



REVISTA INTERDISCIPLINAR ENCONTRO DAS CIÊNCIAS  
V.1. N.1. 2018

## SISTEMA RASTREADOR E CENTRALIZADOR DE NOTÍCIAS DOS PORTAIS CREDE'S DA SECRETARIA DE EDUCAÇÃO DO CEARÁ

SYSTEM TRACKER AND NEWS CENTER OF CREDE'S PORTALS OF THE EDUCATION  
DEPARTMENT OF CEARÁ

José Ailton Batista da Silva<sup>1</sup> | Rangel Henrique Félix<sup>2</sup> | Suyanne Hellen Holanda Pinheiro<sup>3</sup> | Tulio Vidal Rolim<sup>4</sup> | Pedro Luís Saraiva Barbosa<sup>5</sup>

### RESUMO

A internet por ser um veículo massivo de difusão de notícias, é amplamente usado por universidades, empresas e setores públicos através do uso de portais e redes sociais. Esse modelo de publicação de notícias também é usado pelas Coordenadorias Regionais de Educação do Ceará (CREDE's), as quais estão distribuídas em vinte cidades diferentes por todo o estado. Porém, ainda assim o usuário encontra problemas na obtenção rápida de notícias por se tratar de diversos portais separados e não saber quando houveram publicações/modificações em todos esses portais. O objetivo deste trabalho é apresentar uma ferramenta web responsável por fazer o rastreamento das últimas notícias de todos esses portais governamentais/educacionais do Estado do Ceará e centralizá-las em uma única página, além de notificar usuários cadastrados usando um serviço de mensagens instantâneas.

### PALAVRAS-CHAVE

Bots. CREDE. Notícias. Rastreador.

### ABSTRACT

The internet as a massive vehicle to broadcast news it is wide used by universities, companies and public sectors by the use of portals and social nets. This model of publication of news is used by the Coordenadorias Regionais de Educação do Ceará (CREDE's) which are distributed for twenty cities in all State. However even so the user finds problem to obtain fastest news because there are many portals separated and don't know when has publications in all those portals. In this context the aim of this paper is to introduce a web tool responsible for crawling of last news of all those portals and show them in a unique web page and also notify logged people using a chat service.

### KEYWORDS

Bots. CREDE. News. Tracker.

### INTRODUÇÃO

As tecnologias da informação estão sendo bastante usadas pelo governo através do chamado Eletronic Government (e-gov), muitos dos processos e atividades intrínsecas estão sendo virtualizados, além da comunicação com o público externo (PINHO, 2008). O conceito de e-gov pode ter vários sentidos, porém é comum vinculá-lo com o uso da internet provendo serviços

disponíveis aos cidadãos (PRADO; RIBEIRO; DINIZ, 2012). A construção de portais governamentais é uma das formas usadas nesse processo de informatização pelo qual os governos oferecem serviços e mostram suas realizações (PINHO, 2008; TRAUNMULLER; WIMMER, 2001).

A Secretaria de Educação do Ceará (SEDUC), nesse contexto, também disponibiliza para a população um portal por meio do qual faz sua publicidade acerca dos seus projetos educacionais. Porém, assim como nos outros estados, a SEDUC possui instâncias intermediárias entre as escolas. Na capital estão as três Superintendências Regionais das Escolas de Fortaleza (SEFOR), já no interior são as vinte Coordenadorias Regionais de Desenvolvimento da Educação (CREDE's) distribuídas por alguns municípios. Entre as atribuições das CREDE's estão o acompanhamento e monitoramento das políticas de educação nas escolas, além do suporte técnico, pedagógico e administrativo tal como contratação de professores temporários (VIDAL; VIEIRA, 2016; FREITAS, 2017). A Tabela 1 mostra a localização das CREDE's, e a quantidade de escolas que estão vinculadas, destas destacamos as escolas profissionalizantes (EEEP) e as de ensino Regular.

**Tabela 1:** Distribuição das 20 CREDE's nos respectivos municípios

CREDE	MUNICÍPIO	EEEP	ESCOLA REGULAR	TOTAL ESCOLA
1	Maracanã	11	50	75
2	Itapipoca	9	31	42
3	Acará	3	17	30
4	Camocim	3	10	15
5	Tianguá	6	27	36
6	Sobral	8	38	47
7	Canindé	2	11	15
8	Baturité	2	18	22
9	Horizonte	4	10	15
10	Russas	5	21	27
11	Jaguaripe	2	11	14
12	Quixadá	3	16	20
13	Crateús	6	21	37
14	Senador Pompeu	3	10	14
15	Tuaú	2	9	13
16	Iguatu	4	13	18
17	Icó	4	11	16
18	Crato	6	22	29
19	Juazeiro	5	20	27
20	Brejo Santo	4	18	24

**Fonte:** Adaptado de Freitas (2017).

Cada uma dessas instâncias (CREDE's) possui seu próprio portal, o qual realiza divulgação dos processos das escolas que estão vinculadas, além dos próprios projetos, parcerias com os

municípios e seleções públicas. Porém, essa estrutura descentralizada dos portais gera alguns problemas na obtenção de informações rápidas para os usuários, além disso o portal da SEDUC não dispõe de uma área onde possam encontrar informações de todas as instâncias, para isso, é necessário estar navegando por diversos portais para obter as notícias de todas as CREDE's. Essa tarefa torna-se difícil, pois há necessidade de ficar anotando quais sites já foram navegados, além disso, gera muito desperdício de tempo para os usuários que possuem interesse por essas informações.

A partir dessa problemática, o presente trabalho tem por objetivo apresentar uma possível solução utilizando um sistema de mineração de páginas HTML (web crawler) capaz de obter as notícias de todos esses portais de forma rápida e exibi-las em uma única página.

## **MATERIAIS E MÉTODOS**

Para a criação desse sistema primeiramente foram feitas algumas análises nos portais das CREDE's a fim de encontrar padrões que pudessem melhorar a construção do crawler. No entanto, de todos os vinte portais, foram escolhidos sete aleatoriamente, verificando o conteúdo visível para o público. Analisando o código HTML foi possível detectar que os portais usam o Content Manager System (CMS) Joomla que permite um alto grau padronização dos sites e facilidade de publicação de conteúdos (JOOMLA, 2017).

Segundo Coelho (2011), as principais características do Joomla é sua capacidade de personalização sem a necessidade do uso de programação e que além de ser completo em termos de funcionalidades constitui-se uma solução viável na construção de portais educacionais.

Por usar um CMS todos os portais analisados estão com o layout praticamente idênticos, o que facilita para os usuários, pois conseguem usar os conhecimentos de navegação adquiridos em um dos portais em todos os outros, melhorando na sua curva de aprendizagem. Porém algumas divergências foram encontradas, como por exemplo em alguns dos portais não há uma página específica para as notícias mais antigas estando algumas agrupadas no menu principal ou mesmo não existindo.

Além do problema citado, acima as páginas com listas de notícias só contêm o título e link, organizado por ordem cronológica, mas sem campo especificando data publicada necessitando-se acessar o conteúdo da notícia para obter informações da data de publicação. Optou-se então por fazer o rastreamento da seção de últimas notícias das páginas iniciais por estarem mais padronizadas e presentes em todos os portais analisados, conforme exibido na Figura 1. Essas seções em geral só apresentam de dez a quinze itens, mas é o suficiente para se manter atualizado

no portal além de estar disponível a data de publicação da notícia melhorando a performance do algoritmo de rastreamento.

**Figura 1:** Seção de últimas notícias dos Portais CREDE's.



Fonte: SEDUC (2017).

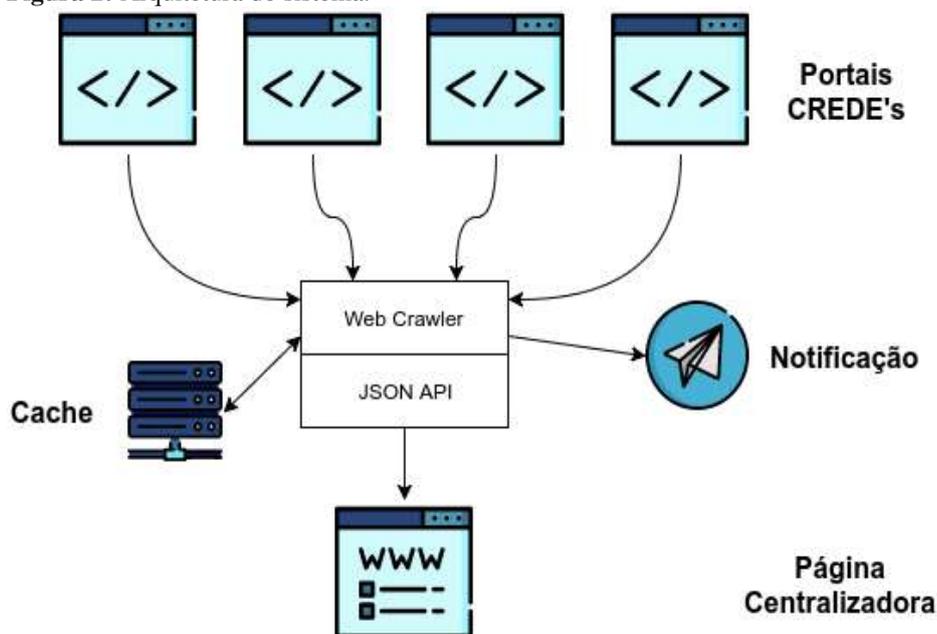
O próximo passo foi definir a arquitetura do sistema e quais tecnologias seriam usadas, visando que a versão inicial do sistema deveria disponibilizar na internet uma página acessível a qualquer momento. Essa página deveria ser hospedada de forma barata para garantir que o serviço seja executado gratuitamente. Baseando-se nesse requisito e outros citados mais adiante a linguagem escolhida para o desenvolvimento foi a PHP por apresentar diversos aspectos de mineração de páginas disponíveis nativamente na linguagem, além de ter diversos frameworks profissionais no mercado como por exemplo: Symfony e Zend (PHP, 2017).

A biblioteca Goutte foi usada por já fornecer diversas funções de rastreamento prontas para extração de dados e respostas HTML/XML, além de usar componentes do framework Symfony facilitando a tarefa de mineração de informações (POTENCIER, 2017). O framework Symfony possui várias classes capazes de fazer uma aplicação web de maneira rápida, por seus componentes serem desacoplados e reusáveis, dois desses componentes foram cruciais para desenvolver tal aplicação (BELÉM et al., 2014).

A Figura 2 apresenta a arquitetura do sistema, o rastreador em PHP, responsável por analisar todas as páginas iniciais dos vinte portais CREDE's. Na varredura cada página é passada por um filtro para capturar somente a seção de últimas notícias. Após conseguir o conteúdo da seção é passado um novo filtro para capturar cada item obtendo o título da notícia, o link e a data de publicação onde são armazenados em forma de uma estrutura de dados na memória do servidor. Essas estruturas são unidas e guardadas em um arquivo de cache para evitar fazer o procedimento novamente, melhorando a performance para o usuário, que a cada duas horas tem o cache descartado e um novo rastreamento será feito.

A página centralizadora das notícias foi desenvolvida usando tecnologia AJAX e faz requisições a uma API previamente desenvolvida para fornecer dados no formato JSON (MCFARLAND, 2011). Foi utilizado o framework Bootstrap do Twitter para que a página ficasse responsiva e disponível a diversos tamanho de telas (SPURLOCK, 2013). Por último, cada vez que o sistema captura as notícias e gera a estrutura de dados é feito uma comparação com o cache anteriormente guardado, resultando nas ultimas notícias lançadas nos portais, essas novas notícias são então enviadas para um serviço de bate-papo.

**Figura 2:** Arquitetura do sistema.



**Fonte:** Elaborado pelos autores (2017).

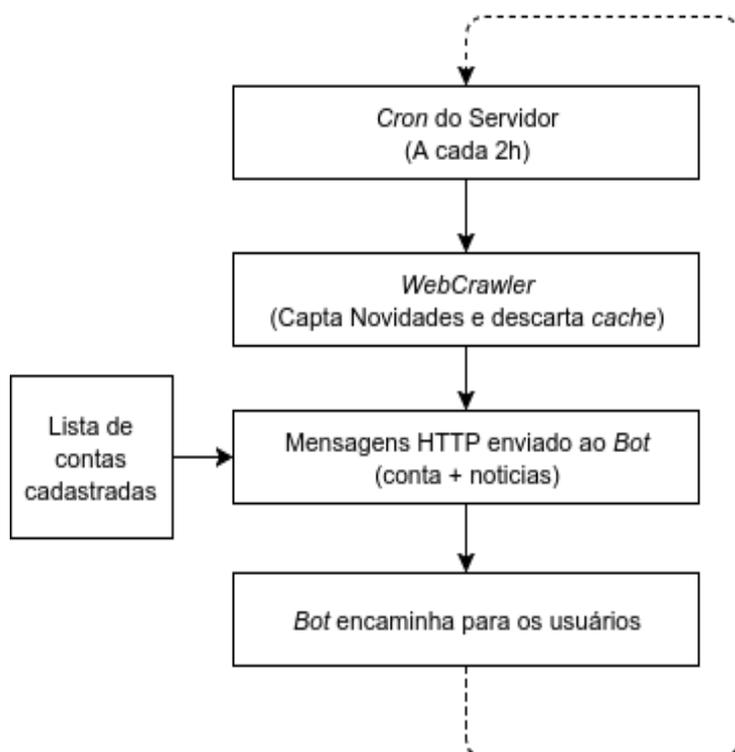
O sistema faz uso de um aplicativo de bate-papo para notificar usuários cadastrados sobre novas notícias. O Telegram é uma aplicação open source, caracterizado por seu código ser aberto para todos, assim como sua API (TELEGRAM, 2017). O aplicativo Telegram possui uma API gratuita para a criação de bots, funciona em dispositivos Android, iPhone e Windows Phone, além de fornecer notificação push alertando o usuário a novas mensagens (MORAIS et al., 2017). Para fazer as notificações foi criado um bot conforme Figura 3. Bot's são simplesmente contas Telegram operadas por software e muitas vezes eles têm características de AI (Inteligência Artificial), além disso, são contas especiais que não precisam de um número adicional de telefone para funcionar. Essas contas prover uma interface para código executando em algum lugar de seus servidores. Para usar não é preciso conhecimento sobre o protocolo MTPProto, pois os seus servidores dispõem uma interface HTTPS simples oferecida através de sua API (TELEGRAM, 2017).

Os Bot's podem fazer diversas ações como ensinar, jogar, procurar, transmitir, lembrar, conectar e integrar-se a outros serviços ou mesmo passar comandos para a internet das coisas

(TELEGRAM, 2017). Neste caso, o bot está integrado com o sistema facilitando a notificação de cada nova atualização.

As notificações usando WhatsApp foi descartada nessa primeira versão do software por este não possuir API oficial além de ter a política de proibir qualquer serviço não caracterizado como sendo um ser humano e apesar de existir bibliotecas de terceiros capaz de fornecer esse recurso, muitas delas estão depreciadas pelas constantes mudanças na aplicação do WhatsApp (TRAVASSO SARINHO, 2017).

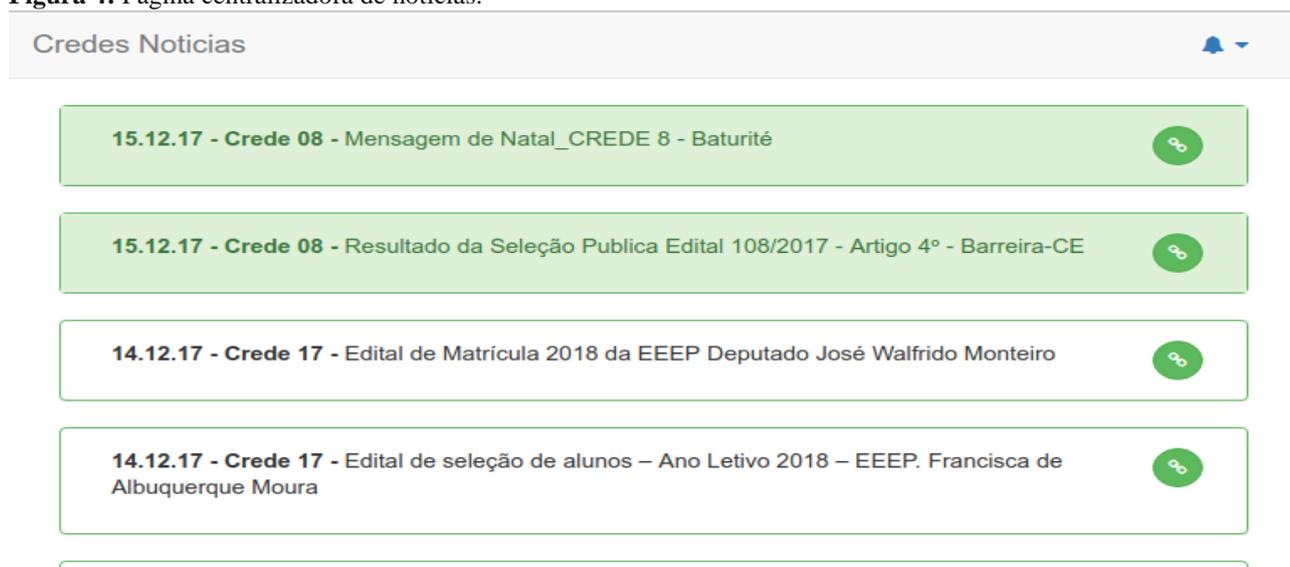
**Figura 3:** Notificação usando *Bot* do Telegram.



**Fonte:** Elaborado pelos autores (2017).

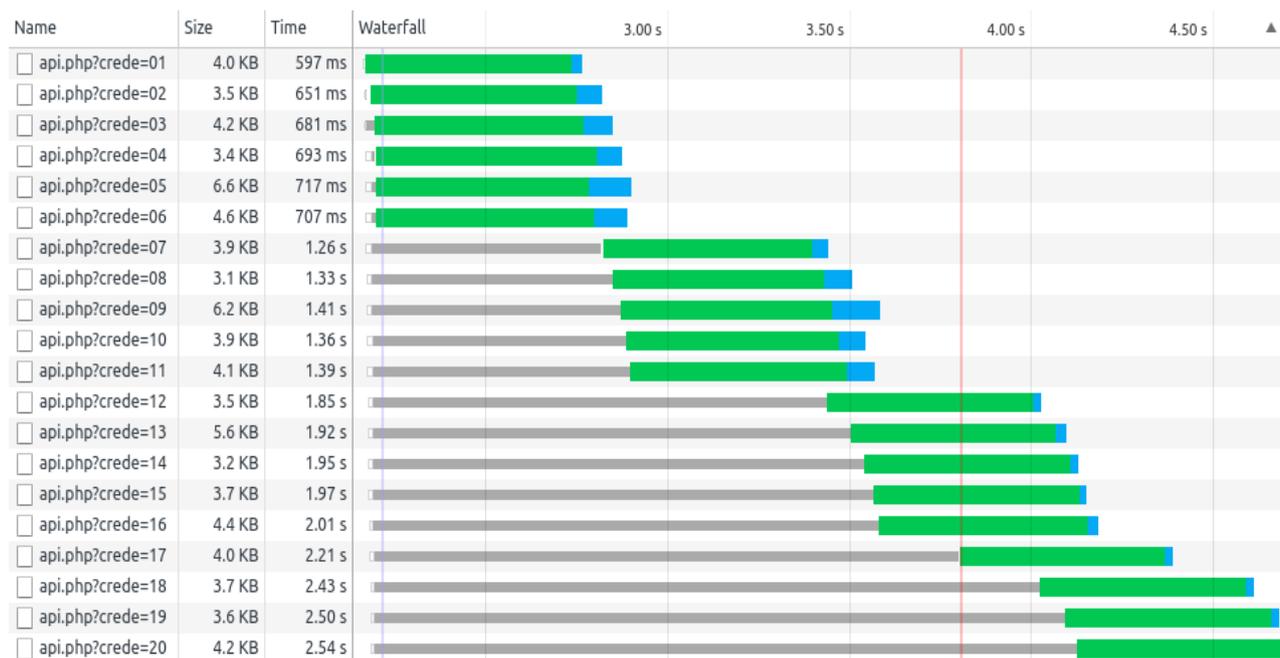
## RESULTADOS E DISCUSSÃO

Para a validação do projeto pode-se observar na Figura 4 uma captura de tela do sistema Crides Notícias, o mesmo organiza os itens por ordem cronológica, cada item contém a data de publicação, o portal de origem, título da notícia e um link que redireciona para a publicação original. Para melhor visualização, as postagens do dia ficam em destaque com fundo esverdeado. Além disso foi implementada na página uma opção de filtro persistente, permitindo que o usuário receba notícias somente das CREDE's que ele escolher, esse filtro fica persistido usando a tecnologia do HTML5 Web Storage não necessitando estar filtrando cada vez que acessar a aplicação (MOZILLA, 2017).

**Figura 4:** Página centralizadora de notícias.

**Fonte:** Elaborado pelos autores (2017).

Para melhor performance do sistema, a JSON API permite fazer requisições individuais das notícias por CREDE garantido que filtragens escolhidas com poucos portais não precisem receber todo o conteúdo do cache. Para facilitar o uso do AJAX na página centralizadora optou-se por usar a biblioteca jQuery (SILVA, 2008; MCFARLAND, 2011), sendo possível fazer múltiplas chamadas ao servidor quase que em paralelo. Conforme exibido na Figura 5 podemos observar que foram feitas vinte requisições HTTP ao servidor que obteve respostas agrupadas em média de cinco requisições por grupo, apesar das últimas aparecerem um pouco defasadas.

**Figura 5.** Múltiplas requisições AJAX usando jQuery.

**Fonte:** Elaborado pelos autores (2017).

O rastreador em PHP envia as novas notícias usando a API fornecida pelo Telegram que automaticamente encaminha para os celulares cadastrados. Usando essa opção não há necessidade que acessar a página centralizadora, porém caso o usuário apague a conversa com bot terá que acessá-la pois não há re-envio de mensagens. Até o presente momento da edição deste artigo as notificações via bot só estão sendo enviadas para três números de celulares o que faz a aplicação ter uma boa performance, porém testes futuros precisam ser feitos para garantir que o sistema seja escalável caso esse número de celulares aumente. A Figura 6 apresenta uma notificação do CredeNoticiasBot, que foi o Bot desenvolvido para automatizar o processo de notificações para o usuário. Nesse exemplo podemos ver uma notificação de um edital de seleção de alunos em uma EEEP da CREDE17. A notificação é feita enviando duas mensagens, a primeira é o título conforme retirado do portal de origem com a data da publicação, na segunda é enviado um link para acesso da publicação original, o Telegram automaticamente encapsula o link deixando-o em destaque e exibindo o domínio.

**Figura 6:** Tela da conversa com *Bot* notificador do Telegram.



**Fonte:** Elaborada pelos autores (2017).

## CONSIDERAÇÕES FINAIS

Este artigo apresentou o desenvolvimento de um sistema que permitisse obter as notícias de todos os portais CREDE's tanto via navegador, quanto usando o aplicativo de chat Telegram. O sistema usou as linguagens base da web em conjunto com o PHP para fornecer um serviço gratuito. Tal aplicação apesar de estar na sua versão inicial já funciona da maneira esperada permitindo o melhor acompanhamento dos informes publicados nos portais citados.

Para trabalhos futuros esperamos melhorar a eficiência da aplicação e fazer o rastreamento de portais educacionais de outros estados vizinhos, bem como fazer um filtro de notícias personalizadas para determinado usuário. Esperamos também fazer com que o usuário possa interagir com o bot pedindo informações sobre as CREDE's e não só servir como interface de recebimentos das mensagens o qual está atualmente.

## REFERÊNCIAS

BELÉM, Patrick Helder Alvarenga et al. Escolha de um Framework para a linguagem de programação PHP através do método AHP clássico. In: **Anais X Congresso Nacional de Excelência em Gestão**. 2014.

COELHO, Marcos Antônio Pereira et al. O uso do CMS Joomla e suas ferramentas hipertextuais na produção de sites educativos e de material didático online. **Texto Livre: Linguagem e Tecnologia**, v. 4, n. 2, p. 38-46, 2011.

FREITAS, Maria Marlene Vieira et al. A superintendência escolar na rede pública estadual de ensino do Ceará: perspectivas e desafios. 2017.

JOOMLA. **About Joomla!**. Disponível em: <<https://www.joomla.org/about-joomla.html>>. Acesso em: 12 dez. 2017.

MCFARLAND, David Sawyer. **JavaScript & jQuery: the missing manual**. " O'Reilly Media, Inc.", 2011.

MORAIS, Ceres GB et al. Donuts: um bot como instrumento facilitador do processo de ensino-aprendizagem na disciplina "Construção de Algoritmos". **Revista Eletrônica Argentina-Brasil de Tecnologias da Informação e da Comunicação**, v. 1, n. 7, 2017.

MOZILLA. **Web Storage API**. Disponível em: <[https://developer.mozilla.org/en-US/docs/Web/API/Web\\_Storage\\_API](https://developer.mozilla.org/en-US/docs/Web/API/Web_Storage_API)>. Acesso em: 12 dez. 2017.

PHP. **Manual do PHP**. Disponível em: <[http://php.net/manual/pt\\_BR/](http://php.net/manual/pt_BR/)>. Acesso em: 12 dez. 2017.

PINHO, José Antonio Gomes de. Investigando portais de governo eletrônico de estados no Brasil: muita tecnologia, pouca democracia. 2008.

POTENCIER, F. **FriendsOfPHP/Goutte**. Disponível em: <<https://github.com/FriendsOfPHP/Goutte>>. Acesso em: 12 dez. 2017.

PRADO, Otávio; RIBEIRO, Manuella Maia; DINIZ, Eduardo. Governo eletrônico e transparência: olhar crítico sobre os portais do governo federal brasileiro. **Estado, sociedade e interações digitais: expectativas democráticas**, p. 13-39, 2012.

SEDUC. **Governo do Estado do Ceará**. Disponível em: <<http://www.seduc.ce.gov.br/>>. Acesso em: 12 dez. 2017.

SILVA, Maurício Samy. **jQuery: a biblioteca do programador JavaScript**. São Paulo: Novatec Editora, 2008.

SPURLOCK, Jake. **Bootstrap: Responsive Web development**. " O'Reilly Media, Inc.", 2013.

TELEGRAM. **Telegram APIs**. Disponível em: <<https://core.telegram.org/api>>. Acesso em: 12 dez. 2017.

TRAVASSOS SARINHO, Victor. LibrasZap-Um Jogo Baseado em Mensagens Instantâneas para Avaliação de Conhecimentos na Língua Brasileira de Sinais. **Revista Brasileira de Informática na Educação**, v. 25, n. 1, 2017.

TRAUNMULLER, Roland; WIMMER, Maria. Directions in e-Government: Processes, portals, knowledge. In: **Database and Expert Systems Applications, 2001. Proceedings. 12th International Workshop on**. IEEE, 2001. p. 313-317.

VIDAL, Eloísa Maia; VIEIRA, Sofia Lerche. **Políticas de Ensino Médio no Ceará**. 2016.

Recebido em: 16 de Dezembro de 2017

Aceito em: 23 de Dezembro de 2017

<sup>1</sup> Graduado em Tecnologia em Mecatrônica Industrial pelo Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE). Tem experiência na área de programação em C/C++ e Linguagens para Web. Campeão da Competição de Robótica Móvel The Freescale Cup Brazil (2014). Participante (6º lugar) da Competição de Robótica Móvel: The Freescale Cup WorldWide 2015 na Alemanha. Atualmente trabalha como Suporte Operacional em Hardware, Software e Administração de Redes Linux na Coordenadoria Regional de Desenvolvimento da Educação (CREDE 17). E-mail: ailton.ifce@gmail.com

<sup>2</sup> Graduado em Redes de Computadores e Técnico em Informática e em Eletrotécnica (Ênfase em Sistemas Elétricos Industriais) pelo Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE). Atua na área de Suporte Operacional em Hardware, Software e Administração de Redes Linux na Coordenadoria Regional de Desenvolvimento da Educação da 11ª Região do Ceará (CREDE 11). Seu interesse de pesquisa está relacionado a Redes de Computadores, Softwares Livres e Tecnologias de Informação e Comunicação. E-mail: rangel.crede11@gmail.com

<sup>3</sup> Graduada em Licenciatura em Química pelo Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE). Bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES. E-mail: suyanne.hellen@gmail.com

<sup>4</sup> Mestrando em Ciência da Computação pela Universidade Federal do Ceará (UFC). Docente no curso de Análise e Desenvolvimento de Sistemas pela Faculdade Vale do Salgado (FVS). Coordenador do Núcleo de Extensão em Análise e Desenvolvimento de Sistemas (NEXAS). Revisor da Revista IEEE Latin América (IEEEEA). Atualmente pesquisa nas seguintes linhas: Banco de Dados Ligados, Web Semântica e Bancos de Dados Móveis. E-mail: tuliovidal@fvs.edu.br

<sup>5</sup> Mestre em Engenharia de Software pelo Centro de Estudos Avançados do Recife (CESAR). Professor EBTT Efetivo do Instituto Federal de Educação, Ciência e Tecnologia do Ceará - Campus Cedro (IFCE). Tem experiência na área de Engenharia de Software atuando principalmente nos seguintes temas: Engenharia de Software e Educação e Tecnologia. E-mail: pedro.barbosa@ifce.edu.br