, a visão otimista de Lévy, acerca de uma ética do coletivo, e a despeito das inúmeras contribuições da grande rede no estabelecimento de uma nova sociedade, muito mais informada e comunicativa, problemas como o uso desta mesma rede para a distribuição de pornografia, para a calúnia, a difamação e roubo de informações têm servido a propósitos criminais preocupantes. É nesse sentido que ações relacionadas com a questão da liberdade de expressão e privacidade das comunicações, apesar de garantidas pelas constituições das nações democráticas, configuram-se em riscos potenciais e servem como sinais de alerta para a nova realidade. É importante pontuar, porém, que essas questões não são, em essência, meros produtos da internet. A *web* como qualquer outro veículo de comunicação, pode ser utilizada de forma criminosa e/ou indevida.

Concluindo Vaz, afirma que a ética nasce da reflexão crítica do indivíduo sobre o comportamento humano e funda-se na consciência crítica, investiga os valores, interpreta, discute e problematiza. A ética indaga sobre os princípios e o comportamento moral, sempre com vistas ao bem-estar da vida em sociedade. Porém, no mundo contemporâneo, a prática da consciência crítica acerca dos problemas da virtualização dos meios de informação e comunicação ainda caminha na esteira do seu próprio desenvolvimento. Nesse contexto, o tema “ética” articula-se de forma imanente ao

² *Inteligência coletiva*: “é uma inteligência distribuída por toda a parte, incessantemente valorizada, coordenada em tempo real, que resulta em uma mobilização efetiva das competências e cuja base e objeto são o reconhecimento e o enriquecimento mútuos das pessoas” (LÉVY *apud* RAMAL, 2002, p. 2005).

plano da educação, suscitando, ainda, outras questões como o papel da educação e do educador na atual conjuntura e as questões éticas que devem ser discutidas quando ocorre a adoção da internet e de seus aplicativos por docentes e discentes. São indagações importantes e, apesar do grande movimento provocado pelo avanço da internet em todo o mundo.

Para Rios (2003), a dimensão ética refere-se à orientação da ação, constituída nos princípios do respeito e da solidariedade, buscando o encontro da realização do bem da coletividade, o que por sua vez envolve a competência, o fazer bem. Não é possível falar de competência sem relacioná-la com a moral (conjunto de valores, normas, princípios, regras que orientam a ação do homem na sociedade) e com a ética (reflexão crítica relativa à moral).

No âmbito legal, o problema do plágio e da cópia é tratado de forma geral pela lei 9.610, de 19 de fevereiro de 1998, que dispõe sobre os direitos autorais.

Pode-se dividir a detecção de plágio em dois tipos: detecção de plágio em códigos fontes e em documentos de texto. Para detecção em códigos fontes, existe a técnica de contagem de atributos e técnica de métrica de estruturas, conforme Xin Chen et al. (2004).

Encontra-se uma literatura maior referente a plágio em códigos fontes do que em documentos escritos. Em parte, isso pode acontecer devido ao fato de que detectar plágio em código fonte é mais simples do que em linguagens naturais. A gramática completa da linguagem de programação pode ser definida e especificada. No entanto, a linguagem natural³ é muito mais complexa e ambígua. Neste trabalho será abordada a detecção de plágio em documentos de texto, ou seja, linguagem natural.

³ *Linguagem Natural*: o termo é usado para distinguir as línguas faladas por seres humanos e usadas como instrumento de comunicação daquelas que são linguagens formais construídas. (Wikipédia, 2008)

Conforme Garschagen (2009), os motivos que se atribuem ao crescimento do plágio textual são vários, entre eles estão:

- 1- A má formação educacional e intelectual de alunos, professores e demais profissionais da área;
- 2- A diluição ética do que é e do que não é lícito fazer;
- 3- A facilidade trazida pela internet, que coloca à disposição, em escala geométrica, muitos textos para quem quiser copiar;
- 4- A falta de tempo e pressão para produzir trabalhos.

Perissé afirma que um dos maiores problemas da atual vida escolar e acadêmica é a cópia de textos da internet. Alunos de todos os níveis escolares fazem uso de forma indiscriminada do copiar/colar, ou seja, copiam e colam textos sem nenhum constrangimento. Esses alunos, segundo o autor, “apresentam excelentes trabalhos sobre qualquer assunto: física, semiologia, história, literatura, filosofia [...] Não escrevem, não elaboram. Vão ao *Google*⁴, conhecidíssimo instrumento de busca, e encontram o material pronto” (PERISSÉ, 2006, p. 12).

Mas, o que fazer a respeito? Como o professor deve agir em relação a essa questão? Perissé responde que é preciso considerar que:

Em nossa Idade Mídia, a internet tornou-se meio indispensável para o aprendizado, para a comunicação, para a vida profissional, para o divertimento etc. É um ambiente riquíssimo de possibilidades, e, conforme declarou recentemente Vinton Cerf, o seu idealizador, “99% das aplicações que serão usadas na rede no futuro ainda nem foram inventadas”. Ao mesmo tempo, como qualquer realidade humana, possui “brechas” perigosas, abertas por nós próprios. Uma delas tem permitido que os estudantes pesquisem sem pesquisar. [...] Sabemos que cada espaço possui as suas regras, para o bem ou para o mal. Cada casa possui as suas leis internas [...] cada um desses âmbitos requer de nós certas atitudes, uma determinada conduta, um tipo de linguagem, e sobre essas atitudes

⁴ *Google*: constitui em um site de busca de imagens, vídeos e notícias. Mais recentemente tem sido utilizado como um escritório on line, possibilitando a empresa gerenciar rotinas operacionais simples e ainda encaminhar e-mail. (VAZ *apud* COSTA, 2006, p. 2006).

e conduta precisamos estar atento, porque podemos perder o rumo e cultivar hábitos que não condizem com a própria essência desses ambientes ou com as regras maiores da condição humana. (PERISSÉ, 2006, p. 12).

No campo virtual, diz o autor, existe uma lei invisível, mas aceita por muitos, de que todos podem apropriar-se de tudo que está “acessável/acessível” e que não há nada mais natural do que copiar o texto de um *site* sem citar a fonte. É comum também não considerar que a mesma obra copiada pode ter sido fruto de grande sacrifício de um autor honesto, mas também pode já ter sido “roubada” de outro. Assim, diz Perissé (2006), são copiados parágrafos ou páginas inteiras:

Um estudante com leitura medíocre ou nula torna-se, depois de algumas clicadas, interlocutor dos diálogos de Platão, especialista nos tratados de Kant e profundo conhecedor dos ensaios de Umberto Eco [...] Os plagiadores mais inocentes copiam e colam sem pensar duas vezes (aliás, sem pensar nenhuma vez!). Já os que aprenderam a arte de enganar e enganar-se mudam palavras, inserem outras, misturam informações de várias procedências, alteram aqui e ali, tornando difícil alguém descobrir a origem de sua “inspiração” (PERISSÉ, 2006, p. 12).

Alguns especialistas, segundo Perissé, explicam que a prática do plágio muitas vezes é ocasionada pelo medo que o estudante tem de errar, de não fazer um bom trabalho, ou mesmo, de não alcançar a qualidade exigida pelo docente. Outros estudiosos, diz o autor, acreditam que o problema encontra-se nos próprios professores, que solicitam trabalhos descritivos e não estimulam a criatividade, o que acaba por induzir à prática da cópia de textos da internet.

Exigir que os trabalhos sejam feitos à mão como forma de contornar essa questão, ou, dar zero, humilhar em público, ameaçar com expulsão não são a melhor solução:

A palmatória e outros recursos violentos jamais conseguiram fazer o elogio da virtude. O mouse ético não se instala numa pessoa mediante gritos ou ameaças. Também não é possível regredir, expulsando o computador de casa, ou criando mecanismos para filtrar ou vigiar. Tudo isso pode ser driblado: onde há opressão aumenta a esperteza. [...] A mão que segura o mouse pertence a uma pessoa, e é

esta pessoa que devemos entender melhor e educar para o melhor. O melhor, neste caso, consiste em fazer o aluno experimentar o prazer do estudo. O estudo como descoberta (PERISSÉ, 2006, p. 14).

A internet, diz o autor, deve ser entendida como espaço contínuo para a ampliação da percepção de mundo. O professor deve mostrar ao estudante a importância de aprender e buscar novas idéias a partir da própria internet, como aprender um novo idioma, fazer entrevistas com escritores que estiverem *on line* e aprender a pesquisar usando o vasto universo de revistas eletrônicas, de dissertações de mestrado, teses de doutorado, ensaios, clássicos da literatura etc. “O mouse ético é o mouse inteligente. Quem descobre o prazer de pensar por conta própria descobre igualmente a necessidade de respeitar o pensamento, e os textos, de outrem” (PERISSÉ, 2006, p. 14).

Desta forma, o mundo de oportunidades de leitura, a diversidade de textos que a internet oferece pode ser um caminho para desenvolver novas formas de pensar e repensar, de criar e desenvolver e estimular conhecimentos originais e criativos.

Liu et al. classificam dois tipos de plágio: o intra-corporal e o extra-corporal. O plágio intra-corporal é aquele no qual um sujeito copia a tarefa de outro quando ambos estão realizando uma mesma tarefa. Já o extra-corporal é aquele no qual o sujeito copia de fontes externas, como por exemplo: livro, artigo de revista, monografias ou internet.

3.2 Plágio em documentos escritos

A identificação de plágio em textos escritos cabe ao professor ou tutor. Caso estejam familiarizados com o estilo de escrever do aluno, eles podem ser capazes de identificar irregularidades no trabalho se comparada a outros trabalhos do mesmo só que mais antigos ou até mesmo identificar vocabulários e linguagens diferentes utilizadas.

Inicialmente, estas características podem identificar um plágio em potencial. Clough (2000), aponta outras características suspeitas de plágios em documentos escritos são:

- **Uso de vocabulário:** comparação de vocabulário com vocabulário conhecido. Quanto maior a diferença, ou seja, quanto mais palavras novas o documento possuir, menor a probabilidade de cópia.
- **Mudança de vocabulário:** caso o vocabulário utilizado mude constantemente dentro de um mesmo texto, isto pode indicar um caso de cópia.
- **Texto incoerente:** se o texto não é consistente, pode ser o indício de cópia.
- **Pontuação:** a pontuação varia muito de texto para texto. Se ela for similar em dois documentos, pode ser um caso de cópia.
- **Quantidade de similaridade entre textos:** quanto maior a similaridade de termos comuns como nomes e definições, maior a suspeita.
- **Erros de gramática comuns:** como erros de gramáticas comuns não são prováveis de acontecer em textos revisados, a presença destes em dois documentos separados pode indicar uma cópia.
- **Estrutura do texto:** quando dois textos possuem estruturas similares, como parágrafos ou seqüência de seções e capítulos.
- **Seqüências longas de texto conhecido:** texto conhecido (frases famosas, por exemplo) sem referências.
- **Ordem de similaridade entre textos:** se dois textos de documentos diferentes possuem seqüências similares com somente algumas características diferentes (palavras, pontuação).
- **Dependência de certas frases e palavras:** um autor prefere utilizar certas palavras em particular.

- **Preferência no uso de sentenças longas ou curtas:** cada autor prefere utilizar sentenças longas ou curtas.
- **Capacidade de leitura (Readability) do texto escrito:** utilizando métricas como o índice Gunning FOG, uma pontuação é dada ao documento. É difícil que dois autores diferentes tenham a mesma pontuação.
- **Referências pendentes:** referências que aparecem no texto, mas não na bibliografia.

Para tudo isso é necessário determinar precisamente o estilo do autor. Uma boa forma de se fazer isso pode ser vista em (MCENERY; OAKES, 1999).

Para determinar o estilo de escrita de um autor, seria necessário utilizar algumas das características citadas acima. Técnicas de estatísticas também foram definidas por (ASBJORN; MACDONELL, 1996), que envolvem contar a frequência que algumas características ocorrem, elaborando assim um perfil de escrita. Estas incluem:

- Tamanho médio de sentenças (palavras).
- Tamanho médio de parágrafos (sentenças).
- O uso de voz passiva (expressado em porcentagem).
- O número de preposições (porcentagem em relação ao número total de palavras).
- A frequência de palavras de função⁵ utilizadas.

Existe também uma série de palavras e frases que são comuns em vários documentos diferentes, como nomes, datas, localizações, termos específicos e termos comuns.

Toreki (1998), descreve várias formas de plágio em linguagens naturais:

⁵ *Palavras de baixo significado léxico:* artigos, preposições, pronomes, etc.

- Copiar diretamente da fonte: copiar sem colocar aspas indicando que é uma citação.
- Falsa paráfrase: reescrever uma sentença com as próprias palavras, mas sem utilizar aspas indicando que é uma citação.
- Submeter trabalho alheio: copiar o trabalho de outra pessoa, com passagens idênticas.
- Não referenciar a fonte: novas informações que são apresentadas às pessoas que não são familiares com o assunto devem ser referenciadas apropriadamente.
- Cópia da Internet: a cópia de um conjunto de parágrafos de uma variedade de meios eletrônicos e colocadas todas juntas para fazer um documento.

Para Gonçalves (2009), um recurso muito utilizado nos textos acadêmicos são as chamadas citações indiretas ou paráfrases. Parafrasear é reescrever, com as palavras, o pensamento de um autor, com o intuito de deixar o texto mais objetivo. Grandes trechos de uma obra podem ser citados como paráfrase, evitando-se longa e desnecessária cópia.

A paráfrase exige cuidados, como manter-se fiel à informação e à idéia do texto original parafraseado, além de se fazer remissão à fonte, sempre.

Cuidado com este ponto, pois uma paráfrase não é plágio. No entanto, um texto parafraseado, sem a devida fonte, torna-se plágio. Este pode ser voluntário (proposital) ou involuntário, fruto de uma citação indireta mal feita.

Exemplos de paráfrases serão abordados no capítulo 6, onde a amostra do autor Umberto Eco (1998), foi utilizada para os testes deste trabalho. O próprio exemplo oferecido pelo autor que pretende mostrar o caminho para evitar o plágio parece evidenciar que os limites entre paráfrase e cópia são muito pouco claros. Esclarece que, uma “paráfrase quase textual que evita o plágio” é algo que se vale exclusivamente da

utilização de aspas e de algumas poucas alterações vocabulares. A “paráfrase honesta”, modo indicado pelas instituições como exemplo a ser seguido, parece bem próxima da “textual”, contando apenas com a inclusão de notas de referência. O problema do plágio para uma instituição de ensino se resolveria pela mera inclusão de aspas e referências; em outro espaço caberia perguntar se de fato o objetivo de uma pesquisa escolar (levar o aluno a pesquisar diferentes fontes, refletir sobre o tema e redigir um texto, resultado dessa reflexão) estaria melhor contemplado por um trabalho que tivesse simplesmente tido o cuidado de incluir as referências de acordo com a norma adotada pela instituição de ensino. Apesar da importância deste questionamento deve ficar claro que foco deste trabalho esta na criação de mecanismos de detecção de plágio.

Há ainda mais uma série de exemplos de plágio, como utilizar sinônimos para escrever uma mesma frase, redução de um parágrafo em uma frase, alteração da ação verbal do texto de voz passiva para voz ativa e vice e versa, inversão da ordem da frase no texto, ausência do caracter de controle “enter” e do ponto final.

Este último exemplo é comum ocorrer nos editores de textos eletrônicos, uma vez que o digitador voluntariamente ou involuntário não os utiliza, influenciando na detecção de plágio automatizada.

3.3 Detectando o plágio em AVAs

Verificar o plágio nas tarefas submetidas nos Ambientes Virtuais de Aprendizagem (AVAs) esbarra na dificuldade de comparar as tarefas de todos os alunos. Manualmente, seria necessário comparar a tarefa de um aluno com a de todos os outros, o que para uma turma de 50 alunos torna-se extremamente demorada. Para Franco et al. (2009), este problema faz com que os professores optem por tarefas nas quais os AVAs avaliam automaticamente os alunos como, por exemplo: alternativas de

múltiplas escolhas, associação de colunas e verdadeiro ou falso, deixando de lado as questões dissertativas e a produção de textos.

A fim de aumentar as opções do professor, o plágio em questões dissertativas e produções textuais podem ser detectados através de softwares. De acordo com Maurer (2006), os métodos de descoberta de plágio através de softwares geralmente são divididos em três categorias: comparação entre documentos, busca por parágrafo suspeito na internet e a estilometria.

A comparação entre documentos é a mais comum. Nessa categoria, os documentos envolvidos são comparados entre si. Essa comparação pode ser feita de várias maneiras, de acordo com a implementação de cada software. Nos softwares mais simples, a comparação é feita de palavra em palavra. Já nos softwares mais complexos, a comparação é feita por parágrafos.

A busca por parágrafo suspeito na internet é geralmente implementada com o uso de ferramentas de busca como, por exemplo: o Google e o Yahoo⁶. O sucesso deste método só ocorrerá com textos publicados na internet e que estejam disponíveis sem custo algum para os usuários. Por exemplo, as buscas em artigos de revistas na maioria das vezes exigem que o usuário seja assinante da revista.

A estilometria analisa o estilo da escrita do texto através de comparações com documentos previamente escritos pelo mesmo autor. Este método é o mais complicado, pois envolve técnicas sofisticadas de inteligência artificial para a confecção do software. Porém, se o plágio for parafraseado, o estilo do autor original deixa de existir.

Neste trabalho, será adotada a comparação entre documentos das questões dissertativas do AVA TelEduc, detectando o plágio nas formas intra e extra-corporal.

⁶ **Yahoo:** relativo ao google. Site de busca de imagens, vídeos e notícias na internet.

3.4 Softwares detectores de plágio

Os softwares para detectar plágio evoluíram sistematicamente nos últimos anos, e podem ser encontrados em diversos formatos, dos simples comparadores de palavra por palavra até os mais sofisticados analisadores que fazem a varreduras mais rápidas em documentos armazenados em base local ou em *sites da web*, afirmam (SANTANA; JOBERTO, 2006). Para os autores, é importante refletir sobre a necessidade de adoção de um sistema detector de plágio pelas instituições de ensino, especialmente as de nível superior, considerando que essa é uma ferramenta importante, capaz de auxiliar docentes e discentes na certificação da autenticidade de determinado documento.

Conforme o relatório desenvolvido por Scaife (2007), são apresentados alguns softwares detectores de plágio. Os softwares foram divididos em duas categorias em função da distribuição: comercial e livre.

3.4.1 Software Comerciais

- **Copycatch:** utilizado para comparar documentos localmente disponíveis em banco de dados. Também oferece a versão on-line que estende as capacidades de detecção do plágio na Internet, usando a Interface de Programação de Aplicativos (API) do Google.
- **Docoloc:** serviço on-line que oferece pesquisa, classificação e capacidade do Google API. O usuário do serviço envia o documento que precisa ser avaliado para um servidor que o analisa e envia um e-mail ao usuário com os fragmentos encontrados na internet.
- **Ephorus:** com mecanismo semelhante ao Docoloc.
- **Eve2 - Essay Verification Engine:** com mecanismo semelhante ao Docoloc.

- **GPSP - Glatt Plagiarism Screening Program:** armazena informações sobre o estilo da escrita de cada aluno. O autor de uma submissão suspeita tem que passar por um teste onde deve preencher uma palavra a cada cinco em todo o texto. O número de preenchimentos corretos e o tempo necessário para a conclusão do teste fornecem a hipótese de plágio.
- **MyDropBox:** conta com os mesmos recursos do Docoloc, porém utiliza arquivos de parceiros institucionais, todos protegidos por senha. O serviço usa busca proprietária e algoritmos estruturados que gera em média, relatórios em dois minutos. O software também é integrável com AVAs.
- **Turnitin:** concorrente do MyDropBox, contendo as mesmas funcionalidades.

3.4.2 Software Livres

- **Copyscape:** com a mesma idéia central do Docoloc, porém totalmente gratuita.
- **DOC Cop:** realiza testes on-line utilizando identificação do cliente. O acesso é gratuito.
- **Plagiarism Checker:** ferramenta que simplesmente utiliza os buscadores *Google* ou *Yahoo*, para procurar frases desejadas. Esta ferramenta é extremamente simples e muito limitada.
- **Praise - Plotted Ring of Analysis of Similarity Exploration:** detecta a semelhança entre documentos. O resultado da análise pode ser visualizado pela ferramenta Vast para uma análise mais minuciosa do resultado.
- **Vast – Visualisation and Analysis of Similarity Tool:** software que proporciona de forma interativa a visualização entre dois documentos. É utilizado em conjunto com a ferramenta Praise para a detecção e investigação de

similaridade.

- **Urkund:** baseado em detecção on-line, oferece um serviço automatizado para detecção de plágio. Utiliza e-mail padrão para o sistema de submissão de documentos e visualização dos resultados.
- **WCopyfind:** detecta palavras ou frases de tamanho definido dentro de um repositório local de documentos.
- **Sherlock:** Encontra semelhanças entre documentos textuais, através de assinaturas digitais. Os textos devem estar armazenados em arquivos de texto puro e as assinaturas podem ou não ser armazenadas no disco rígido, a fim de acelerar comparações futuras. Também faz parte da ferramenta BOSS, que é um sistema de submissão on-line de tarefas de estudantes de computação.

Foram realizados alguns testes dos softwares detectores de plágio apresentados pelo relatório de Scaife, não foi possível realizar testes em todos, pois os softwares comerciais exigem licenciamentos. Mas, a análise dos softwares livres detectores de plágio foi possível e conclui-se que estes com exceção do software Sherlock, não são integrado em Ambientes Virtuais de Aprendizagem.

A demora do retorno da resposta das análises dos testes dos softwares livres detectores de plágio, é preocupante, relatórios dos resultados encontrados são enviados por e-mail, após a análise, houve software que o relatório foi enviado no dia posterior à análise. Não atende a necessidade do professor em descobrir se houve plágio. O professor necessita de *feedback* em tempo real.

Dos softwares apresentados o único integrado ao AVA é o Software Sherlock, este apresenta os resultados das análises em tempo real.

O software Sherlock tem dois modos de operação: pode descobrir plágio em tarefas de código fonte, enquanto que o outro modo pode descobrir plágio em tarefas de linguagem natural.

A Universidade de Warwick no Reino Unido desenvolveu uma ferramenta de gerenciamento de curso chamada BOSS Online Submission System, a qual permite somente, que as submissões sejam realizadas com segurança pelos alunos on-line, contendo uma seleção de ferramentas eficientes.

O software Sherlock de detecção de plágio foi integrado com o software BOSS.

O projeto do BOSS, desenvolvido pelo Departamento de Ciência da Computação, de Pesquisa e o Fundo de Desenvolvimento Pedagógico da Universidade Warwick, é capaz de detectar plágio em linguagem natural, (uma construção de bloco de palavras para comunicação de idéias), sendo mais difícil do que na linguagem artificial; ou seja, na linguagem de máquina, como visto no início deste capítulo.

A proposta do projeto é atingir a detecção de diversas situações de plágio.

Neste trabalho a ferramenta foi integrada ao TelEduc, por ser este o ambiente de aprendizagem e utilizada em situações reais. Entretanto, outro fator para o desenvolvimento desta ferramenta é minimizar a dificuldade dos profissionais de educação automatizando o processo de avaliação em EaD, disponibilizando uma ferramenta livre que integra ao Ambiente TelEduc para detectar plágio e atendendo também a realidade educacional brasileira.

Entre os softwares apresentados, optou-se por implementar os princípios do software Sherlock devido à disponibilidade de sua documentação, a integração com Ambiente Virtual de Aprendizagem e o rápido retorno dos resultados encontrados. No próximo capítulo será esclarecido o seu funcionamento.

4 O funcionamento do Sherlock

Este tópico tem como objetivo apresentar a funcionalidade e aplicação do software Sherlock.

O software Sherlock criado por Pike (2007) encontra semelhanças entre textos armazenados em arquivos do tipo texto puro, conforme dito anteriormente. Para verificar a semelhança, o software analisa certa quantidade de palavras para cada linha do texto e gera uma assinatura digital que identifica essas palavras. Este procedimento de geração da assinatura digital é repetido até o final do documento.

Ao terminar esta etapa, o Sherlock possuirá as assinaturas digitais que identificam todo o texto. Para comparar um texto com outro, o mesmo procedimento é realizado com o outro texto a fim de se obter também as assinaturas digitais.

Finalmente, para determinar a semelhança entre os dois textos, o Sherlock compara as assinaturas digitais dos textos e retorna a porcentagem de semelhança entre eles. A comparação entre os textos é realizada da seguinte maneira:

$$f1 = tamanhoDoArquivo1 = A + B$$

Equação 1

$$f2 = tamanhoDoArquivo2 = A + C$$

Equação 2

Onde A é a seção similar e B ou C são dissimilares. A similaridade é dada por:

$$Similaridade = 100 \times \frac{A}{(f1 + f2 - A)}$$

Equação 3

Porém, substituindo a Equação 1 e a Equação 2 na Equação 3, tem-se:

$$\textit{Similaridade} = 100 \times \frac{A}{(A + B + A + C - A)}$$

$$\textit{Similaridade} = 100 \times \frac{A}{(A + B + C)}$$

Equação 4

Na Equação 4, caso A, B e C sejam iguais, tem-se que a similaridade será 33%. Isto é desejável uma vez que o Sherlock determina a taxa de similaridade como uma fração da soma das similaridades com as dissimilaridades.

Outra informação importante sobre o funcionamento do Sherlock está na quantidade de comparações a serem realizadas para certa quantidade de textos. Uma vez que o Sherlock compara os textos em pares e todos os textos devem ser comparados entre si, a quantidade de comparações a serem realizadas será dada por:

$$C \binom{m}{2} = \frac{m!}{2! \times (m-2)!}$$

Equação 5

Onde m é a quantidade de textos a serem comparados. Desta forma, nota-se que é indesejável comparar um texto A com um texto B se B já foi comparado com A.

O software Sherlock possui os seguintes parâmetros a serem informados pelo usuário antes que seja realizada a comparação:

- **Zero bits (z):** controla a granularidade da comparação. Quanto maior o número,

mais superficial será a comparação, porém mais rápida. Quanto menor o número, mais exata a comparação, porém mais lenta e isso pode dificultar a detecção de plágio, pois pequenas mudanças no texto serão percebidas pelo software e não serão tratadas como semelhança.

- **Chain length (n):** controla quantas palavras formam uma assinatura digital. Isto também contribui para a granularidade da comparação. Quanto maior o número, maior a exatidão. Entretanto, a comparação será mais lenta.
- **Threshold (t):** controla o quanto similar devem ser os textos antes de serem processados.

Para (LANCASTER; CULWIN, 2004) a granularidade é o número de caracteres de um texto ou número de palavras de uma sentença usadas para a análise.

O sucesso (ou fracasso) ao detectar o plágio com o Sherlock está intimamente ligado aos valores utilizados nestes parâmetros.

4.1 Teste de eficácia do Sherlock*

Para verificar a eficácia do algoritmo do Sherlock, seus princípios foram implementados num software que será chamado de agora em diante de Sherlock*, para diferenciá-lo do software originalmente desenvolvido por seus criadores. A partir deste desenvolvimento, foi criada uma amostra baseada num texto contendo um total de 108 palavras. O texto foi então dividido entre 9 alunos, onde o texto do primeiro aluno tinha as 12 primeiras palavras do texto da amostra, o segundo tinha as primeiras 24 palavras e assim sucessivamente até que o nono aluno tivesse todas as 108 palavras em seu texto. O software Sherlock* foi aplicado nas respectivas amostras:

Texto Aluno 1

Informativa - porque informa a existência de problema no processo de aprendizagem durante

Texto Aluno 2

Informativa - porque informa a existência de problema no processo de aprendizagem durante o curso, tanto para o professor como para o aluno. Reguladora – Porque

Texto Aluno 3

Informativa - porque informa a existência de problema no processo de aprendizagem durante o curso, tanto para o professor como para o aluno. Reguladora - Porque permite que os problemas informados por ela seja analisado, re-planejado e

Texto Aluno 4

Informativa - porque informa a existência de problema no processo de aprendizagem durante o curso, tanto para o professor como para o aluno. Reguladora - Porque permite que os problemas informados por ela seja analisado, re-planejado e corrigido. Exemplo prático: Se os formadores de um curso virtual recebem várias

Texto Aluno 5

Informativa - porque informa a existência de problema no processo de aprendizagem durante o curso, tanto para o professor como para o aluno. Reguladora - Porque permite que os problemas informados por ela seja analisado, re-planejado e corrigido. Exemplo prático: Se os formadores de um curso virtual recebem várias mensagens contendo as mesmas dúvidas sobre uma determinada atividade, isto indica que

Texto Aluno 6

Informativa - porque informa a existência de problema no processo de aprendizagem durante o curso, tanto para o professor como para o aluno. Reguladora - Porque permite que os problemas informados por ela seja analisado, re-planejado e corrigido. Exemplo prático: Se os formadores de um curso virtual recebem várias mensagens contendo as mesmas dúvidas sobre uma determinada atividade, isto indica que a atividade precisa ser analisada para que se identifique as causas dos

Texto Aluno 7

Informativa - porque informa a existência de problema no processo de aprendizagem durante o curso, tanto para o professor como para o aluno. Reguladora - Porque permite que os problemas informados por ela seja analisado, re-planejado e corrigido. Exemplo prático: Se os formadores de um curso virtual recebem várias mensagens contendo as mesmas dúvidas sobre uma determinada atividade, isto indica que a atividade precisa ser analisada para que se identifique as causas dos problemas, que podem ser dentre outros, alguma orientação pode não esta clara,

Texto Aluno 8

Informativa - porque informa a existência de problema no processo de aprendizagem durante o curso, tanto para o professor como para o aluno. Reguladora - Porque permite que os problemas informados por ela seja analisado, re-planejado e corrigido. Exemplo prático: Se os formadores de um curso virtual recebem várias mensagens contendo as mesmas dúvidas sobre uma determinada atividade, isto indica que a atividade precisa ser analisada para que se identifique as causas dos problemas, que podem ser dentre outros, alguma orientação pode não esta clara, o material estar mal elaborado ou o que está sendo solicitado necessita

Texto Aluno 9

Informativa - porque informa a existência de problema no processo de aprendizagem durante o curso, tanto para o professor como para o aluno. Reguladora - Porque permite que os problemas informados por ela seja analisado, re-planejado e corrigido. Exemplo prático: Se os formadores de um curso virtual recebem várias mensagens contendo as mesmas dúvidas sobre uma determinada atividade, isto indica que a atividade precisa ser analisada para que se identifique as causas dos problemas, que podem ser dentre outros, alguma orientação pode não esta clara, o material estar mal elaborado ou o que

está sendo solicitado necessita de um conhecimento anterior sobre o tema, e este deveria ser revisado antes da aplicação de tal atividade.

Desta forma, as porcentagens de plágio esperadas, entre o primeiro aluno e os outros oito alunos, baseados no funcionamento do Sherlock* deveriam ser:

- **Aluno 1 e aluno 2:** 50%
- **Aluno 1 e aluno 3:** 33,4%
- **Aluno 1 e aluno 4:** 25%
- **Aluno 1 e aluno 5:** 20%
- **Aluno 1 e aluno 6:** 16,7%
- **Aluno 1 e aluno 7:** 14,3%
- **Aluno 1 e aluno 8:** 12,5%
- **Aluno 1 e aluno 9:** 11,1%

Executou-se o Sherlock* para a amostra criada, e o resultado é apresentado na Figura 1.

Os seguintes valores foram utilizados nos parâmetros do software:

- **Zero bits (z):** 0, pois deseja-se que toda a assinatura seja considerada.
- **Chain length (n):** variado de 4 a 12, pois deseja-se que a assinatura seja criada utilizando de 4 até 12 palavras.
- **Threshold (t):** 0%, pois deseja-se que todo o resultado seja considerado.

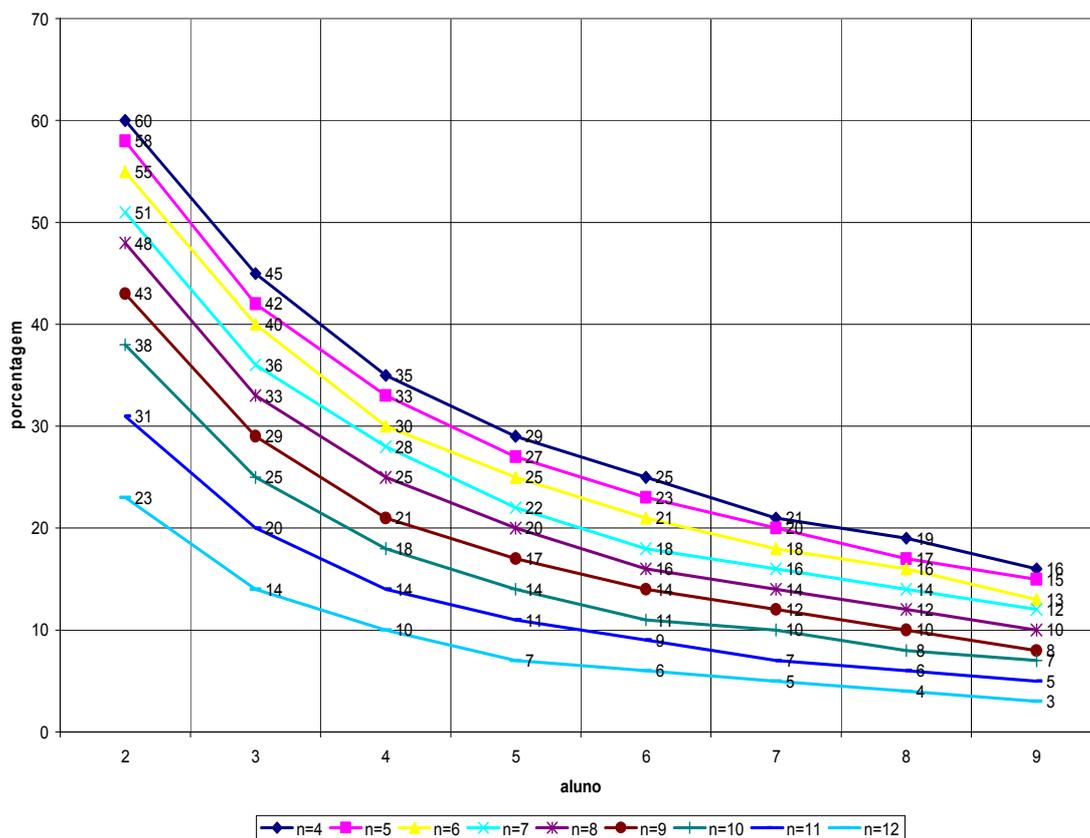


Figura 1 - Resultado da avaliação do Sherlock*

Nos resultados obtidos, nota-se que o valor encontrado não foi exatamente o esperado.

Para obter, a porcentagem de similaridade entre o aluno 1 e os demais alunos, executou-se em cada texto todos os *chain length* (n) possíveis, variando o n de 4 a 12 palavras, conforme a configuração do parâmetro.

Nota-se também que para os valores de n entre 4 e 6, os mesmos estão mais próximos do valor desejado do que os outros resultados onde n é maior do que 6.

4.2 Implementação do Sherlock*

A implementação foi direcionada para uma aplicação voltada à internet utilizando a mesma tecnologia do TelEduc (NIED, 2008), o *PHP*.

Uma interface amigável para o usuário foi desenvolvida com o propósito de permitir tanto ao coordenador, aos tutores e aos professores do curso Teste no TelEduc selecionarem o curso, a questão desejada, visualizarem e analisarem os resultados de uma forma rápida e eficiente, como mostrada a seguir.

O diagrama de atividades da interface é ilustrado na Figura 2.

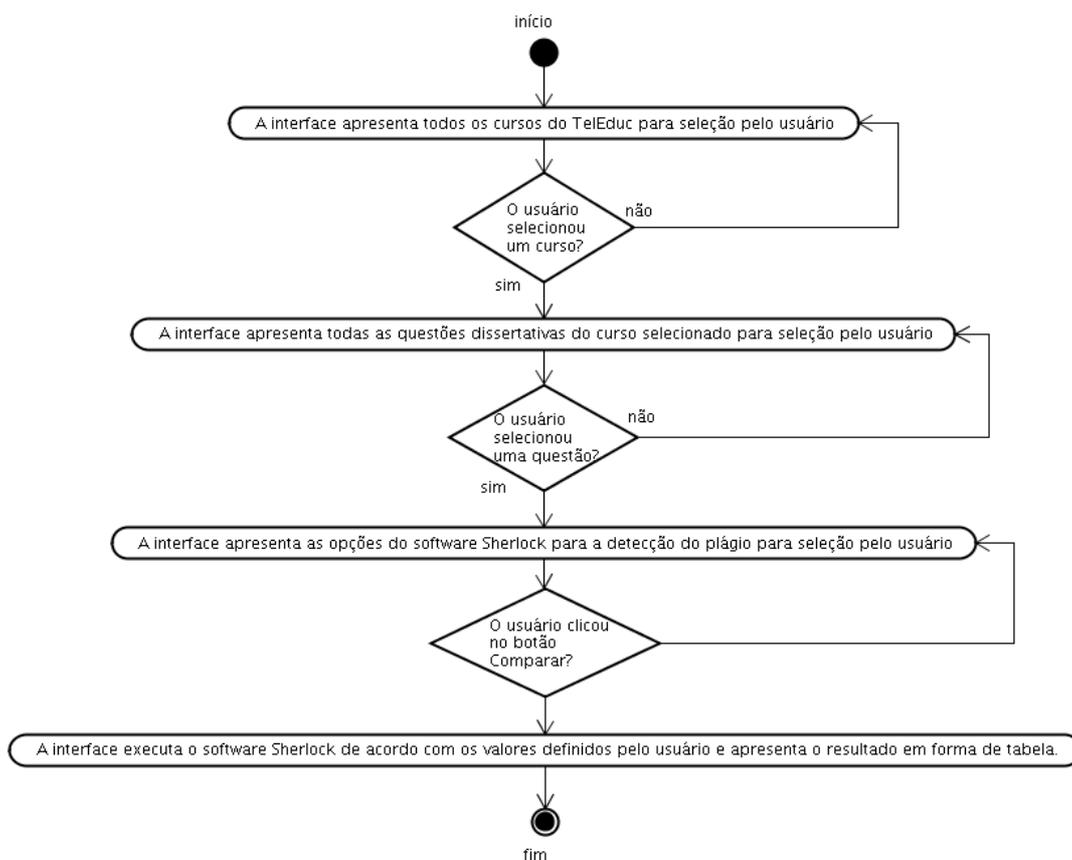


Figura 2 - Diagrama de atividades da interface

No primeiro estágio, a interface pesquisa o banco de dados do TelEduc e apresenta para o usuário todos os cursos disponíveis para seleção.

Após a seleção do curso pelo usuário, a interface pesquisa quais são as questões dissertativas que foram aplicadas no curso escolhido e as apresenta para o usuário selecionar uma entre as encontradas.

Uma vez selecionada a questão, a interface apresenta as opções do software Sherlock e a interface está pronta para iniciar a comparação.

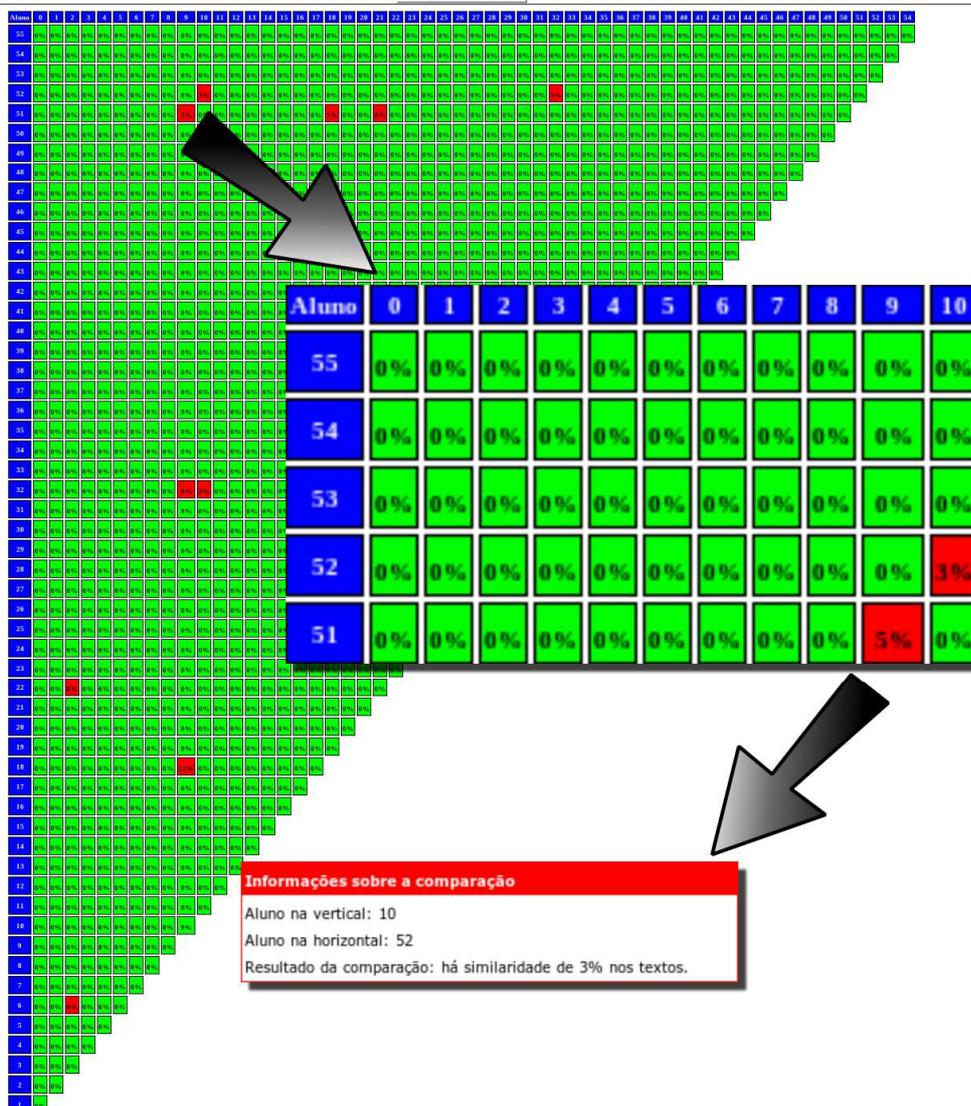
| Araponga - Software Detector de Plágio em Ambientes Virtuais de Aprendizagem | |
|---|--|
| M. Sc. José Renato Castro Milanez, Flávia Aparecida Oliveira Santos Isaac José Mendonça de Serpa Rodrigues, Prof. Dra. Lucía Regina Horta Rodrigues Franco | |
| Copyright(c) 2009 NEAD UNIFEI | |
| Seleção do curso do TelEdu | |
| Categoria | Teste |
| Status do curso | Em andamento |
| Curso | 801 - Curso de teste do suporte EAD UNIFEI |
| Informe o login e senha do curso selecionado | |
| Login | flaviaoliveira |
| Senha | •••••••• |
| Seleção do tipo de avaliação | |
| Tipo de avaliação encerrada | Portfólios individuais não avaliados |
| Lista de avaliações encerradas | Superteste |
| Lista de questões dissertativas para a avaliação selecionada | Nenhuma questão encontrada |
| Lista de alternativas da questão dissertativa selecionada | Nenhuma alternativa encontrada |
| Informe o algoritmo a ser utilizado | |
| Algoritmo | Sherlock |
| Informe as opções do Sherlock | |
| Zero bits | 3 |
| Chain length | 4 |
| Threshold | 0% |
| Comparar avaliações dos participantes | |
| Nada analisado! | |

Figura 3 - Interface do Sistema

Por fim, quando o usuário clica no botão <comparar>, a interface executa o software Sherlock* utilizando as opções desejadas por ele e apresenta o resultado em forma de tabela.

A Figura 4 apresenta a interface com a tabela comparativa.

| Informações do TelEduc | |
|-----------------------------------|--|
| Curso | Design Instrucional para EaD Virtual - Pólo Cambuí - 2007 - Módulo 4 |
| Categoria | Especialização |
| Status do curso | Encerrado |
| Enunciado da questão dissertativa | <p>Avaliação Formativa tem características informativa e reguladora. O que você entende desta afirmação? Dê um exemplo prático.</p> |
| Gabário da questão dissertativa | <p>Informativa - porque informa a existência de problema no processo de aprendizagem durante o curso, tanto para o professor como para o aluno. Reguladora - Porque permite que os problemas informados por ela seja analisado, re-planejado e corrigido. Exemplo prático: Se os formadores de um curso virtual recebem várias mensagens contendo as mesmas dúvidas sobre uma determinada atividade, isto indica que a atividade precisa ser analisada para que se identifique as causas dos problemas, que podem ser dentre outros, alguma orientação pode não esta clara, o material estar mal elaborado ou o que está sendo solicitado necessita de um conhecimento anterior sobre o tema, e este deveria ser revisado antes da aplicação de tal atividade.</p> |
| Opções do Sherlock | |
| Zero bits | 3 |
| Chain length | 4 |
| Threshold | 0% |
| Comparar novamente | |



Tempo levado para a detecção: 15 segundo(s)

Copyright(c) José Renato Castro Milanez / Flávia Oliveira

Figura 4 - Tabela comparativa

Para facilitar a identificação na tabela comparativa, foi utilizada a cor verde para indicar a ausência de indícios de plágio e a cor vermelha, a porcentagem de indícios de plágio encontrada entre as duas respostas dos alunos. Quando o usuário deixa o cursor

do mouse sobre o valor encontrado, uma mensagem informa quais são os alunos envolvidos e o índice encontrado. Este recurso é muito útil, pois o tamanho da tabela tende a ser amplo, impossibilitando que seja mostrada em uma única tela. Ao clicar no valor encontrado, uma janela se abre apresentando os dois textos para comparação e análise visual.

4.3 Dados Experimentais e Resultados Obtidos do Sherlock*

Para verificar um caso real com o software Sherlock*, foi utilizada uma questão dissertativa do curso de especialização em Design Instrucional para EaD Virtual - Pólo Cambuí - 2007 - Módulo 4, oferecido pela Educação a Distância da Universidade Federal de Itajubá. O curso continha 50 alunos, totalizando 56 respostas à questão a ser analisada.

Partindo do número total de respostas e utilizando a Equação 5, obtém-se o número de iterações:

$$C\binom{56}{2} = \frac{56!}{2! \times (56! - 2!)} = 1540$$

O software Sherlock* teve seus parâmetros configurados da seguinte maneira:

- **Zero bits (z):** 3
- **Chain length (n):** 4
- **Threshold (t):** 0%

Após a análise, que durou cerca de 15 segundos para um microcomputador baseado no processador AMD Athlon X2 +6000 com 4 Gbytes de memória RAM e um disco rígido Serial ATA II de 7200 RPM, foram obtidos os resultados apresentados na Figura 4.

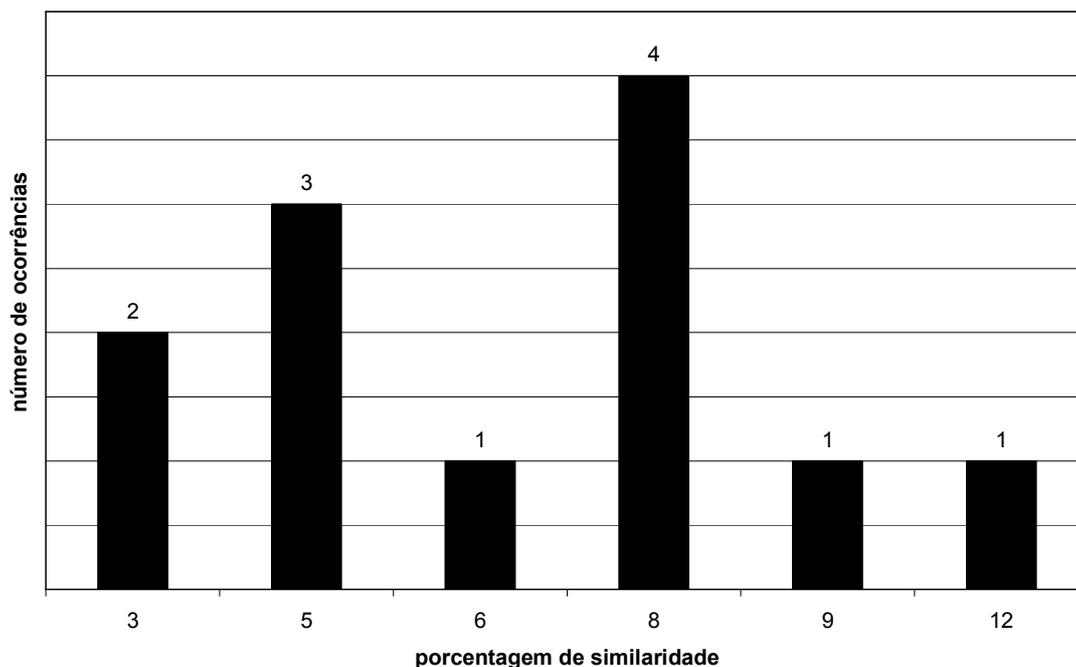


Figura 5 - Resultado dos dados experimentais

O gráfico da Figura 5 indica os resultados do Software Sherlock*, detectando 2 alunos (números de ocorrências) que apresentaram 3% de porcentagem de similaridade, 3 alunos apresentaram 5%, 1 aluno apresentou 6%, 4 alunos apresentaram 8%, 1 apresentou 9 e por fim 1 aluno com 12% de porcentagem de similaridade.

De acordo com Franco et al. (2009), conclui-se que o Software Sherlock* é ineficiente, considerando as 1540 iterações, conforme a Equação 5, numa turma de 50 alunos detectou-se pequena quantidade de ocorrências e com valores abaixo de 12% de porcentagem.

4.4 Conclusão do Sherlock*

O software Sherlock*, foi testado através de uma amostra criada neste trabalho e apresentou resultados ineficientes no que diz respeito à detecção de plágio intra-corporal. De acordo com os resultados apresentados pela Figura 1, foi possível verificar a existência de grandes variações na detecção do plágio quando o parâmetro *chain length* é alterado. Antes de se concluir que houve o plágio, deve-se avaliar atentamente os dois textos envolvidos.

De acordo com os dados experimentais e os resultados obtidos do curso de especialização em Design Instrucional para EaD Virtual - Pólo Cambuí - 2007 - Módulo 4, verificou-se que o Sherlock* encontrou uma baixíssima incidência de plágio, apresentado pela Figura 5. Os índices encontrados foram baixos para uma turma de 50 alunos.

Porém, ao analisar o funcionamento do Sherlock*, nota-se que o mesmo não possui outros tipos de verificação de plágio como, por exemplo: a substituição de palavras por sinônimos. Também não possui um tratamento especial dos textos que analise somente palavras que não possuem caracteres especiais latinos como, por exemplo: a cedilha ou o trema. Esta análise é particularmente muito importante, pois o Sherlock não será enganado pela omissão ou inclusão desses caracteres no texto, aumentando as chances de encontrar o plágio.

Há ainda outras situações não detectadas pelo Sherlock* como, por exemplo: alteração da ação verbal do texto de voz passiva para voz ativa e vice e versa, inversão da ordem da frase no texto, ausência do caracter de controle “enter” e ponto final.

5 Proposta e Implementação (software)

Este capítulo tem a finalidade de apresentar o desenvolvimento do sistema de detecção de plágio em documentos escritos em linguagem natural chamado Araponga e seus detalhes da implementação.

5.1 Implementação

Após pesquisa bibliográfica e testes realizados na implantação do Software Sherlock*, foi proposto o desenvolvimento do Software Araponga, que auxilia automaticamente na detecção de plágio efetuando melhorias em relação ao software Sherlock*.

O sistema desenvolvido possui uma arquitetura típica de sistemas Web: sistema operacional Mandriva Linux versão 2008.1, linguagem PHP versão 5.2.5 orientada a objetos e banco de dados relacional MySQL versão 5.051.a. Este último é o responsável pelo armazenamento dos dados do curso, dos usuários e o tipo de avaliação.

Para realizar a pesquisa é necessário que os arquivos sejam do tipo texto puro codificados em ASCII, pois para analisar arquivos do tipo binário como doc e pdf, é necessário um tratamento especial para recuperar o texto, uma vez que podem existir formatações tais como figuras e tabelas, e caracteres de controle que dificultam a busca por textos.

Os arquivos tipo texto são extraídos do banco de dados SQL para depois serem analisados. É realizada uma pesquisa no banco de dados do TelEduc, que será analisado pelo software Araponga, que por sua vez faz o tratamento de texto antes de iniciar a comparação de matrizes.

A Figura 6 ilustra a estrutura básica do software, onde o usuário pode pesquisar os textos e os comparar. Além disso, o usuário pode escolher também o software Sherlock* para fazer a pesquisa de possíveis casos de plágio.

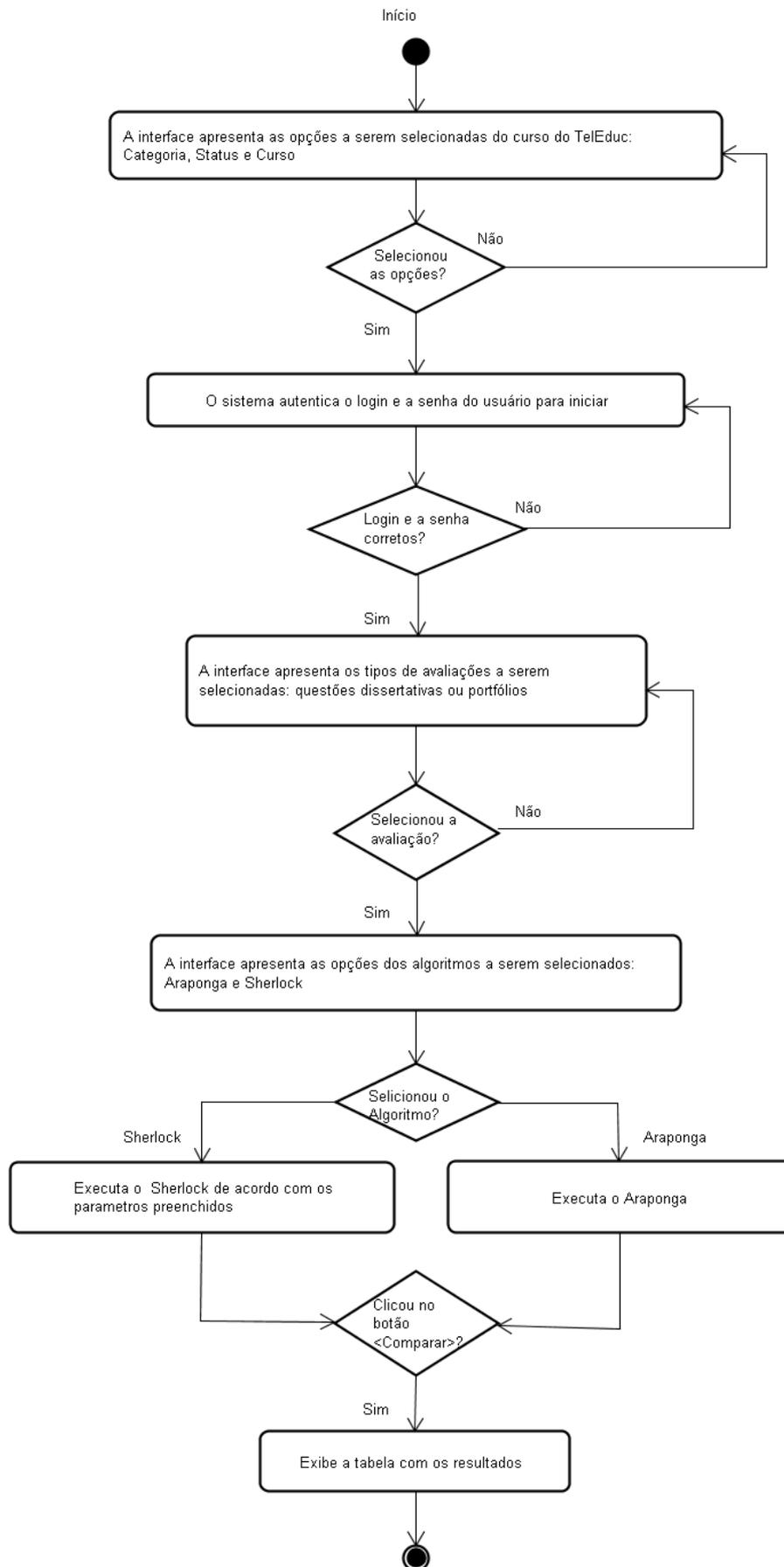


Figura 6 - Estrutura básica do Sistema

| | |
|-----------------------|---|
| Remoção de Conectivos | o, ao, na, no, que, da, de, do, das, dos, com, para, por, esta, este, pela, pelo, dela, dele, num, numa, deste, desta, disto, dessa, desse, daquela, daquele, daquilo, aquilo, em, as, os, aos, ou, se, e, um, uma; |
|-----------------------|---|

Tabela 1 – Tabela de Pré-Processamento

O próximo tratamento é a divisão de frases. Cada frase é identificada por um vetor. Cada elemento de um vetor é comparado com todos os outros elementos de todos os outros vetores. Caso haja coincidência adiciona-se uma unidade em uma variável que expressará a quantidade de palavras idênticas que aparecem nas duas frases.

Esta análise é executada em todos os textos da matriz de comparação de todos os textos dos alunos da turma.

A Figura 7 apresenta os detalhes do algoritmo de tratamento de texto para melhor compreensão do seu funcionamento.



Figura 7 - Diagrama de atividades do Algoritmo de Tratamento

Após o tratamento do texto é realizada a comparação entre matrizes: a matriz de palavras do texto original comparada com toda a matriz de palavras do texto suspeito, retornando o grau de similaridade entre as duas matrizes. Verificando se a similaridade é maior que a diferença, a fim de obter a porcentagem de plágio encontrada.

O cálculo da porcentagem em cada frase suspeita é dado por:

$$\text{Porcentagem} = 100 \times \frac{A}{B}$$

Equação 6

Onde A é a quantidade da intersecção dos dois textos, ou seja, contém o número de valores que estão presentes em comum no texto original e no texto suspeito.

E o B é quantidade de palavras do texto suspeito.

Para usar o software é necessário que os textos estejam disponíveis no Portfólio e associados à avaliação correspondente, conforme na Figura 8.

The screenshot shows a web browser window with the URL http://www.ead.unifei.edu.br/~teleduc/cursos/aplic/index.php?cod_curso=801. The page title is "Curso de teste do suporte EAD UNIFEI". The main content area is titled "Portfólio - Portfólio de Grupo / Ver Item do Portfólio" and shows a list of portfolio items. The selected item is "Teste Araponga/Sherlock" with a date of "16/09/2009 14:55:58" and a status of "Totalmente Compartilhado". The text of the item is:

O próprio Cohn, já citado, recorda ainda que "a vinda do Anticristo deu lugar a uma tensão ainda maior". As diversas gerações viviam em constante expectativa do demônio destruidor, "cujo reino seria de fato um caos sem lei, uma era consagrada à rapina e ao saque, à tortura e ao massacre, mas também prelúdio de um termo ansiado, a Segunda Vinda e o Reino dos Santos". As pessoas estavam sempre alerta e atentas aos sinais que, segundo os profetas, acompanhariam e anunciaram o último "período de desordens". Ora, sublinha Cohn, uma vez que estes sinais incluíam "maus governantes, discórdia civil, guerra, seca, fome, carestia, peste, cometas, mortes imprevistas de pessoas eminentes e uma crescente pecaminosidade geral, numa houve dificuldade em detectá-los."

Below the text is a button labeled "Avaliar Participante".

Figura 8 - Texto associado na avaliação

Quando o usuário deixa o cursor do mouse sobre o valor encontrado na tabela, uma mensagem informa quais são os alunos envolvidos e o índice encontrado, da mesma forma que é exibido no Software Sherlock*. A Figura 9 apresenta as informações.

The screenshot shows the Araponga software interface. At the top, it displays the title 'Araponga - Software Detector de Plágio em Ambientes Virtuais de Aprendizagem' and the authors: 'M. Sc. José Renato Castro Milanez, Flávia Aparecida Oliveira Santos, Isaac José Mendonça de Serpa Rodrigues, Prof. Dra. Lucia Regina Horta Rodrigues Franco'. Below this is the copyright information: 'Copyright(c) 2009 NEAD UNIFEI' and a prompt to 'Selecione o curso do TelEdu'. The interface includes several dropdown menus for 'Categoria' (Teste), 'Status do curso' (Em andamento), 'Curso' (801 - Curso de teste do suporte EAD UNIFEI), 'Tipo de avaliação encerrada' (Portfólios individuais não avaliados), 'Lista de avaliações encerradas' (Teste Araponga - Não Apagar/ Não inserir item / Não Fazer teste / Não mexer), 'Lista de questões dissertativas para a avaliação selecionada' (Nenhuma questão encontrada), 'Lista de alternativas da questão dissertativa selecionada' (Nenhuma alternativa encontrada), and 'Algoritmo' (Araponga). A button 'Comparar avaliações dos participantes' is visible. Below the button is a table with the following data:

| | Aluno 1 | Aluno 2 | Aluno 3 | Aluno 4 |
|---------|---------|---------|---------|---------|
| Aluno 5 | 100% | 32% | 86% | 97% |
| Aluno 6 | 87% | 27% | 83% | |
| Aluno 7 | 77% | 29% | | |
| Aluno 8 | 43% | | | |

An arrow points to the '100%' value in the first row. A red box highlights the '100%' value, and a detailed comparison window is open over it. The window title is 'Informações sobre a comparação'. It contains the following text:

Aluno na vertical: Aluno 1
 Aluno na horizontal: Aluno 5
 Resultado da comparação: há similaridade de 100% nos textos.

At the bottom of the window, it shows '0,10 (s)' and 'Memória utilizada: 256,00 (Kb)'.

Figura 9 - Tabela e as informações sobre a comparação

Ao clicar no valor da similaridade o Software Araponga abre uma janela exibindo o texto original e o suspeito, facilitando assim a visualização. Além disto, na mesma janela na parte inferior são apresentados os resultados: a frase original relevante, a frase suspeita relevante, o número de palavras iguais entre as duas frases e a porcentagem na frase suspeita, conforme a Figura 10.

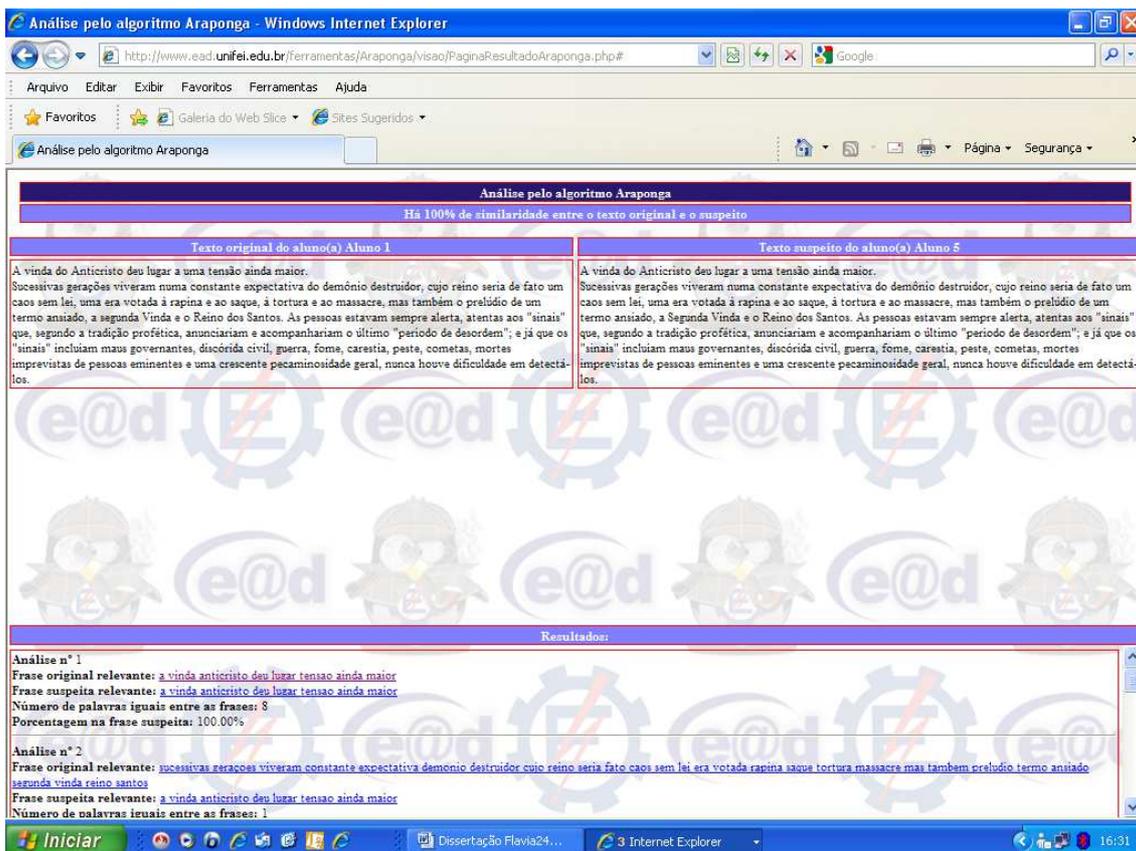


Figura 10 - Análise do Software Araoponga

Clicando em qualquer uma das frases, original ou suspeita, o Software exibe outra janela direcionando para o site do Google, exibindo a pesquisa extra-corporeal encontrada na internet.

A Figura 11 ilustra a pesquisa extra-corporeal.

Sherlock para o TelEduc x Araponga x A vinda do Anticristo deu lugar a ... x

Web Imagens Mapas Notícias Orkut Gmail mais v Efetuar login

Google A vinda do Anticristo deu lugar a uma tensão ainda maior Pesquisar Pesquisas avançadas Preferências

Pesquisar: a web páginas em português páginas do Brasil

Web Resultados 1 - 10 de aproximadamente 8.690 para A vinda do Anticristo deu lugar a uma tensão ainda maior (0,13 segundos)

CITAÇÃO BIBLIOGRÁFICA - REFERÊNCIA NUMÉRICA
 Formato do arquivo: Microsoft Word - [Ver em HTML](#)
A vinda do Anticristo deu lugar a uma tensão ainda maior. Sucessivas gerações viveram numa constante expectativa do demônio destruidor, cujo reino seria de ...
www.unimep.br/~gmarques/referenciabibliocitacoes.doc - [Páginas Semelhantes](#)

Valdínei Pereira Garcia
 Formato do arquivo: PDF/Adobe Acrobat - [Ver em HTML](#)
A vinda do Anticristo deu lugar a uma tensão ainda maior. Sucessivas gerações viveram numa constante expectativa do demônio destruidor, cujo ...
www.sapientiajus.com.br/tabelas/METODOLOGIA-DO-TRABALHO-ACADEMICO.pdf - [Páginas Semelhantes](#)

Pró Tensão: Comentários TUA VIDA É UMA LINDA CANÇÃO DE AMOR
Maior que esse dia. Parabéns mesmo, César; vc caprichou e encantou o mundo (não ... ó papo bom! **deu** saudade de conversar com vc...na sua casa nova não tem msn não? é ... A característica que define o **anticristo** é o orgulho - o qual, ... Ele **ainda** vai pensar para responder! O gênio por escrito concluiu o que pensou ...
www.apostolos.com/mint-comments.cgi?entry_id=23770 - 12k - [Em cache](#) - [Páginas Semelhantes](#)

Breve Investigação do Estudo: "Nova Era - A Vinda do Anticristo ..."
 "Nova Era - A Vinda do Anticristo" do Pr. Vanderlei Dornelles ... a História Universal sabem que este conflito se **deu** por causa das disputas coloniais ... de rivalidade, **tensão** e instabilidade conhecido como "Guerra Fria". Nova Era não é **ainda** uma nova forma de governo diferente da democracia e do comunismo" ...
www.adventistas.com/janeiro2001/nova_era_dornelles.htm - 34k - [Em cache](#) - [Páginas Semelhantes](#)

A SEGUNDA VINDA DE JESUS
 No momento de **maior** **tensão** política e espiritual da grande tribulação, ... O destino final do **anticristo** e do falso profeta será o lago de fogo (Apocalipse 19:11-21). ... a **Deus** e dos que não obedecem ao evangelho de nosso Senhor Jesus Cristo. ... Em primeiro **lugar**, o peso espiritual maligno que assola a Terra, ...

Figura 11 - Tela da Pesquisa Extra corporal

6 Testes e Avaliação dos Resultados do Araponga

Este tópico tem como objetivo avaliar os resultados do software Araponga e exibi-los através das simulações.

Para se avaliar a eficiência do Araponga, vários testes foram feitos e são abordados a seguir.

6.1 Teste com as Amostras de Umberto Eco

Neste teste, foram utilizadas as amostras descritas por Umberto Eco (1998), que se preocupou em escrever em seu trabalho diferenças entre plágio, paráfrase e falsa paráfrase com o trecho do livro *Os Fanáticos do Apocalipse* de Norman Cohn. Para o autor é necessário não confundir paráfrase honesta com falsa paráfrase (citação sem aspas), que constitui uma modalidade de plágio. Seguem os respectivos textos:

Texto 1 - Original

A vinda do Anticristo deu lugar a uma tensão ainda maior.

Sucessivas gerações viveram numa constante expectativa do demônio destruidor, cujo reino seria de fato um caos sem lei, uma era votada à rapina e ao saque, à tortura e ao massacre, mas também o prelúdio de um termo ansiado, a Segunda Vinda e o Reino dos Santos.

As pessoas estavam sempre alerta, atentas aos "sinais" que, segundo a tradição profética, anunciariam e acompanhariam o último "período de desordem"; e já que os "sinais" incluíam maus governantes, discórdia civil, guerra, fome, carestia, peste, cometas, mortes imprevistas de pessoas eminentes e uma crescente pecaminosidade geral, nunca houve dificuldade em detectá-los.

Texto 2 - Paráfrase honesta

A esse respeito, Cohn é bastante explícito. Debruça-se sobre a situação de tensão típica desse período, em que a expectativa do anticristo é, ao mesmo tempo, a do reino do demônio, inspirado na dor e na desordem, mas também prelúdio da chamada Segunda Vinda, a Parúsia, a volta do Cristo triunfante.

Numa época dominada por acontecimentos sombrios, saques, rapinas, carestia e pestes, não faltavam às pessoas os "sinais" correspondentes aos sintomas que os textos proféticos haviam sempre anunciado como típicos da vinda do Anticristo.

Texto 3 - Falsa paráfrase

Segundo Cohn... [segue-se uma lista de opiniões expressas pelo autor em outros capítulos].

Por outro lado, cumpre não esquecer que a vinda do Anticristo deu lugar a uma tensão ainda maior.

As gerações viviam na constante expectativa do demônio destruidor, cujo reino de fato um caos sem lei, uma era consagrada à rapina e ao saque, à tortura e ao massacre, mas também o prelúdio à Segunda Vinda ou ao Reino dos Santos. As pessoas estavam sempre alerta, atentas aos sinais que, segundo os profetas, acompanhariam e anunciariam o último "período de desordem": e, já que esses "sinais" incluíam os maus governantes, a discórdia civil, a guerra, a seca, a fome, a carestia, as pestes e os cometas, além das mortes imprevistas de pessoas importantes (e uma crescente pecaminosidade geral), nunca houve dificuldade em detectá-los.

Texto 4 – Paráfrase quase textual que evita o plágio

O próprio Cohn, já citado, recorda ainda que "a vinda do Anticristo deu lugar a uma tensão ainda maior".

As diversas gerações viviam em constante expectativa do demônio destruidor, "cujo reino seria de fato um caos sem lei, uma era consagrada à rapina e ao saque, à tortura e ao massacre, mas também o prelúdio de um termo ansiado, a Segunda Vinda e o Reino dos Santos".

As pessoas estavam sempre alerta e atentas aos sinais que, segundo os profetas, acompanhariam e anunciariam o último "período de desordens".

Ora, sublinha Cohn, uma vez que estes sinais incluíam "maus governantes, discórdia civil, guerra, seca, fome, carestia, peste, cometas, mortes imprevistas de pessoas eminentes e uma crescente pecaminosidade geral, nunca houve dificuldade em detectá-los".

Estes textos acima, assim como uma cópia do Texto 1, foram postados nos portfólios de alunos fictícios nomeados como Aluno 1, Aluno 2, Aluno 3, Aluno 4 e Aluno 5. Os respectivos portfólios foram associados à atividade de avaliação do Software Araponga no curso disponibilizado no Ambiente do TelEduc para realização dos testes.

Ao executar o Software Araponga nos textos da mesma amostra, foram obtidos os seguintes resultados apresentados na Figura 12.

Araponga - Software Detector de Plágio em Ambientes Virtuais de Aprendizagem

M. Sc. José Renato Castro Milanes, Flávia Aparecida Oliveira Santos
Isaac José Mendonça de Serpa Rodrigues, Prof. Dra. Lucia Regina Horta Rodrigues Franco

Copyright(c) 2009 NEAD UNIFEI

Selecione o curso do TeL@duc

| | | | | |
|--|--|---------|---------|---------|
| Categoria | Teste | | | |
| Status do curso | Em andamento | | | |
| Curso | 801 - Curso de teste do suporte EAD UNIFEI | | | |
| Informe o login e senha do curso selecionado | | | | |
| Login | flaviaoliveira | | | |
| Senha | ***** | | | |
| Selecione o tipo de avaliação | | | | |
| Tipo de avaliação encerrada | Portfólios individuais não avaliados | | | |
| Lista de avaliações encerradas | Teste Araponga - Não Apagar/ Não iserir item / Não Fazer teste / Não mexer | | | |
| Lista de questões dissertativas para a avaliação selecionada | Nenhuma questão encontrada | | | |
| Lista de alternativas da questão dissertativa selecionada | Nenhuma alternativa encontrada | | | |
| Informe o algoritmo a ser utilizado | | | | |
| Algoritmo | Araponga | | | |
| Comparar avaliações dos participantes | | | | |
| Aluno 5 | Aluno 1 | Aluno 2 | Aluno 3 | Aluno 4 |
| 100% | 32% | 86% | 97% | |
| Aluno 4 | 87% | 27% | 83% | |
| Aluno 3 | 77% | 29% | | |
| Aluno 2 | 43% | | | |

Tempo de execução: 0,10 (s) Memória utilizada: 256,00 (Kb)

Figura 12 - Índice de similaridade entre trabalhos com o Software Araponga

Os índices de similaridades exibidos variam de 0 a 100, onde 0 é dissimilaridade e 100 é similaridade total ou 100%.

De acordo com os resultados exibidos na Figura 11, o índice de similaridade entre os textos dos Aluno1 e Aluno 5 aponta para plágio total ou seja 100%. Este resultado está correto, pois esses textos foram duplicados propositalmente para o teste, sendo idênticos.

Com exceção dos índices do Aluno 2, que apresentaram valores baixos de similaridade, os demais índices indicaram valores altos de similaridade. Os índices do aluno 2 foram abaixo da média porque o texto 2 é uma paráfrase honesta (uso de aspas), o que não impediu a detecção pelo Software Araponga, devido ao pré-processamento, onde o Software Araponga exclui os caracteres inválidos (no caso as aspas) para

analisar a comparação entre os textos, que conseqüentemente detectou as similaridades encontradas entre os pares de alunos.

O texto do aluno 3 é uma falsa paráfrase e o Araponga detectou todos os seus índices apontando 77%, 83% e 86% de similaridade entre os textos.

E por último, o texto do aluno 4 onde o autor escreve uma paráfrase tentando evitar plágio, não impediu que o Software Araponga detectasse 97% de similaridade.

Já a mesma amostra testada com o Software Sherlock*, alcançou resultados diferentes do Software Araponga como, por exemplo: os pares dos textos dos alunos 1 e aluno 3 e do aluno 3 com os alunos 4 e alunos 5 indicaram abaixo de 33% de similaridade. Enquanto que no Software Araponga a média encontrada nos mesmos pares foi de 82% de porcentagem de similaridade.

O texto do aluno 2, onde se encontra a paráfrase honesta, não identificou nenhuma porcentagem de similaridade, validando assim a pesquisa de Eco, que no seu posicionamento, quando a referência é citada não é identificado como plágio. Diferentemente da proposta deste trabalho, que é detectar plágio qualquer que seja a semelhança.

O índice de similaridade no par de textos dos alunos 1 e 5 também foram 100%, uma vez que os textos foram duplicados propositalmente.

Observam-se na Figura 13, em destaque na cor vermelha, os altos índices de similaridades entre pares dos documentos e na cor verde os pares dos documentos com textos divergentes apresentando altos índices de dissimilaridade, ou baixos índices de similaridades.

Segue apresentada na Figura 13 a tabela de resultados do Software Sherlock*.

Seleção do curso de TeLDuc

Categoria: Teste

Status do curso: Em andamento

Curso: 801 - Curso de teste do suporte EAD UNIFEI

Informe o login e senha do curso selecionado

Login: flaviaoliveira

Senha:

Seleção do tipo de avaliação

Tipo de avaliação encerrada: Portfólios individuais não avaliados

Lista de avaliações encerradas: Teste Araponga - Não Apagar/ Não iserir item / Não Fazer teste / Não mexer

Lista de questões dissertativas para a avaliação selecionada: Nenhuma questão encontrada

Lista de alternativas da questão dissertativa selecionada: Nenhuma alternativa encontrada

Informe o algoritmo a ser utilizado

Algoritmo: Sherlock

Informe as opções do Sherlock

Zero bits: 3

Chain length: 4

Threshold: 0%

Comparar avaliações dos participantes

| Aluno | Aluno 1 | Aluno 2 | Aluno 3 | Aluno 4 |
|---------|---------|---------|---------|---------|
| Aluno 5 | 100% | 0% | 25% | 40% |
| Aluno 4 | 40% | 0% | 24% | |
| Aluno 3 | 25% | 0% | | |
| Aluno 2 | 0% | | | |

Tempo de execução: 0,23 (s) Memória utilizada: 256,00 (Kb)

Figura 13 - Índice de similaridade do Software Sherlock entre as amostras

Os resultados obtidos com o Software Sherlock*, tanto de similaridade quanto de dissimilaridade, são divergentes em relação aos resultados do Software Araponga.

O gráfico da Figura 14 permite visualizar melhor esses resultados entre os dois softwares, comparando apenas o texto do aluno 1 com os textos dos outros alunos.

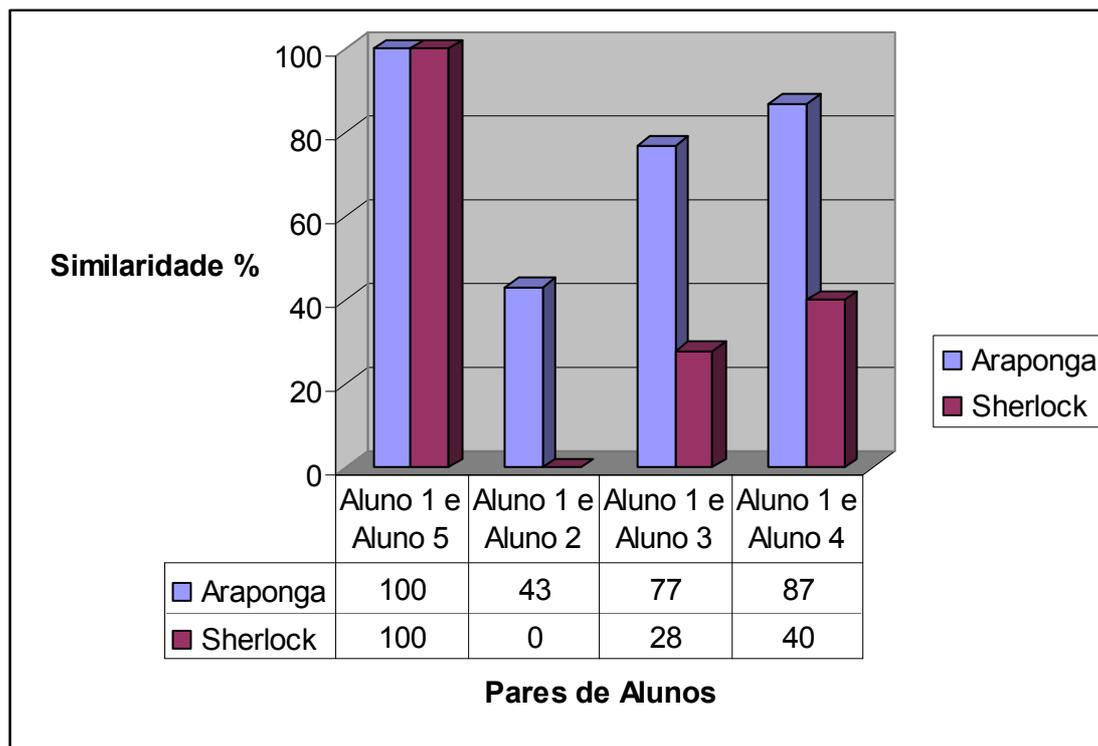


Figura 14 - Gráfico Comparativo dos Softwares (amostra Eco)

Observa-se que o Software Araponga detectou similaridade maior em três casos. Os resultados foram esperados, uma vez que o Sherlock* não analisa todas as palavras de uma mesma frase (parâmetro chain length), tão pouco faz análise de frequência das palavras em todas as frases.

Outra vantagem do Araponga sobre o Sherlock* é o tempo de execução, no algoritmo Araponga o tempo foi de 0,06 segundos e no algoritmo Sherlock* foi 0,13 segundos, sendo assim o algoritmo Araponga é muito mais rápido.

Vale ressaltar que esta amostra foi executada com apenas cinco alunos participantes do teste, numa turma de 50 alunos, por exemplo, o tempo de execução seria bem maior.

O Algoritmo do Sherlock* não faz a pesquisa extra-corporal, diferentemente do Software Araponga que com apenas um clique sobre a frase suspeita ou sobre a frase original, abre o browser exibindo as URL encontradas, conforme citado no capítulo anterior.

6.2 Teste com Plágios com diferenças em Voz Passiva e Ativa

Uma melhoria implementada no Araponga deveu-se a percepção de que os outros softwares não identificavam plágios com textos diferentes somente na escrita em voz passiva com os escritos em voz ativa.

Assim, para um dos testes, os mesmos conteúdos dos textos na voz ativa foi redigido na voz passiva, alterando apenas a ação verbal, e estes textos foram testados nos dois softwares. O Software Araponga confirmou a hipótese de mais eficiente, detectando 76% de similaridade, enquanto que o Software Sherlock* não detectou nenhuma percentagem. A Tabela 2 exhibe as frases que foram avaliadas no teste.

| Voz Ativa | Voz Passiva |
|---|--|
| Branca de Neve mordeu a maçã envenenada. | A maçã envenenada foi mordida pela Branca de Neve. |
| Peter Pan derrotou o Capitão Gancho. | O Capitão Gancho foi derrotado por Peter Pan. |
| Ali Babá pronunciou as palavras mágicas. | As palavras mágicas foram pronunciadas por Ali Babá. |
| Alice seguiu o Coelho Branco. | O Coelho Branco foi seguido por Alice. |
| O Advogado assumiu o caso. | O caso foi assumido pelo advogado. |
| Maria fez uma boa prova. | Uma boa prova foi feita por Maria. |
| O diretor da escola agradou as crianças. | As crianças foram agradadas pelo diretor da escola. |
| O caçador matou o jacaré. | O jacaré foi morto pelo caçador. |
| Os animais do zoológico atacaram os visitantes. | Os visitantes foram atacados pelos animais do zoológico. |
| O barulho acordou toda a vizinhança. | Toda a vizinhança foi acordada pelo barulho. |

Tabela 2 – Tabela Voz Ativa e Passiva

Nota-se que na primeira linha da tabela 2 há quatro palavras iguais entre as frases: branca, neve, maçã, envenenada. Aplicando a equação 6

$Porcentagem = 100 \times \frac{A}{B}$, onde A é igual a 4 e B é igual a 5, implica-se

$Porcentagem = 100 \times \frac{4}{5}$, obtendo 80% de similaridade na primeira frase suspeita, e

assim segue a análise em todas as frases até completar o texto inteiro, retornado a porcentagem total de similaridade entre os dois textos.

6.3 Teste com Plágio com frases invertidas

Para o teste das frases invertidas foram utilizados três portfólios de alunos com mesmo texto do teste da voz ativa, alterando apenas a ordem das frases. As frases foram mescladas diferentemente para cada aluno, e mais uma vez, foi identificada a melhor eficiência do Software Araponga em relação ao Software Sherlock*, apresentando a porcentagem de similaridade de 98% a 100% enquanto que o Software Sherlock* apresentou as similaridades de 66% a 83%.

6.4 Teste com plágio redigido sem “enter”

Num portfólio, é comum o digitador não terminar a frase ou parágrafo corrente com um “enter”, por esquecimento ou por não ver necessidade uma vez que não vai continuar a digitar mais nada. E como estudado, a ausência deste caracter de controle confunde a maioria dos programas até então disponíveis. Tal situação, então, foi prevista no Araponga.

O teste do texto regido sem “enter” foi realizado com dois textos com o mesmo conteúdo, porém um escrito numa linha só sem “enter” e outro em várias linhas separadas por “enter”, retornando o resultado de 100% para o Software Araponga e 66% para o Software Sherlock.

Observem a Tabela 3:

| Texto sem “Enter” | Texto com “Enter” |
|--|---|
| A casa é branca, as portas e janelas são azuis, o telhado vermelho. O carro preto e a moto prata. As crianças se chamam Clara e João, elas brincam no parquinho do quintal, que tem uma enorme piscina com cascata | A casa é branca. As portas e janelas são azuis. O telhado vermelho. O carro preto e a moto prata. As crianças se chamam Clara e João, elas brincam no parquinho do quintal, que tem uma enorme piscina com cascata. |

Tabela 3 – Tabela do Texto sem Enter e com Enter

Neste caso para calcular a porcentagem da frase suspeita o Software Araponga analisou-se o texto sem “enter” até o primeiro ponto final da frase e comparou-se com a primeira frase do texto com “enter”, detectando 100% de similaridade na frase suspeita.

6.5 Teste com Plágios sem ponto final e com “enter”

Pelo mesmo motivo que a situação anterior, o esquecimento do ponto final ocorre com frequência e confunde os demais softwares. Razão pela qual sua diferenciação foi implementada também no Araponga.

Então, por último, foram avaliadas as frases sem o ponto final e com “enter”. A finalidade deste teste foi para verificar se o Software consegue identificar quando o aluno erra ao redigir, esquecendo de colocar ponto final na frase. O Software Araponga

uniu as duas frases para fazer a comparação, alcançado a similaridade de 100% enquanto que o Software Sherlock* detectou 72% isso porque ele não analisa a frase inteira.

O gráfico da Figura 15 ilustra a eficiência do Software Araponga sobre o Software Sherlock* nas avaliações realizadas.

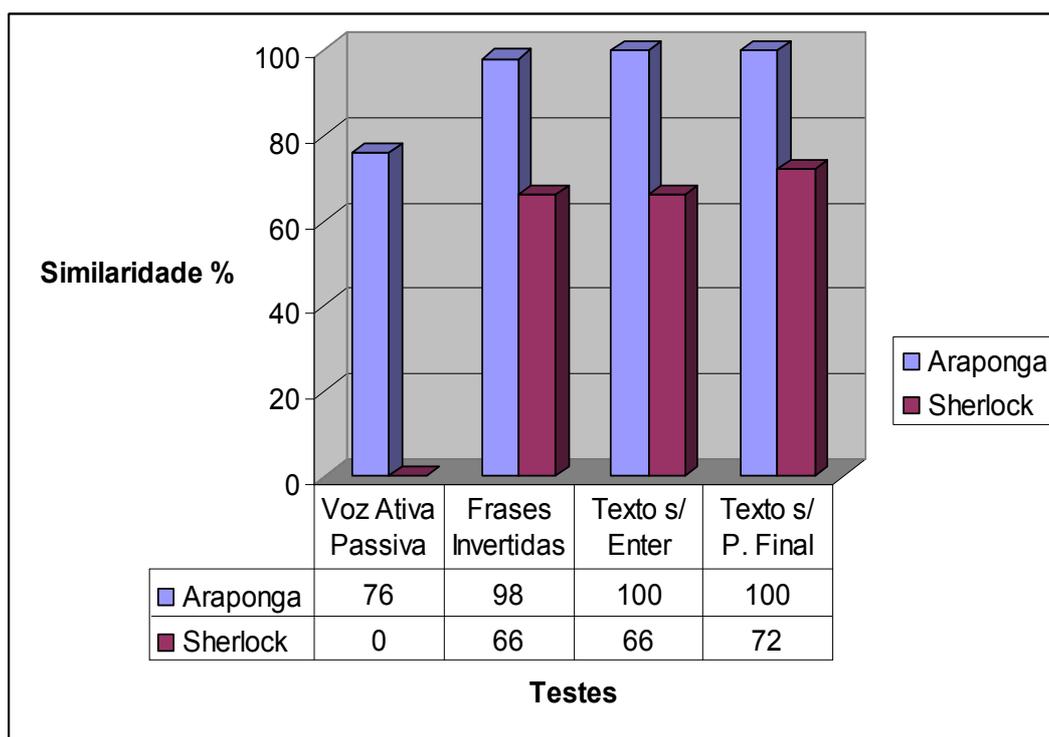


Figura 15 - Gráfico Comparativo dos Softwares

Deste modo, as várias situações foram detectadas pelo Araponga com índices de similaridade maiores que os índices obtidos no Sherlock*. No caso de voz passiva/ativa, por exemplo, o Sherlock avalia como 0% de similaridade, o que é um erro altíssimo.

Ao concluir o levantamento bibliográfico para a investigação desta dissertação, constatou-se que os Softwares disponíveis de detecção de plágio, inclusive o Software Sherlock apresentaram algumas situações antes não detectados, tais como: textos redigidos na ação verbal da voz ativa e passiva, em frases invertidas, em textos redigidos sem “enter” e em textos sem ponto final e com “enter”.

Essas deficiências foram solucionadas no Software Araponga, obtendo-se êxito nos resultados, concretizando a justificativa deste trabalho que é alcançar a detecção de plágio em diversas situações.

7 Conclusão e Trabalhos Futuros

Neste capítulo são descritas as conclusões referentes ao software e seu emprego. Finalmente são abordadas sugestões para o aperfeiçoamento da aplicação e recomendações para trabalhos futuros.

Plágio em tarefas escolares é um problema que está avançando com o tempo. Embora seja complexo desenvolver ferramentas generalizadas para detecção de plágio, fazer o mesmo para detectar de plágio em tarefas de programação é bem mais simples, uma vez que é mais fácil identificar sua estrutura do que a de uma dissertação, por exemplo.

Acredita-se que a etapa de pré-processamento pode ser a mais importante do que a aplicação do algoritmo em si.

Para testar isso, foram desenvolvidas seis etapas de pré-processamento: separação de frases, remoção de caracteres, remoção de linha em branco, remoção conectivos, remoção palavras repetidas e remoção em valores em branco.

Os resultados obtidos são positivos, já que os conjuntos de simulação pré-processados oferecem uma taxa de semelhança para textos plagiados maior do que os não pré-processados.

Nos resultados encontrados para os casos em que foi feito o pré-processamento com o software Araponga a detecção é maior do que nos casos em que o mesmo não foi realizado o pré-processamento no software Sherlock. Em particular, para texto puro, o software consegue identificar textos idênticos, sem perda de informação. Isto confirma hipótese da importância do pré-processamento.

O objetivo proposto neste trabalho que é desenvolver uma ferramenta para auxiliar o avaliador na detecção de plágio, disponibilizando um software com interface amigável e integrado ao TelEduc foi alcançado. Esta automatização do processo amplia a confiabilidade dos resultados do processo de ensino aprendizagem.

Apesar das técnicas apresentadas serem eficientes, é sempre necessária a intervenção humana para ser ter certeza de que o caso marcado como suspeito é de fato um plágio e não um mal entendido. Todavia, quanto menos ocorrer à intervenção humana, mais eficiente o algoritmo é considerado.

Ao usar aplicativos baseados no servidor para avaliar trabalho de aluno é aconselhável informar os alunos sobre o envio eletrônico de controle de autenticidade. Tais serviços mantêm uma versão de impressão digital do trabalho do estudante no banco de dados, que, por sua vez, é utilizado para verificar processos. Para IPR (2008), isto pode ser considerada uma violação de direitos autorais de propriedade intelectual dos alunos. Há casos de alunos ameaçando processar judicialmente por negligência, porque a instituição não forneceu as declarações políticas sobre as suas proibições e tratamento de plágio, conforme Heids (2003).

O Sistema será utilizado como ferramenta oficial de detecção de plágio pelo NEaD da Universidade Federal de Itajubá – UNIFEI.

O NEaD da Universidade utiliza o ambiente virtual de aprendizagem TelEduc, através do qual professores e alunos possuem aulas virtuais. O sistema provê diversas funcionalidades, como entrega de trabalhos pela WEB, o que permite aos alunos enviar seus trabalhos via navegador e aos professores que os recebam também via navegador. Esta é uma boa fonte de dados para o Sistema. Desta forma, é possível utilizar o Sistema em outros ambientes virtuais, como por exemplo: o Moodle, Tidia-Ae, etc desde que os textos a serem analisados sejam textos puros.

As ferramentas de detecção de plágio limitam-se em textos que a informação está disponível na internet ou em outras fontes eletrônicas:

- (1) Quando várias tentativas são feitas para combater ferramentas de plágio, por exemplo, utilizando paráfrase, com a ajuda de ferramentas de sinônimo, variações ou expressões diferentes para o mesmo conteúdo.
- (2) Quando o plágio está baseado em documentos, que eletronicamente não estão disponíveis (só em forma impressa, ou em arquivos que não são acessíveis para a ferramenta usada).
- (3) Quando o plágio atravessa limites de idiomas.

Trabalhos Futuros

Alguns trabalhos futuros podem ser realizados para melhoria e complementos da proposta implementada, tais como:

- Pré-processamento de textos com caracteres de controle criados em diferentes editores de texto, possibilitando a aplicação do software na comparação entre arquivos postados;
- Integração em diferentes ambientes virtuais de aprendizagem;
- Integração com diversos algoritmos de detecção de plágio, nos quais o usuário escolheria via interface gráfica qual opção ele deseja utilizar;
- Integração de outros mecanismos de buscas além do Google, possibilitando ao usuário a escolha;
- Mecanismos de controle de acesso, para impossibilitar o acesso à base de dados por usuários maliciosos;

Além disso, constatou-se que, dadas as condições, e as limitações da implementação, por exemplo: a falta de um algoritmo com técnicas utilizadas por

(WHITE; JOY,2004), detectando plágio em linguagem natural. Outra melhoria seria o uso de técnicas de inteligência artificial propostas por Engels et al. 2007.

8 Referências

ASBJORN, A. P. S.; MACDONELL, S. *Software forensics: old methods for a new science*, in proceedings of software engineering. Education and practice (se:e&p'96). IEEE Computer Society Press, 1996.

CAMPOS, G. H. B. *Avaliação em cursos on-line*. Revista TI. Disponível em <http://www.timaster.com.br/revista/colunistas/ler_colunas_emp.asp?cod=522> Acesso em: dezembro de 2008.

COSTA, Eric. *Escritório com o google – G-mail, agenda e outros serviços com a cara da sua empresa*. Revista Info Exame. Ano 21, n.º 247. São Paulo: Editora Abril, out. 2006. p. 82-83.

ECO, Umberto. *Como se faz uma Tese*. São Paulo: Editora Perspectiva, 14ª Edição, 1998.

ENGELS, S.; LAKSHMANAN, V.; CRAIG, M. *Plagiarism detection using feature-based neural networks*. ACM SIGCSE Bulletin, New York, v. 39, n. 1, p. 34-38, mar. 2007.

FRANCO, R. H. L.; SANTOS, O. A. F.; MILANEZ, C. R. J. *Implantação de um software detector de plágio para análise das questões dissertativas do ambiente virtual de aprendizagem TelEduc*. Revista Brasileira de Aprendizagem Aberta e a Distância, v.7, p.5, 2009.

GARSCHAGEN, Bruno. *Universidade em tempos de plágio*. S.d. Disponível em: <<http://www.fev.edu.br/canais/docentes/publica/principal>>. Acesso em: 24 fev. 2009.

GONÇALVES, J. A. T. *Metodologia da Pesquisa: citação indireta e paráfrase*. Disponível em <<http://metodologiadapesquisa.blogspot.com/2009/06/citacao-indireta-parafrase.html>>. Acesso em setembro 2009.

HADJI, C. *Avaliação desmistificada*. Porto Alegre: Artmed, 2001.

HEIDS, J.LIS. *Inappropriate use technical investigation process*. 2003. Disponível em <<http://www.jisclegal.ac.uk>>. Acesso em: junho 2008.

HOFFMANN, Jussara. *Pontos & Contrapontos: do pensar ao agir em avaliação*. 2ª ed. Porto Alegre: Mediação, 2000.

IPR - INTELLECTUAL PROPERTY RIGHTS: Overview, April 2008. Disponível em: <<http://www.jisclegal.ac.uk/pdfs/IPROverview.pdf>>. Acesso em: junho 2008.

JOY, M.; LUCK, M. *Plagiarism in programming assignments*. IEEE Transactions on Education, v. 42, n. 2, p. 129-133, maio 1999.

LANCASTER, T.; CULWIN, F. *Using freely available tools to produce a partially automated plagiarism detection process*. Proc. of the 21st ASCILITE Conference. 2004

LÉVY, Pierre. *A inteligência coletiva: por uma antropologia do ciberespaço*. 3. ed. São Paulo: Edições Loyola, 2000.

_____. *A Conexão Planetária: o mercado, o ciberespaço, a consciência*. Tradução de Carlos Irineu da Costa. São Paulo: Editora 34, 2001.

LIU, Y.T.; ZHANG, H.R.; CHEN, T.W.; TENG, W.G. *Extending web search for online plagiarism detection*. IEEE International Conference on Information Reuse and Integration, Las Vegas, p. 164-169, ago. 2007.

MAURER, H.; KAPPE, F.; ZAKA, B. *Plagiarism - A survey*. Journal of Universal Computer Science, v. 12, n. 8, p. 1050-1084, ago. 2006.

MCENERY T.; OAKES, M. *Authorship identification and computational stylometry*. Internal report, Department of Linguistics, Lancaster University, 1998.

MORAN, J.M. *Internet no ensino*. Comunicação & Educação. São Paulo: Paulinas, v. 14, jan./abr. 2000.

NÚCLEO DE INFORMÁTICA APLICADA A EDUCAÇÃO - NIED, Universidade Estadual de Campinas - UNICAMP. TelEduc. Disponível em: <<http://www.teleduc.org.br>>. Acesso em: maio de 2008.

OTSUKA, J.L.; FERREIRA, T.B; LACHI, R. L.; ROCHA, H.V. *Um Modelo de Suporte à Avaliação Formativa no Ambiente TelEduc*. Revista Brasileira de Informática na Educação, v 12, novembro 2003.

PAUL CLOUGH. *Plagiarism in natural and programming languages: an overview of current tools and technologies*. Department of Computer Science, University of Sheffield, Internal Report, 2000.

PERISSÉ, Gabriel. *O Professor do Futuro*. Rio de Janeiro: Thex Editora, 2002.

_____. *O conceito de plágio criativo*. Revista Videtur. Porto (Port.) / S. Paulo, v. 18, p. 9-19, 2003.

_____. *Nossos filhos não sabem usar o mouse ético*. Revista Máxxima, n. 1, ago., 2006. p.12 e 14. Disponível em: <<http://www.revistamaxxima.com.br/>>. Acesso em: 02 ago. 2008.

PERRENOUD, P. *Avaliação: da excelência à regulação das aprendizagens entre duas lógicas*. Porto Alegre: Artes Médicas, 1999.

PIKE, R. *The Sherlock Plagiarism Detector*. Março de 2007. Disponível em: <<http://www.cs.su.oz.au/~scilect/sherlock>>. Acesso em: maio de 2008.

RAMAL, Andrea Cecilia. *Avaliar na cibercultura*. Revista Pátio, Porto Alegre: Artmed, fev., 2000.

_____. *Educação na cibercultura: hipertextualidade, leitura, escrita e aprendizagem*. Porto Alegre: Artmed, 2002.

RIOS, Terezinha Azerêdo. *Compreender e ensinar: por uma docência da melhor qualidade*. 4. ed. São Paulo: Cortez, 2003.

_____. *Ética e competência*. 14. ed. São Paulo: Cortez, 2004.

ROSALES, F.; GARCIA, A.; RODRIGUES, S.; PEDRAZA, J.L.; MENDEZ, R.; NIETO, M. M.; *Detection of plagiarism in programming assignments*. IEEE Transactions on Education, v. 51, n. 2, p. 174-183, maio 2008.

SANTANA, Joseval de Melo; JOBERTO, Sérgio Barbosa Martins. *Um Sistema para Detecção de Plágio em Ambiente de Aprendizado Virtual*. Disponível em <<http://www.nuppead.unifacs.br/artigos/Paper-JobertoMartins-fpve03%20Joberto.pdf>>. Acesso em: 10 set. 2008.

SCAIFE, B. *Evaluation of Plagiarism Detection Software: Plagiarism Detection Software Report For JISC Plagiarism Advisory Service, Manchester, ver. 1.5, n. 11147, set. 2007.*

TOREKI ROB. *Plagiarism: definitions, examples and penalties*. Dezembro 1998. Disponível em: <<http://www.chem.uky.edu/Courses/common/plagiarism.html>> Acesso em: junho 2008.

UNIVERSITY OF WARWICK. *History of BOSS*. Outubro de 2006, Disponível em: <<http://www.dcs.warwick.ac.uk/boss/history.html>> Acesso em: maio de 2008.

VAZ, Telma Romilda Duarte. *O avesso da ética: ciberespaço e a questão do plágio e da cópia no ensino superior presencial*. 2006. 130 f. Dissertação (Mestrado em Educação) – Programa de Pós-Graduação em Educação, UNINOVE, São Paulo, 2006.

WHITE, D. R.; JOY, M. S. *Sentence-based natural language plagiarism detection*. ACM Journal on Educational Resources in Computing, Reino Unido, v. 4, n. 4, art. 2, dez. 2004.

WIKIPÉDIA. *Origem*: Wikipédia, a enciclopédia livre. Disponível em: <<http://pt.wikipedia.org/wiki/Wikip%C3%A9dia>> Acesso em: 28 fev. 2008.

XIN CHEN; BRENT FRANCA; MING LI; MCKINON B.; AMIT SEKER. *Shared information and program plagiarism detection*. IEEE Transactions on Education, 2004.