UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CAMPUS FRANCISCO BELTRÃO
CURSO DE LICENCIATURA EM INFORMÁTICA

Felipe Theodoro Guimarães

Desenvolvimento de um Web Crawler para Obtenção e Reprodução de Vídeos REA

Francisco Beltrão, Paraná

2017

Felipe Theodoro Guimarães

Desenvolvimento de um Web Crawler para Obtenção e Reprodução de Vídeos REA

Trabalho de Conclusão de Curso, apresentado a Universidade Tecnológica Federal do Paraná – Campus Francisco Beltrão, como parte das exigências para a obtenção do título de Licenciado em Informática.

Orientador: Prof. MSc. Wellton Costa De Oliveira

Coorientador: Profa. Doutora Maici Duarte Leite

Francisco Beltrão, Paraná

2017

Felipe Theodoro Guimarães
Desenvolvimento de um Web Crawler para Obtenção e Reprodução de Vídeos REA/ Felipe Theodoro Guimarães. – Francisco Beltrão, Paraná, 2017-
41 p. : il ; 30 cm.

Orientador: Prof. MSc. Wellton Costa De Oliveira

monografia – , 2017.

1. Web Crawler. 2. Recursos Educacionais Abertos. I. Wellton Costa de Oliveira. II. Universidade Tecnológica Federal do Paraná. III. Campus Francisco Beltrão. IV.Desenvolvimento de um Web Crawler para Obtenção e Reprodução de Vídeos REA

CDU 02:141:005.7

Felipe Theodoro Guimarães

Desenvolvimento de um Web Crawler para Obtenção e Reprodução de Vídeos REA

Trabalho de Conclusão de Curso, apresentado a Universidade Tecnológica Federal do Paraná – Campus Francisco Beltrão, como parte das exigências para a obtenção do título de Licenciado em Informática.

Francisco Beltrão, 24 de Outubro de 2017

Prof. MSc. Wellton Costa De Oliveira
UTFPR (Orientador)

Profa. Doutora Maici Duarte Leite
UTFPR (Co-orientadora)

Prof. Doutor Michel Albonico
UTFPR (Membro Convidado para Banca)

Prof. Doutor Eng. Francisco Antonio Fernandes Reinaldo
UTFPR (Preside a Banca)

Folha de Aprovação assinada encontra-se arquivada na Coordenação do Curso.

Dedico este trabalho à minha família,
especialmente minha mãe Silvia,
meus avós Reinaldo e Jandira Guimarães,
e aos futuros profissionais da educação.

AGRADECIMENTOS

Primeiramente agradeço a Deus por minhas conquistas, dentre elas eu evidencio o presente trabalho.

Agradeço à minha família por todo apoio e investimento durante os meus anos de graduação.

Agradeço também ao meu Orientador Prof. Msc. Wellton, pela sabedoria com que me guiou nesta trajetória, deste o projeto deste trabalho até sua conclusão, e por toda paciência e vontade de me incentivar a seguir na docência e não desistir deste trabalho.

Também tem meu agradecimento a minha Co-orientadora Profa. Doutora Maici por compartilhar seu conhecimento e me ajudar a fazer um trabalho melhor.

Enfim, um último agradecimento para todos os colegas que cruzaram meu caminho na universidade, aos que caminharam comigo, mas especialmente aos que tornaram-se meus amigos e fizeram desta tarefa mais suportável, e aos meus professores. Muito Obrigado.

Feliz aquele que transfere o que sabe
e aprende o que ensina.
(Cora Coralina)

RESUMO

GUIMARÃES, Felipe Theodoro. Desenvolvimento de um Web Crawler para Obtenção e Reprodução de Vídeos REA. 2017. 37 f. Monografia (Trabalho de Conclusão de Curso) - Curso Superior de Licenciatura em Informática, Universidade Tecnológica Federal do Paraná, Câmpus Francisco Beltrão. Francisco Beltrão, 2017.

Os REA são materiais digitais que diferem dos materiais comuns, como os livros e cadernos, esses materiais estão disponíveis na *internet*, gratuitamente e em uma infinidade de assuntos e idiomas, e vão além de livros, como *e-books*, eles também englobam vídeos, aplicativos, *softwares*, músicas, simulados e etc.. Porém esse tipo de mateirial é um tanto recente e não é muito popular na vida dos profissionais da educação, mesmo sabendo que o principal meio de estudo dos jovens é a *internet*, buscando exemplos diferentes e mais aula em vídeo ou apenas vídeos que possuem as informações necessárias naquele momento. O presente trabalho tem o objetivo de disponibilizar um tipo de REA, os vídeos, do Repositório de Outras Coleções Abertas (ROCA), em um ambiente *online* de vizualição de vídeos educacionais abertos, denominado neste trabalho de REA *Player*. Cada vídeo do repositório ROCA tem uma página que especifíca suas informações e disponibiliza um *link* para *download* do vídeo. A solução pensada para conseguir pegar todos os vídeos e disponibilizá-los de forma *online* no REA Player, seria através de um *WebCrawler*, que fosse capaz de rastrear todas as páginas e coletar os vídeos. Assim, foi desenvolvido um *WebCrawler*, na linguagem de programação PHP, para realizar o rastremento no ROCA e armazenar os vídeos coletados dentro de um banco de dados e do banco eles ficarem disponíveis na *web* mais especificamente no REA *Player*. Acredita-se que o movimento REA pode contribuir para educação, ele gira em torno de uma educação colaborativa, acessível e diversificada, que esta inserida nas tecnologias, ou seja, esta inserida também na sociedade e suas transformações. Este trabalho resulta em uma maneira mais simples para acessar e assistir aos vídeos do ROCA, através do REA *Player*

Palavras-chave: REA. WebCrawler. Educação. Tecnologia na educação. Vídeos.

ABSTRACT

GUIMARÃES, Felipe Theodoro. Development of a WebCrawler for Obtaining and Reproducing REA videos. 2017. 37 f. Monografia (Trabalho de Conclusão de Curso) - Curso Superior de Licenciatura em Informática, Universidade Tecnológica Federal do Paraná, Câmpus Francisco Beltrão. Francisco Beltrão, 2017.

Open educational resources (OER) are digital materials which differ from common materials, such as books and notebooks. These materials are available on the internet, free of charge and in a multitude of subjects and languages, as well as books, for instance e-books. They also encompass applications, software, music, simulation and so on. However, this type of material is somewhat new and it is not very popular for today's education professionals, even knowing that the main means of study for young people is through the internet, seeking different examples and more video lessons. The present work aims to provide a kind of OER, which are the videos, from the Repository of Other Open Collections, named ROCA, in an online environment of viewing open educational videos, called in this paper as REA Player. Each video in the ROCA repository presents a page that specifies your information and provides a link to download the video. A solution to take all the videos and make them available online in the REA Player, it would be through a WebCrawler, that should be able to track all the pages and collect the videos. Thus, a WebCrawler was developed in PHP programming language to perform the tracking in the ROCA and store the videos inside a database. And from the database they become available in the web, more specifically in the REA Player. It is believed that the REA movement is important for education, it revolves around a collaborative, accessible and diversified education that is inserted in the technologies, what means, it is also inserted in the society and its transformations. This work results in a simpler way to access and watch the ROCA videos through the REA Player.

Keywords: OER. WebCrawler. Education. Technology for education. Videos.

LISTA DE ILUSTRAÇÕES

LISTA DE ABREVIATURAS E SIGLAS

CC	Creative Commons
CSS	Cascading Style Sheets (Folha de estilos em cascata)
HTML	Hypertext Markup Language (Linguagem de Marcação de Hipertexto)
OA	Objetos de Aprendizagem
OER	Open Educational Resources
PHP	Php: Hypertext Preprocessor (Pré-processador de Hipertexto)
REA	Recursos Educacionais Abertos
ROCA	Repositório de Outras Coleções Abertas
URL	Localizador Uniforme de Recursos
UTFPR	Universidade Tecnológica Federal do Paraná

SUMÁRIO

1 INTRODUÇÃO

O presente capítulo apresenta uma contextualização sobre os principais temas deste trabalho, os Recursos Educacionais Abertos e o *WebCrawler*. Seguem descritos os capítulos objetivos, a justificativa e por fim é apresentada a estrutura completa deste trabalho.

1.1 CONSIDERAÇÕES INICIAIS

A educação é um processo importante na busca de uma sociedade melhor. Há uma mudança entre os métodos tradicionais de ensino e a forma com a qual as pessoas conseguem informações hoje (SANTOS, 2016), elas não dependem mais somente de um professor ou de livros. A fim de melhorar a educação busca-se recursos que possam dar apoio para a geração de crianças e jovens presentes na escola e nas universidades. Neste contexto, existe a necessidade de utilizar a tecnologia e os recursos educacionais para inovar o ensino e a aprendizagem.

Um exemplo de recursos educacionais são os Objetos de Aprendizagem (OA), que consistem em utilizar materiais digitais que possuem a finalidade exclusiva de ensinar. Partindo da linha de ensino dos OAs, são desenvolvidos os Recursos Educaionais Abertos (REA), eles podem abranger cursos completos, livros, vídeos, *softwares*, em suma, ou qualquer outra ferramenta que auxilie o ensino e a aprendizagem.

A cultura digital está diretamente relacionada a educação (MILL, 2010). Hoje em dia é possível buscar conteúdos educacionais com praticamente qualquer aparelho eletrônico que possua acesso a *internet*. Por isso, levando em consideração a geração atual de crianças e adolescentes, considerados socialmente como "nativos digitais", procura-se novos métodos de ensino que acompanhem a velocidade com a qual essa geração recebe informações pela *internet*.

Após realizar pesquisas no repositório ROCA, notou-se que não é possível assistir aos vídeos *online*, através do próprio repositório, ou seja, é necessário fazer *download* dos vídeos e reproduzí-los em algum *software* de reprodução de vídeos, ao contrário de sites muito populares como *YouTube*, *Vimeo* ou *Google Vídeos*. O repositório ROCA possui vídeos de REA armazenados que podem contribuir para pesquisa e educação.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

O objetivo deste trabalho é facilitar o acesso aos vídeos educacionais do ROCA através de um site *player* usando um *Web Crawler*.

1.2.2 Objetivos Específicos

- Desenvolver um *Web Crawler* para buscar e extrair as URL's de vídeos no *site* ROCA;
- Armazenar as URL's obtidas com o *Web Crawler* em um banco de dados;
- Desenvolver um *site* com *player* para os vídeos do banco de dados.

1.3 JUSTIFICATIVA

O interesse de trabalhar com os vídeos de Recursos Educacionais Abertos baseia-se na cultura audiovisual que predomina atualmente. Segundo Santos (2016), os número de estudantes que buscam vídeos na *internet* estão aumentantdo, graças a *sites* como o *YouTube* e o *Google Videos*, que popularizaram muito os vídeos, e também a cada dia ganham mais usuários.

Apesar de populares, os *sites* mencionados acima não são, sempre, de cunho educativo, e o foco deles é o entretenimento. Por isso busca-se comtemplar o meio digital com um ambiente que possua foco na educação, e nele serem inseridos vídeos com fins educacionais, vídeos estes provenientes do repositório ROCA. O ROCA reúne conteúdos de outras três importantes comunidades de repositórios, sendo elas: Recursos Educacionais Abertos (REA), Trabalho de Conclusão de Curso e Especialização (TCCE) e Trabalho de Conclusão de Curso de Graduação (TCC).

O relacionamento entre vídeos *online* e a educação está na base da revolução educacional (PANTO; COMAS-QUINN, 2013). O principal símbolo e meio do ensino tradicional, a escola, não conseguiu se adaptar com a mesma velocidade que as novas tecnologias impõe atualmente na sociedade, e ainda encontra-se despreparada às mudanças do mundo (PINHEIRO, 2014).

Para automatizar a indexação dos vídeos disponibilizados no repositório ROCA será utilizado o conhecimento em programação adquirido no decorrer da graduação para criar um robô de internet: *Web Crawler* para acessar o site do ROCA e extrair os dados necessários, excluindo assim a necessidade de indexação manual pelo autor, facilitando a coleta dos vídeos para a construção do site mencionado na seção Objetivos do trabalho.

1.4 ESTRUTURA DO TRABALHO

Este Capítulo fez a Introdução ao trabalho, descrevendo seus objetivos, justificativa e a organizção;

O Capítulo 2 apresenta o Referencial teórico dos temas centrais do trabalho;

No Capítulo 3, faz uma descrição sucinta dos materiais e métodos utilizados para o desenvolvimento do trabalho;

No Capítulo 4 são apresentados os resultados obtidos no trabalho e realiza a discussão;

Finalmente, no Capítulo 5 apresenta-se as conclusões para finalizar o trabalho e escrever as últimas considerações juntamente com sugestões para trabalhos futuros;

2 REFERENCIAL TEÓRICO

Este capítulo apresentará o levantamento teórico sobre os assuntos envolvidos neste trabalho como Objetos de aprendizagem, recursos educacionais abertos e *Web Crawler*.

2.1 OBJETOS DE APRENDIZAGEM

A aprendizagem evolui constantemente. Devido ao avanço tecnológico e os meios de comunicação, surgem também mudanças no ambiente escolar e no modo em que o ser humano tem acesso a informações que antes eram de exclusividade do professor ou da escola. O computador pessoal, depois a *internet* sendo popularizada, possibilitou as pessoas a acessarem informações disponíveis *online*, ter acesso a livros, vídeos e muito mais. Os conteúdos pedagógicos também ganharam mais espaço na internet, através da publicação de trabalhos científicos, *e-books* e o ensino a distância.

Um objeto de Aprendizagem (OA), é uma ferramenta, institucional ou não, que possui finalidades pedagógicas, com amplas possibilidades de cursos, assuntos e qualquer tipo de ensino. Este tipo de ferramenta pode ser criada em qualquer mídia (virtual, textual, vídeo, áudio, etc.) como apresentações de *slides* e *softwares* (AGUIAR; FLÔRES, 2014), e ainda, para Wiley (2000) é "um recurso digital que pode ser reutilizado para apoiar a aprendizagem". Este tipo de ferramenta pode ser uma grande aliada no trabalho do professor e na aprendizagem do aluno, pois em tese, um OA deve ser flexível e de fácil atualização.

A escolha de um OA não determina a metodologia com a qual ele será utilizado em uma aula ou atividade. Segundo Aguiar e Flôres (2014), determinar também a metodologia mais apropriada de acordo com o OA é muito importante para o desenvolvimento do pensamento crítico do aluno. Sendo uma atividade prática no computador ou no celular, na qual o aluno pode pensar em resolver os problemas através de estratégias ou hipóteses, o OA permite uma linha pedagógica também construtivista, que torna o professor um mediador do "conhecimento" que carrega o OA.

Segundo Aguiar e Flôres (2014), para um OA ser adequadamente usado em aula, o(a) professor(a) deve considerar as seguintes características:

- Linguagem apropriada para os alunos;
- Abordagem dos conceitos conforme seus interesses;
- Veracidade das informações;

- Atualização das informações.

Para compor um OA, segundo Singh (2001), ele deve ser estruturado com objetivos, conteúdos instrucionais e prática e *feedback*. Para fazer uso de um OA recomenda-se a busca por repositórios *online* de OA ou REA (COSTA, 2015), como exemplo temos o ROCA.

2.2 RECURSOS EDUCACIONAIS ABERTOS

Um REA (Recurso Educacional Aberto) pode ser classificado como um OA, porém sua ideologia e características diferem. A concepção do REA exige a possibilidade de compartilhamento, reutilização e gratuidade.

Wiley (2000) definiu o conceito de *Open Content* e também desenvolveu a *Open Content License/Open Publication License*, como forma de agregar mais ferramentas pedagógicas aos movimentos de código aberto para conteúdos (OCDE, 2015). Os REAs são também conhecidos em inglês pela sigla OER (*Open Educational Resources*).

O movimento REA é recente, teve seu início em 2001, quando o MIT (Instituto Tecnológico de Massachussets, em português) iniciou um projeto denominado *MIT Open Courseware*, que consiste em uma página *Web* com cursos e materiais didáticos, produzidos pelos professores do MIT e sob licenças da *Creative Commons* (CC) (PANTO; COMAS-QUINN, 2013). Neste contexto, em 2005 foi criado um consórcio internacional chamado *Open Courseware Consortium*, com objetivo de ampliar o acesso a materiais didáticos e melhorar a qualidade da educação.

Em 2002 a UNESCO utilizou pela primeira vez o termo *Open Educational Resource* (Recurso Educacional Aberto, em português), em uma conferência e o definiu como "materiais digitalizados oferecidos livres e abertos para utilização e reutilização no ensino, na aprendizagem e na pesquisa" (OCDE, 2008).

O REA gira em torno da prática colaborativa e aberta, onde tudo poderá ser reutilizado ou estudado livremente, daí vem a importância da licença *Creative Commons* e a tecnologia, pois é necessário que o REA tenha sua licença claramente identificada. Os REAs abrangem complementos para aulas tradicionais ou não tradicionais, para cursos a distância, conteúdos de mídia visual, auditiva e etc. E seus três princípios fundamentais são:

- Os materiais devem ter valor educacional;
- Um conteúdo somente é REA se, for totalmente aberto, sem custos e restrições, estando disponível para reutilização, revisão, recontextualização e redistribuição;

- As tecnologias devem ser capazes de dar suporte ao desenvolvimento e as questões pedagógicas dos REAs.

Desde a elaboração da Declaração da Cidade do Cabo para Educação Aberta, na qual a educação recebe liberdade de uso, personalização, melhoria e distribuição (CAPETOWN OPEN EDUCATIONAL DECLARATION, 2007), estabelece que recursos educacionais abertos abrangem planos de aulas, livros, jogos, *softwares*, periódicos e outros materiais que possam apoiar o ensino, ou seja, podendo também ser um OA. Também são considerados recursos as tecnologias abertas, que são *softwares* que criam ou melhoram os conteúdos de aprendizagem. Para os REAs terem seus objetivos realizados, eles devem estar sob licenças que garantem a sua função, no caso a *Creatve Commons* é responsável por esse direito.

A licença CC foi criada a partir da liceça *Copyleft*, onde o autor decide os direitos que serão reservados e alguns que serão liberados. Existe o direito padrão de cópia e reprodução para todos os tipos de licença CC. É importante frisar que um conteúdo, ou obra, aberto e acessível não admite plágio, o autor deve ser reconhecido, referenciado e citado. Infelizmente a maioria das obras *onlines* não estão sob uma licença CC.

Somente é possível atribuir uma licença CC em uma obra na qual você é o autor, e que não esteja inserido na mesma conteúdos de terceiros, principalmente conteúdos com *Copyright*.

O REA tem grande importância para educação em qualquer lugar. São ferramentas digitais que podem ser facilmente compartilhadas e adotadas no cotiano de qualquer professor, instituição e alunos. Apesar do receio que a maioria dos profissionais da educação têm em usar materiais que não são de autoria própria, os conteúdos digitais estão muito acessíveis, e os repósitórios devem especificar a licença de seus materiais. Três características que se destacam sobre os REAs:

- Acessibilidade: facilidade em ser acessada, encontrada, não precisar de senhas e etc. Também está ligada a facilidade de compreensão para pessoas com alguma incapacidade ou limitação: física, neurológica, auditiva, visual e etc.
- Reutilização: afirma a possibilidade do conteúdo em ser usado várias vezes, com ou sem modificações;
- Interoperabilidade: utilizar mídias em formatos que possam ser adaptadas ou modificadas. Está ligado ao quão bem o conteúdo funciona em outros dispositivos, *softwares* e etc.

A educação aberta tem potencialidades no ensino formal e tradicional. Os REA integram ações por uma educação aberta. O principal objetivo destas ferramentas é

facilitar e enriquecer o trabalho do educador e da instituição, bem como espalhar materiais com qualidade que possam ser usados livremente.

Existe também alguns fatores que podem prejudicar a utilização dos REA. Por exemplo, segundo Baker (2007), não foi atribuído um modelo padrão comum entre os países para análise e precisão dos conteúdos educacionais abertos, e ainda os materiais podem não atender aos requisitos de acessibilidade da seção 508[1] ou da *SCORM*[2].

O site oficial www.rea.net.br descreve a confiabilidade dos REA e ainda sobre sua qualidade: "Você pode confiar tanto ou mais em REA quanto confia em conteúdos educacioanis proprietários [...]". Ele ainda conta que existem projetos ganhando espaço, tanto no Brasil quanto no mundo, que são responsáveis por filtrar e analisar a qualidade dos REA lançados atualmente como o *openstax*[3], que foca em livros didáticos abertos. Neste mesmo *site* existe uma lista[4] com vários repositórios onde pode-se encontrar REA com licenciamento unificado, como exemplo: *MOODLE*, *WIKIPEDIA*. Para projetos com REA no Brasil existe uma outra página com mais *links*[5] que incluem sites como *SCIELO Books*, *Wikimedia* Brasil e etc.

É preciso lembrar, cabe ao educador que vai utilizar um REA, analisar o conteúdo e verificar sua eficácia antes de integrá-lo a sua prática educativa.

2.3 WEB CRAWLER

Segundo Pinkerton (2000) quando a *web* ainda era pequena, os pesquisadores conseguiam pesquisar documentos seguindo *links* de hipertextos que levaria de um documento para outro ou poderiam acessar diretamente um documento se soubessem seu endereço correto.

O primeiro motor de busca de texto completo da *web*, o *WebCrawler*, surgiu em 1994. Esta ferramenta contribuiu para o crescimento da *web*, e hoje é o motivo da facilidade com a qual milhões de pessoas conseguem fazer buscas na internet e em alta velocidade. Sua invenção e evolução, de 1994 a 1997, criaram um novo modo de navegar por hipertexto, que é procurando.

O *WebCrawler* foi construído como um serviço da *web* que automatiza a tarefa de cruzamento entre *links* e cria um índice pesquisável da *web*. Ele precisa criar um índice das páginas que ele encontra o objetdo que foi buscado. Para criar este índice automaticante, o *WebCrawler* precisa de uma semente *url*, ou seja, um endereço da *web* que sirva de início,

[1] Lei de Reabilitação de 1973 – Toda tecnologia eletrônica e de informação utilizada pelo governo federal deve ser acessível para pessoas com deficiência.
[2] Conjunto de normas técnicas para o desenvolvimento de materiais virtuais de aprendizagem
[3] Site: https://cnx.org/
[4] Lista de REA no mundo: http://www.rea.net.br/site/mao-na-massa/iniciativas-rea/rea-no-mundo/.
[5] Lista de alguns repositórios no Brasil: reabrasil2011@gmail.com.

e que necessariamente esteja ligado aos próximos endereços que contenham objetos de interesse do pesquisador.

O *Google* popularizou a busca por documentos e sites na *web* através de *links*. Toda busca realizada no *Google* resulta em uma imensa lista de *links* que são os possíveis caminhos para levar ao que o usuário esta procurando, isso graças a utilização de ferramentas como *WebCraler*. Através de um *index*, ou índice, ele rapidamente encontra os recursos procurados, caracterizando assim as duas principais fuções de um *WebCrawler*: construir um índice da *web*; e navegar automaticamente sob demanda.

Componentes de *software* do primeiro *WebCrawler* (PINKERTON, 2000):

- Mecanismo de busca: responsável pelo redirecionamento das buscas e atividades do *Crawler*;
- Agentes: para segurar os documentos obtidos na *internet* são invocados os agentes pelo mecanismo de busca. São empregados vários agentes com objetivo de retornar um documento, ele precisa obter uma *url* e em seguida montar um arquivo com os dados descritivos do documento visitado;
- Banco de dados: funciona com duas peças, contém um índice em *full-text* (texto completo em português) dizendo quais *url's* já foram visitadas e outra com url's que serão visitadas;
- Servidor de consulta: ele implementa ao Índice *WebCrawler*, um serviço de busca disponível via documentos na *web*, através de palavras-chave.

A imagem abaixo ilustra, em tese, o *WebCrawler* de Pinkerton em 1994:

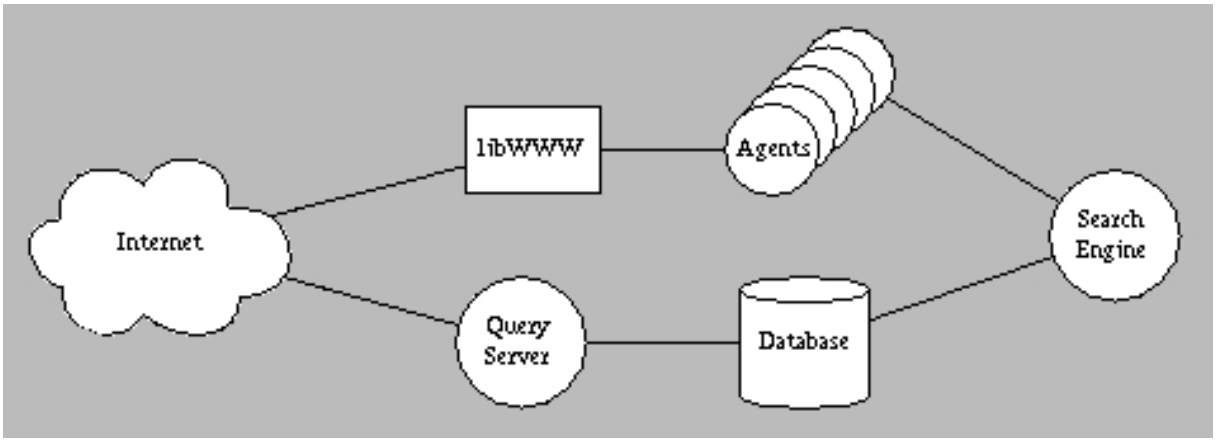


Figura 1 – Modelo do WebCrawler de Pinkerton - 1994

Fonte: The Design of the WebCrawler. Pinkerton (2000).

Atualmente existem dois tipos comuns de *WebCrawler* desenvolvidos, ele pode ser perioódico ou incremental.

1. *WebCrawler* preriódico: é mais limitado quanto a atualização de seu índice. A cada rastreamento que faz ele atualiza seu índice inteiro, substituindo o antigo por um novo;

2. *WebCrawler* incremental: ele opera em um modo estável, ele coleta páginas diariamente, é mais difícil de ser implementado, porém atualiza toda toda sua coleção de indice mesmo enquanto esta operando o rastreamento de páginas.

O *WebCrawler* desenvolvido para o presente trabalho é o do tipo periódico.

3 MATERIAIS E MÉTODO

Neste capítulo será mostrado de maneira sucinta como foi desenvolvido as ferramentas propostas neste trabalho, que são o *Web Crawler* para indexação de dados automaticamente no banco de dados e o player dos vídeos para o usuário final.

3.1 MATERIAIS

Foi utilizado como gerenciador do banco de dados o *software* ***MySQL Workbench***. Banco de dados é um sistema que serve para armazenar conteúdos digitais, como um repositório, onde as pessoas podem recuperar qualquer conteúdo que tenha sido armazenado nele. Seus componentes principais são: dados, *software*, *hardware* e os usuários (DATE, 2004). O *MySQL* é uma ferramenta de criação de banco de dados gratuíta.

Para administração e gerenciamento do conteúdo do banco de dados foi utilizado um Sistema de Gerenciamento de Banco de Dados (SGBD) chamado ***PHPMyAdmin***. Esta ferramenta é livre, gratuita e suporta praticamente todas as operações do *MySQL*.

O desenvolvimento do *WebCrawler* e da página *web* foram implementados através do editor de texto ***Sublime Text***. Ele suporta linguagens de programação *web* e pode ser usado gratuitamente.

Foi necessário fazer o *download* de um programa para criar um servidor local para testar o banco de dados e o *WebCrawler* na internet. O programa escolhido para executar tal tarefa foi o ***WampServer***, que é um ambiente de desenvolvimento que funciona no sistema operacional *Windows* e também é gratuito.

3.2 VISÃO GERAL

Depois de realizar a pesquisa com os vídeos no repositório ROCA, foram coletadas as informações comuns de cada vídeo para poder criar um banco de dados que armazenaria essas informações.

A Figura 2 mostra a visão geral do sistema proposto. A partir da semente URL que é dada ao *WebCrawler*, ele começa seu rastreamento por páginas com vídeo, mais especificamente páginas do repositório ROCA. No passo 1: Do repositório ROCA, são coletadas páginas em HTML. Quando ele acha página com vídeo, é retornado um *link* para o próximo rastreamento. No passo 2: o *WebCrawler* analisa e extrai os dados dos vídeos que serão armazenados no banco de dados. No passo 3: o banco de dados e a página comunicam-se pela *internet* onde os vídeos estarão disponíveis *online*, para enviar os vídeos

e suas informações e também para a página acessar o banco e extrair a informação que se pede.

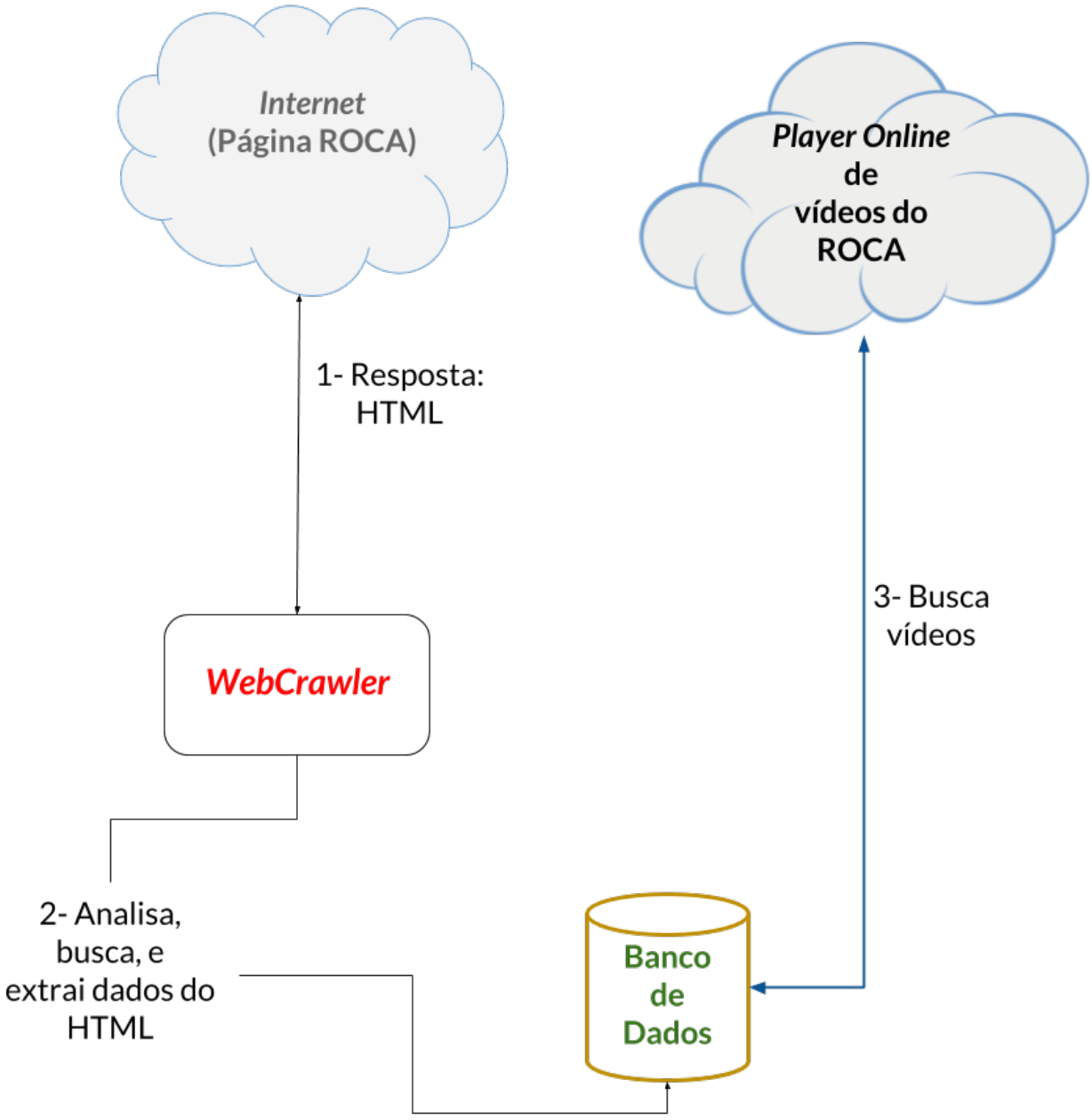


Figura 2 – Visão geral do sistema proposto.

Fonte: Desenvolvida pelo autor.

3.3 CRAWLER

Nesta seção serão descritas, detalhadamente, as partes fundamentais para construção do *Web Crawler*. A Figura 3 mostra o fluxograma proposto.

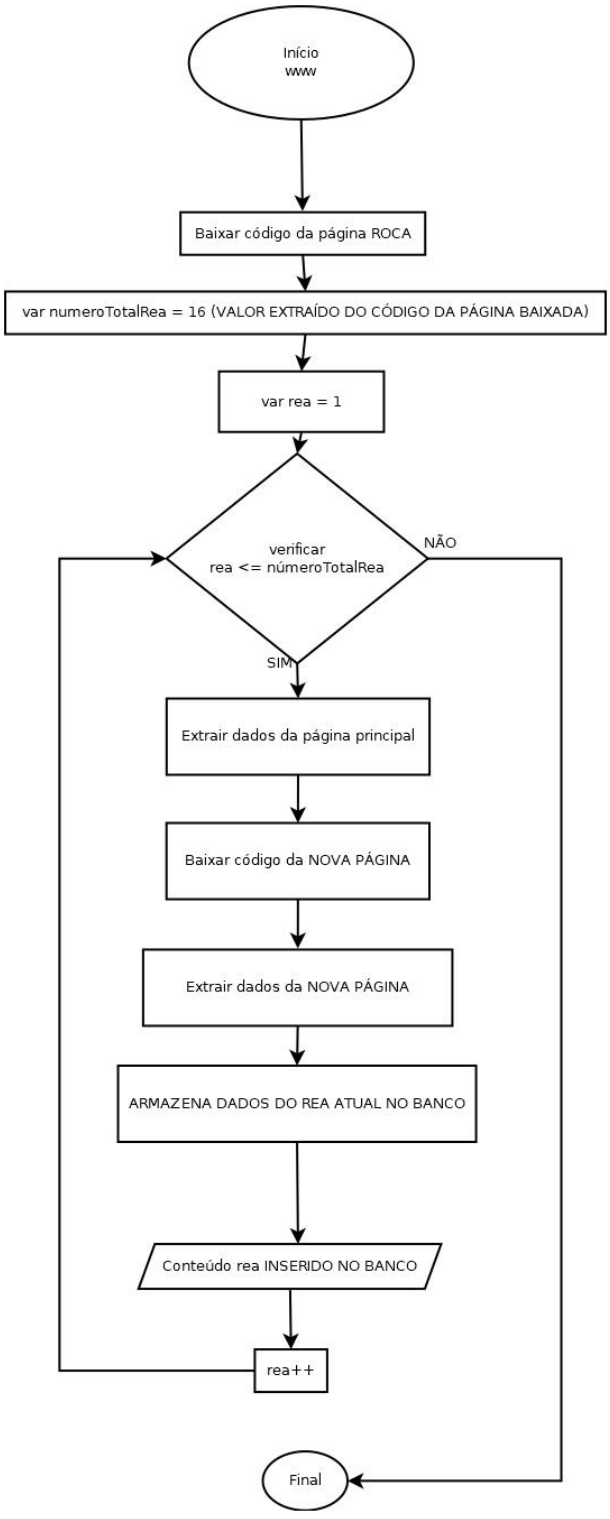


Figura 3 – Arquitetura do *Web Crawler* para REA.

Fonte: Desenvolvida pelo autor.

Para baixar o código da página *web*, primeiro foi informado para o *crawler* qual é a *url* que ele deve acessar para começar seu rastreamento. Ela acessa a página com essa *url*, faz *download* do código fonte da página, seleciona um trecho de código que é delimitado

pelo autor, pois ele sabe que aquele contém a tabela informa o caminho para todos os vídeos do repositório:

```
$pagina_roca =
file_get_contents("http://repositorio.roca.utfpr.edu.br/jspui/browse?
type=tipologia&order=ASC&rpp=20&value=video");
```

Separando o código:

```
$pagina_roca = retirarEntre($pagina_roca, '</th></tr><tr>', "</table>");
```

Foi criada a variável "páginaroca"que chama a função "retirarEntre"e atribui para a variável "páginaroca"o código que esta delimitado entre as *tags* </th></tr><tr> e </table>.

Com a linha seguinte:

```
$separa = explode("<tr>", $pagina_roca);
```

é feita uma quebra quando o *crawler* chega até a *tag* <tr> pois em HTML5 cada "tr" representa uma linha de uma tabela. Até aí o *crawler* conseguiu extrair toda tabela de vídeos da página ROCA.

A função "retirarEntre" foi programa da seguinte maneira:

```
function retirarEntre($conteudo, $inicio, $fim){
    $r = explode($inicio, $conteudo);
    if (isset($r[1])){
        $r = explode($fim, $r[1]);
        return $r[0];
    }
    return '';
}
```

A função recebe os seguintes parâmetros: "conteúdo", "inicio" e "fim", pois são necessários para receberem os dados separados das linhas de código. Uma variável ("r") recebe explode(inicio, conteudo); onde início e conteúdo são as *tags* de delimitação lidas pelo *crawler* para filtrar o código. A seguir, a condição *"if"* que indica o fim do processo de separação, se for verdade que "r" recebeu os valores anteriores, a variável "r" recebe a quebra do "fim" e do código contido em "r", e a função retorna a posição 0 em "r".

O comando totalreas = count(separa) é um contador, que recebe a variável "separa", para mostrar o total de vezes que o *crawler* executou a quebra, representando assim o número de vídeos encontrados dentro da tabela.

Após isso o Agendador tem uma condição para o *crawler* coletar as informações de cada vídeo que a tabela cont, um laço de repetição (FOR) que só para de ser executado enquanto o número total de reas não é alcançado:

```
$rea = 1; for($x=0; $x < $total_reas; $x++){.
```

A variável "sql" é criada com valor vazio: sql = ; para que, futuramente, comece a reunir cada informação dentro dela, conforme o *crawler* vai localizando. Em seguida é necessária uma variável que vai receber a mesma função do início (retirarEntre) para o *crawler* poder filtrar o código:

```
separa1 = retirarEntre(separa[x], "<td headers=\"t1\" >", "</tr>");.
```

Estas *tags* «td headers=ẗ1ö"e «/tr> delimitam uma linha da tabela, a primeira neste caso, para, novamente, ser filtrada e o *crawler* retirar a informação Campus, que é a primeira coluna da tabela:

```
campus = retirarEntre(separa1, "<em>", "</em>");
$sql = $sql . " '$campus', ";
```

em seguida armazenar campus na variável "sql": sql = sql . "'campus', ". A variável "campus" recebeu "retirarEntre" que tinha o código da primeira linha trazido pela variável "separa1", e filtrou o campus delimitado pelas *tags*: «em>"e «/em>", e finalizou adicionando a informação de "campus" dentro de "sql".

As próximas colunas que contém o ano e o nome dos autores seguiram a mesma estrutura do campus para serem coletadas, o que mudou foram as *tags* que delimitavam o código de cada uma das colunas e as variáveis que armazenaram suas informações:

```
$ano  = retirarEntre($separa1, "\">", "</td>");
$sql = $sql . " '$ano', "  ;

$autores = retirarEntre($separa1, "<td headers=\"t4\" >", "</tr>");
$sql = $sql . " '$autores', ";
```

Ainda dentro do laço de repetição, foi implementada uma fila (*QUEUE*) de url's para entrar na página que tem realmente o vídeo hospedado.

Este trecho:

```
$link_pagina_video = retirarEntre($separa1, "<a href=\"", "\">");
```

possui a variável “linkpaginavideo” chamando a função “retirarEntre” que contém “separa1” e a delimitação das *tags* <a href=¨ e "¨>"que filtra o código de “separa1” para levar ao *link* de um vídeo. Com isso, a variável “linkpaginavideo” recebe um endereço do repositório que vai servir para coletar todos os outros vídeos, e mais o *link* da página do vídeo, formando apenas uma url:

```
$link_pagina_video = "http://repositorio.roca.utfpr.edu.br" .
$link_pagina_video;
e $sql = $sql . " '$link_pagina_video', ";
```

Foi criada uma função chama “extrairvideo”, que recebe como parâmetro a variável “link_ pagina_video”. Dentro da função é criada a variável “sql_pagina_video” pois são as informações que posteriormente serão armazenadas no banco de dados. Outra variável: “paginavideo” recebe o conteúdo do parâmetro da função:

```
$pagina_video = file_get_contents( $link_pagina_video );.
```

Contendo isso a variável “paginavideo” pode agora filtrar o código entre as tags: <table class=¨table itemDisplayTable¨>"e «/td></tr></tbody></table>"que informam o *link* do vídeo:

```
$pagina_video = retirarEntre($pagina_video, "
<table class=\"table temDisplayTable\">", "</td></tr></tbody></table>");.
```

Com isso, a “paginavideo” pode ser usada para o *crawler* rastrear as demais informações de cada vídeo: título, palavras-chave, referencia, resumo, *abstract*, descrição e o vídeo.

Para começar a extração das informações mencionadas acima, foi programado da seguinte forma, tomamos como exemplo o título:

```
$titulo = retirarEntre($pagina_video, "<td class=\"metadataFieldValue
dc_title\">", "</td>");
```

foi criada uma variável para o título e atribuida a ela a função “retirarEntre”, com o *link* da página que contém o vídeo (paginavideo) mais as tags para filtrar o título de fato. A seguir é executada:

```
$sql_pagina_video = $sql_pagina_video . " '$titulo', ";
```

para a variável “sql_pagina_video” guardar o título do vídeo . Seguiu-se este método para extrair as demais informações, o que foi necessário alterar foram as variáveis e as *tags* que delimitam cada informação contida no código. Para extração do vídeo foi programada da seguinte maneira:

```
$url_video = @explode("mp4", $pagina_video);
```

onde a variável “urlvideo” recebe a quebra do código em “paginavideo” que é feita em “mp4”. E logo executa:

```
$url_video = @retirarEntre($url_video[2], "<a class=\"btn btn-primary\"
target=\"_blank\" href=\"", " ");
```

para executar a função “retireEntre” e extraia o vídeo.

A variável “dados_da_pagina” agora recebeu todas as informações sobre o vídeo e o vídeo, e a seguir passa a ser guardada também na variável “sql”.

Por fim foi realizado a inserção, dos dados extraidos do ROCA, dentro do banco de dados:

```
$query = "INSERT INTO videos (campus, ano, autores, link_para_pagina, titulo,
palavras_chave, referencia, resumo, abstract, descricao, link_video) VALUES
( $sql )";.
```

É enviada uma consulta ao banco ativo no servidor:

```
$query = mysqli_query($conexao, $query) or die(mysqli_error($conexao));
```

se for bem sucedida é exibida na tela uma mensagem sobre a inserção usando *crawler* e caso a *query* falhe é mostrado um erro na tela:

```
echo "<font color=green>[OK] REA $rea INCLUÍDO COM SUCESSO USANDO CRAWLER
AUTOMATICAMENTE</font><br><br>";
```

3.4 BANCO DE VÍDEOS

Esta seção é dedicada ao banco de dados, utilizado para armazenar todas as informações extraídas sobre o vídeo com o *crawler*.

Os vídeos são armazenados em um banco, para posteriormente poder exibir todos na nova página. A imagem acima mostra a tabela “Vídeos” criada para receber as informações que o *crawler* coletou no repositório ROCA durante seu rastreamento.

Foi criada uma coluna para cada informação e mais o “id” como chave primária para identificação das informações, impedindo que elas fossem duplicadas.

videos

Coluna	Tipo	Nulo	Predefinido	Comentários
id *(Primária)*	int(3)	Não		
campus	varchar(500)	Não		
ano	varchar(500)	Não		
autores	varchar(500)	Não		
link_para_pagina	varchar(500)	Não		
titulo	varchar(500)	Não		
palavras_chave	text	Não		
referencia	varchar(500)	Não		
resumo	text	Não		
abstract	text	Não		
descricao	text	Não		
link_video	varchar(500)	Não		
data_hora	timestamp	Não	CURRENT_TIMESTAMP	

Índices

Nome da chave	Tipo	Único	Pacote	Coluna	Quantidade	Agrupamento (Collation)	Nulo	Comentário
PRIMARY	BTREE	Sim	Não	id	16	A	Não	

Figura 4 – Tabela de vídeos do banco de dados.

Fonte: Desenvolvida pelo autor.

3.5 PLAYER DE VÍDEOS

O *player* de vídeo foi desenvolvido na linguagem PHP e implementado em uma página *web*, desenvolvida em HTML5 e CSS3. Também foi implementada uma lista com os demais vídeos coletados no banco de dados.

O *player* está inserido em uma condição (if), onde ele verifica se o "id" do vídeo foi selecionado, para identificar no banco de dados qual vídeo deve ser mostrado no *player*.

```
if($_GET["id"]!=""){

    $id=$_GET["id"];

    $video_escolhido = mysqli_query($conexao, "SELECT * FROM videos
    WHERE id=$id");
```

Se o "_GET["id"]" é diferente de zero, a variável "id" recebe o id do banco de dados. A variável "videoescolhido" recebe uma consulta na tabela de vídeos do banco de dados especificando o id para ter acesso ao vídeo que deseja assistir. Em:

```
$video = mysqli_fetch_array($video_escolhido);
```

a variável "video" guarda os dados em índices numéricos na matriz do resultado, mysqli_fetch _array() é uma função que pode também guardar os dados em índices associativos, usando os nomes dos campos do conjunto de resultado como chave. Tendo especificado que era para retirar os dados da linha consultada por "video_escolhido".

Foram filtrados da tabela o *link* do vídeo e o título usando a seguinte programação:

```
$link_video = $video["link_video"]; e
$titulo     = $video["titulo"];
```

Tendo isso pode ser apresentado para o usuário final na página o *player* com o vídeo e seu respectivo título:

```
echo "<video src='$link_video' controls width=100% height='100%' autoplay>
</video><br>";
```

para imprimir o *player* e echo <h2>$titulo</h2> para imprimir título.

Também foi programado para aparecer o restante dos vídeos do banco, resumidamente, na página *web*. Em:

```
$videos = mysqli_query($conexao, "SELECT * FROM videos");
```

uma variável denominada "videos" recebeu da consulta sql da tabela videos. Foi criado um laço de repetição para extrair de dentro da tabela vídeos o id e título de todos os vídeos da tabela:

```
while($video = mysqli_fetch_array($videos)){   $id     = $video["id"];
$titulo
= $video["titulo"];
```

E por fim, é disponibilizado o link para cada vídeo coletado do repositório:

```
echo "<div id='miniatura'><a href='player.php?id=$id'><img
src='miniatura.png'>
<br> $titulo </a> </div>";
```

exibindo uma miniatura de imagem com símbolo de *play* e o título do vídeo ao qual corresponde o *link*.

3.6 IMPLEMENTAÇÃO DO SISTEMA

Para ser implementado, o *crawler* foi organizado em partes, para organização do autor.

- Parte 1

```
88
89  //PARTE 1
90  //World Wide Web
91  $pagina_roca = file_get_contents("http://repositorio.roca.utfpr.edu.br/jspui/
        browse?type=tipologia&order=ASC&rpp=20&value=video");
92
93
94  $pagina_roca = retirarEntre($pagina_roca, '</th></tr><tr>', "</table>");
95
96  $separa = explode("<tr>", $pagina_roca);
97
98  $total_reas = count($separa);
99
100 echo "Número de REAs encontrados: " . $total_reas ."<br><br>";
101
```

Figura 5 – Printscreen do da parte 1 do código.

Fonte: Autoria própria.

A parte 1 do código (Figura 5) não está escrita no começo, pois foi organizado de forma que as funções do programa ficassem localizadas na parte superior do código-fonte.

- Parte 2

```
102 //PARTE 2
103 //MULTI-THREADED DOWNLOADER E SCHEDULER (agendador)
104 $rea = 1;
105
106 for($x=0; $x < $total_reas; $x++){
107
108     $sql = "";
109
110     $separa1 = retirarEntre($separa[$x], "<td headers=\"t1\" >", "</tr>");
111
112
113     //CAMPUS
114     $campus = retirarEntre($separa1, "<em>", "</em>");
115     $sql = $sql . " '$campus', ";
116
117     //ANO
118     $ano  = retirarEntre($separa1, "\">", "</td>");
119     $sql = $sql . " '$ano', "  ;
120
121     //AUTORES
122     $autores = retirarEntre($separa1, "<td headers=\"t4\" >", "</tr>");
123     $sql = $sql . " '$autores', ";
124
```

Figura 6 – Printscreen da parte 2 do código.

Fonte: Autoria própria.

A parte 2, representado pela Figura 6, refere-se a construção do agendador. Ele coletou todos os *links* que seriam percorridos pelo *crawler* para chegar aos vídeos.

- Parte 3

```
128     //PARTE 3, LINK PARA ENTRAR NA PAGINA QUE TEM VIDEO
129     //QUEUE URL (FILA DE URLs)
130     $link_pagina_video = retirarEntre($separa1, "<a href=\"", "\">");
131
132
133     $link_pagina_video = "http://repositorio.roca.utfpr.edu.br".$link_pagina_video;
134     $sql = $sql . " '$link_pagina_video', ";
135
136     $dados_da_pagina = extrair_video( $link_pagina_video );
137     $sql = $sql . $dados_da_pagina;
138
```

Figura 7 – Printscreen da parte 3 do código.

Fonte: Autoria própria.

A parte 3 (Figura 7) descreve a criação da fila de url's de onde se extraem os vídeos, e o armazenamento final no banco de dados, depois que todas os vídeos foram coletados.

- Parte 4

```
7
8   //PARTE 4
9   //FUNÇÃO QUE SELECIONA O QUE ESTA DELIMITADO
10  function retirarEntre($conteudo, $inicio, $fim){
11      $r = explode($inicio, $conteudo);
12      if (isset($r[1])){
13          $r = explode($fim, $r[1]);
14          return $r[0];
15      }
16      return '';
17  }
18
```

Figura 8 – Printscreen da parte 4 do código.

Fonte: Autoria própria.

A parte 4 (Figura 8) descreve a criação da função "retirarEntre", responsável por filtrar os códigos rastreados pelo *crawler*.

- Parte 5

  A Figura 9 mostra o código da função "extrair_video".

  Nesta função que acontece o armazenamento das demais informações que o *crawler* extrai do repositório para guardar no banco de dados.

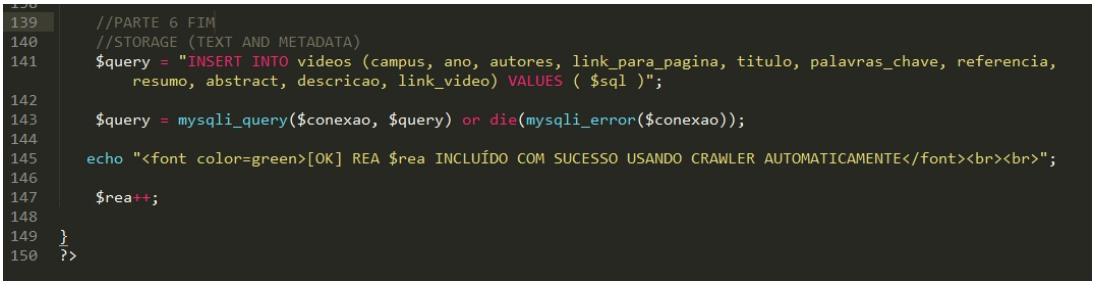
- Parte 6

  A Figura 10 mostra a parte da inserção, de todas as informações sobre o vídeo, incluindo o vídeo, dentro do banco de dados.

```
//PARTE 5
//FUNÇÃO PARA EXTRAIR INFORMAÇÕES E O VIDEO
function extrair_video($link_pagina_video){

    $sql_pagina_video = "";

    $pagina_video = file_get_contents( $link_pagina_video );
    $pagina_video = retirarEntre($pagina_video, "<table class=\"table itemDisplayTable\">", "</td></tr></tbody></table>");

    //TITULO
    $titulo = retirarEntre($pagina_video, "<td class=\"metadataFieldValue dc_title\">", "</td>");
    $sql_pagina_video = $sql_pagina_video . " '$titulo', ";

    //PALAVRAS-CHAVE
    $palavras_chave = retirarEntre($pagina_video, "<td class=\"metadataFieldValue dc_subject\">", "</td></tr>");
    $sql_pagina_video = $sql_pagina_video . " '$palavras_chave', ";

    //REFERENCIA
    $referencia = retirarEntre($pagina_video, "<td class=\"metadataFieldValue dc_identifier_citation\">", "</td>");
    $sql_pagina_video = $sql_pagina_video . " '$referencia', ";

    //RESUMO
    $resumo = retirarEntre($pagina_video, "<td class=\"metadataFieldValue dc_description_resumo\">", "</td>");
    $sql_pagina_video = $sql_pagina_video . " \"$resumo\", ";

    //ABSTRACT
    $abstract = retirarEntre($pagina_video, "<td class=\"metadataFieldValue dc_description_abstract\">", "</td>");
    $sql_pagina_video = $sql_pagina_video . " \"$abstract\", ";

    //DESCRICAO
    $descricao = retirarEntre($pagina_video, "<td class=\"metadataFieldValue dc_description\">", "</td>");
    $sql_pagina_video = $sql_pagina_video . " '$descricao', ";

    //VIDEO MP4
    $url_video = @explode("mp4", $pagina_video);
    $url_video = @retirarEntre($url_video[2], "<a class=\"btn btn-primary\" target=\"_blank\" href=\"", " ");
```

Figura 9 – Printscreen da parte 5 do código.

Fonte: Autoria própria.



```
    //PARTE 6 FIM
    //STORAGE (TEXT AND METADATA)
    $query = "INSERT INTO videos (campus, ano, autores, link_para_pagina, titulo, palavras_chave, referencia,
        resumo, abstract, descricao, link_video) VALUES ( $sql )";

    $query = mysqli_query($conexao, $query) or die(mysqli_error($conexao));

   echo "<font color=green>[OK] REA $rea INCLUÍDO COM SUCESSO USANDO CRAWLER AUTOMATICAMENTE</font><br><br>";

    $rea++;

}
?>
```

Figura 10 – Printscreen da parte 6 do código.

Fonte: Autoria própria.

Tanto o site para o *player* dos vídeos, quanto o *crawler*, foram desenvolvidos na mesma *engine*: o *Sulbime Text*. Foram utilizados os navegadores *web*: *Google Chrome* e o *Mozilla Firefox* para interpretar a ferramenta e poder avaliar sua funcionalidade.

Neste trabalho, o *crawler* se limita a rastrear e armazenar apenas conteúdos de vídeo dentro do repositório ROCA, porém não é apenas neste repositório que existem vídeos educacionais abertos. É possível que um *crawler* possa rastrear toda a *web* para achar conteúdos deste tipo, porém não é o objetivo do presente trabalho.

É interessante, pensando em trabalhos futuros, agregar mais vídeos, de outras fontes e repositórios além do ROCA, para a página. Desde que os conteúdos sejam também REA.

O site foi públicado sobre a licença *CREATIVE COMMONS* (Figura 11) Atribuição-NãoComercial 4.0 International (CC BY-NC 4.0), o que significa que qualquer pessoa esta livre para:

- Compartilhar: copiar e distribuir o material em qualquer meio ou formato; e
- Adaptar: remixar, transformar e construir sobre o material.

Figura 11 – Símbolo da licença CC BY-NC 4.0.

Fonte: creativecommons.org/licenses/by-nc/4.0/.

Sobre os termos da licença, a atribuição indica que você deve dar crédito apropriado, fornecer um *link* para a licença e indicar se alterações foram feitas. Você pode fazê-lo de forma razoável, mas não de alguma forma que sugira que o autor aprove você ou o seu uso. O termo não-comercial indica que ninguém, incluindo o autor, pode usar o material para fins lucrativos.

Por fim o autor, ou qualquer outro usuário, não pode mais aplicar nenhum tipo de restricões ao material que possa restringir outros de fazer algo permitido pela licença.

4 RESULTADOS

Este capítulo apresenta o que foi obtido como resultado do trabalho, que, em princípio, é um *WebCrawler* para rastreamento e extração de vídeos no ROCA e uma página *Web* para exibição dos vídeos. Serão exibidos *printscreens* da página e da programação do *crawler* em seu estado final. Também serão discutidas algumas limitações e futuras ideias para um *crawler*.

4.1 ESCOPO DO SISTEMA

A ferramenta *WebCrawler* foi desenvolvida para facilitar o trabalho de rastrear as páginas no repositório ROCA e extrair as informações do vídeos para enviar ao banco de dados. O trabalho dela é sempre coletar o código fonte das páginas e rastrear os valores que o autor especifica na sua programação. Existem vários tipos de materiais contidos no repositório, por isso é importante especificar de onde a ferramenta deve começar seu trabalho. O acesso dos usuários na página do *player* pode ser feito através de algum navegador da *web*, mas não se restringe ao computador, funciona em navegadores de *tablet* e *smartphones*.

De início, o rastreamento do *crawler* vai apenas para os 16 vídeos educacionais que contém dentro do repositório ROCA, por isso é nesse repositório que o *WebCrawler* deve fazer seu rastreamento, dentro da tipologia Vídeos do site.

A página de exibição dos vídeos deve mostrar o *player* como parte principal da interface, seu título e os demais vídeos listados abaixo.

4.2 APRENSENTAÇÃO DO CRAWLER

O *crawler* possui uma apresentação bem simples porém extramente útil e necessária. Ele apresenta a arquitetura do *crawler* na tela do navegador, onde é executado, e também mostra mensagens para guiar o desenvolvedor sobre seu funcionamento.

Na Figura 12 é mostrada a tela após a execução do crawler.

Como primeiro conteúdo da página, está a figura que representa a arquitetura do *crawler* desenvolvido neste trabalho. Abaixo dela é a impressão do contador de rea que a ferramenta faz. Em verde, a Figura 12, está mostrando a mensagem de confirmação sobre a inserção de um vídeo no banco de dados.

Caso tenha algum erro no código, aparece uma mensagem, especificando o erro e a linha no código em que ele está, é possível ver um exemplo de erro na Figura 13.

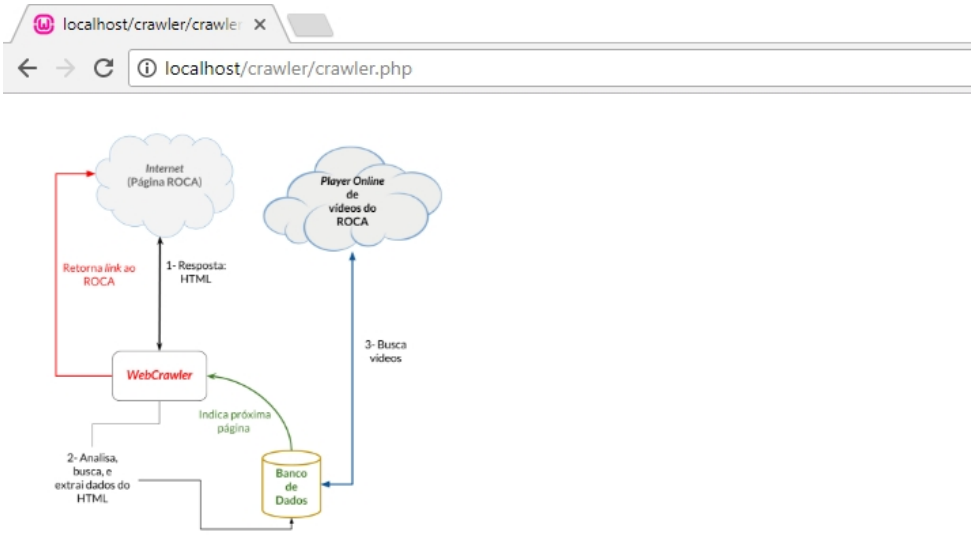


Figura 12 – Crawler sendo executado.

Fonte: Autoria própria.

Figura 13 – Printscreen da tela de erro do Crawler.

Fonte: Autoria própria.

4.3 APRENSENTAÇÃO DO PLAYER

O player foi desenvolvido para ser utilizado de maneira simples. É posto em um site com página única, desenvolvido pelo autor, onde se pode reproduzir o vídeo, encontrar os demais vídeos do ROCA e ter algumas informações sobre o autor (Figura 14).

Figura 14 – Tela do player de vídeo.

Fonte: Autoria própria.

Na parte superior da página, está situado o menu principal, composto por links que fala Sobre o trabalho, os apoios que o trabalho recebeu e o contato do autor do trabalho. Sua funcionalidade principal é focada na exibição dos vídeos. Por este motivo, o *player* está localizado logo abaixo do menu pricipal, com um tamanho que pega 100% da largura da página. No canto superior esquerdo é exibido o nome do site escolhido pelo autor. O *player*, como já explicado anteriormente, é chamado pela tag <video>, ou seja, é o player padrão do navegador, e varia de interface de acordo com o *web browser* de preferência, ou seja, se abrir no *Firefox* será um padrão, se abrir no *safari* será outro, se abrir no *Microsoftr Edge* será outro, se abrir no *Google Chrome* será outro e assim sucessivamente.

Abaixo do *player* pode ser encontrada uma grade com várias imagens em miniatura com um título cada. Estas miniaturas representam os demais vídeos disponíveis para serem assistidos. Ao clicar na miniatura (que é composto por uma imagem com desenho do ROCA e o título do vídeo) o vídeo específico (através do seu identificador) será carregado e exibido como vídeo principal na tela acima. A Figura 15 mostra essa parte.

O *Player* também pode ser visualizado em dispositivos *mobiles* por se tratar de uma ferramenta *web*. A Figura 16 mostra como fica a página em um *smartphone* comum.

Figura 15 – Miniaturas do demais vídeos.

Fonte: Autoria própria.

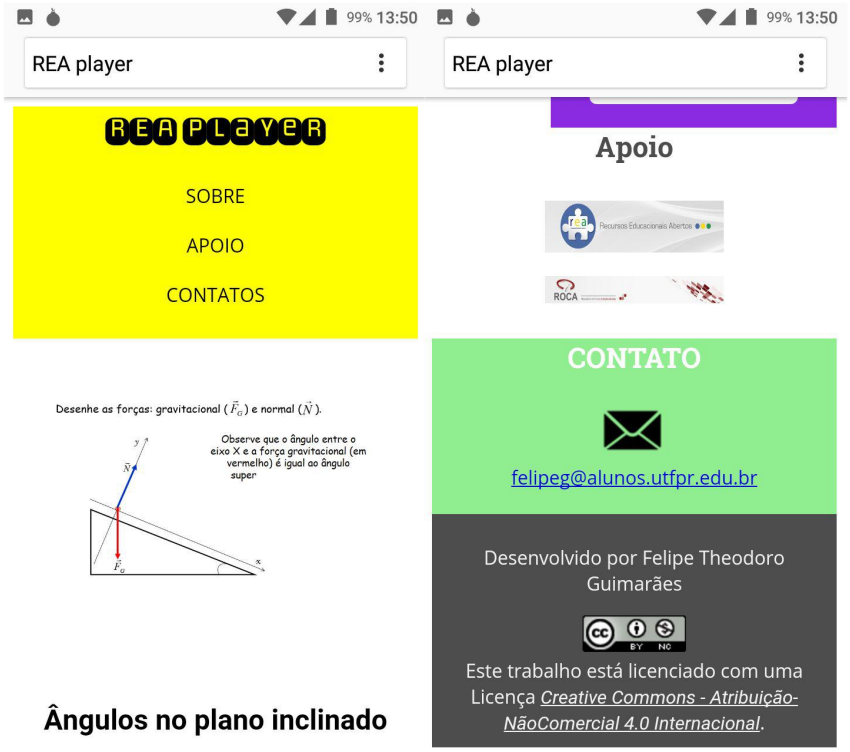


Figura 16 – Player visto no smartphone.

Fonte: Autoria própria.

4.4 DISCUSSÃO

O trabalho apresenta uma ferramenta que possibilita visualizar os vídeos do ROCA (que até então prcisavam ser baixados para serem visualizados). Com o presente trabalho,

a facilidade para quem quer ter acesso aos objetos de aprendizagem fica mais explícita.

O trabalho poderá ser utilizado, modificado, melhorado e distribuído pois ele possui uma licença livre para isso com seu código fonte aberto. Isso permite a possibilidade de compartilhamento de conhecimento para que a ferramenta seja ampliada, não se limitando apenas aos autores deste trabalho. O conhecimento colaborativo é de extrema importância já que trabalha-se em cima de um projeto aberto e livre como o REA. Os códigos podem ser acessados via *github* aqui neste *link* `<https://github.com/REAWebCrawlerPlayer>`

A melhoria das ferramentas propostas deve ser feita, como detalhes de *download* não apenas de vídeos mas de todos os reas que estão armazenadas juntamente com os vídeos, como documentos e imagens. Tamém deve-se implementar uma fucionalidade pra baixar todos os vídeos pois até agora apenas um vídeo está sendo indexado por página, pois sabe-se que existem páginas com dois ou mais vídeos. Também as imagens para as miniaturas estão estáticas e padrões, mas seria mais interessante que um *frame* do vídeo seja capturado e esta imagem serviria como miniatura do vídeo em questão, porém, esta funcionalidade não é trivial de se implementar, mas com a ajuda colaborativa, isso pode se tornar possível.

5 CONCLUSÃO

O aprendizado da geração atual esta cada vez mais ligado às novas tecnologias e o papel do professor se transforma juntamente com a mudança da sociedade atual.

O objetivo geral do trabalho consistia em facilitar o acesso dos usuários finais aos vídeos do repositório ROCA, e para isso também desenvolver um *WebCrawler* que fosse capaz de extrair e armazenar as URL's dos vídeos. A forma de facilitar o acesso aos vídeos, seria disponibilizar todos eles, em um *site*, *online*, sem necessidade de baixar cada um deles para assistir.

Para realizar tal trabalho não foi difícil achar boas ferramentas e que fossem gratuitas. Todas as ferramentas utilizadas são sugeridas pelos professores no curso de Licenciatura em Informática, e também os alunos aprendem, em sala de aula, o modo de usá-las.

Os objetivos do trabalho foram concluídos, uma vez que ele tornou possível assistir aos vídeos *online*, e achá-los rapidamente na mesma página em que esta sendo executado o vídeo. Também foi concluído o desenvolvimento de um *WebCrawler* para extrair as URL's do repositório ROCA e armazenar todas em um banco de dados.

A qualidade do *site* não é alta, e também não é o mais atrativo para a geração atual, pois sabe-se que para ter um *design* de qualidade voltado a um público específico é necessário mais estudo e conhecimento na área de desenvolvimento *web*, pois não é apenas nisso que o curso de Licenciatura em Informática esta focando, mas também existe foco nos processos de ensino e aprendizagem que o futuro licenciado deve dominar para atuar profissionalmente.

Há intenção de que este trabalho seja continuado, pelo autor ou não, por isso a aplicação da licença *CREATIVE COMMONS*, para livre incorporação da ideia e utilização dos Recursos Educacionais Abertos.

5.1 TRABALHOS FUTUROS

Sugere-se que o *player* seja melhorado, não se utilizando mais o *player* padrão de cada navegador chamado através da *tag* <video>, mas que se utilize um *player* em html5 com *javascript* com possibilidade de carregar legendas e de controle melhorado, tal como o *VideoJS*[1]; Sugere-se também a incrementação de mais páginas de Recursos Educacionais abertos e sua igual indexação para o banco de dados e consequente visualização pra o *player*. Também pode ser feito um motor de busca no banco de dados através de palavras-chave

[1] <http://videojs.com/>

ou título do trablaho. Por enquanto há dezesseis REAS incluídos no banco, mas no futuro quando esse número for superior, esta busca se fará necessária. Também é necessário resolver o problema de acentuação que o *crawler* no momento que indexa não consegue armazenar caractere[2] especial.

[2] Sinal ou símbolo usado na escrita, em especial no domínio da informática, in Dicionário Priberam da Língua Portuguesa [em linha], 2008-2013, https://www.priberam.pt/dlpo/caractere [consultado em 23-10-2017].

REFERÊNCIAS

AGUIAR, E. V. B.; FLÔRES, M. L. P. Objectos de Aprendizagem: conceitos básicos. *TAROUCO, Liane Margarida Rockenbach; COSTA, Valéria Machado da; ÁVILA, Barbara Gorziza et al. Objetos de aprendizagem: teoria e prática. Porto Alegre: Evangraf*, 2014. ISSN 1098-6596. Citado na página 15.

BAKER, J. Oer introduction. *The Connexiona Project*, v. 1, n. 1, p. 1–5, 2007. Citado na página 18.

DATE, C. J. *Introdução a sistemas de bancos de dados*. [S.l.]: Elsevier Brasil, 2004. Citado na página 21.

MILL, D. Das inovações tecnológicas às inovações pedagógicas: considerações sobre o uso de tecnologias na educação a distância. *Educação a distância: desafios contemporâneos. São Carlos: Edufscar*, p. 43–57, 2010. Citado na página 12.

OCDE. *Organización para la Cooperación y el Desarrollo Económicos*. 2015. Citado na página 16.

PANTO, E.; COMAS-QUINN, A. The challenge of open education. *Journal of E-Learning and Knowledge Society*, 2013. ISSN 18266223. Citado 2 vezes nas páginas 13 e 16.

PINHEIRO, D. S. *Potencialidades dos Recursos Educacionais Abertos para a educação formal em tempos de cibercultura*. Dissertação (Mestrado) — Universidade Federal da Bahia. Faculdade de Educação, Salvador, 2014. Citado na página 13.

PINKERTON, B. *Webcrawler: Finding what people want*. [S.l.]: University of Washington Seattle, WA, 2000. Citado 2 vezes nas páginas 18 e 19.

SANTOS, I. C. e. a. ANÁLISE DE CRESCIMENTO DO USO DE PLAYERS DE VÍDEO EM UMA STARTUP DE EDUCAÇÃO. *Revista Interdisciplinar do Pensamento Científico*, 2016. ISSN 2446-6778. Citado 2 vezes nas páginas 12 e 13.

WILEY, D. A. Learning Object Design and Sequencing Theory. *Learning Object Design and Sequencing Theory*, 2000. ISSN 1098-6596. Citado 2 vezes nas páginas 15 e 16.