

# API ACCESSIBILITY EXTRACTION\*

API para Extração de Parâmetros em Objetos de Aprendizagem Digital para Validação de Acessibilidade

Juliana de Fatima da Silva  
Departamento de Tecnologia da  
Informação  
Universidade Federal de Santa  
Maria (UFSM)  
Frederico Westphalen, RS, Brasil  
julianafatsil@gmail.com

Maicon André Dall' Asta  
Departamento de Engenharias e  
Ciência da Computação  
Universidade Regional Integrada  
do Alto Uruguai e das Missões  
(URI)  
Frederico Westphalen, RS, Brasil  
dallas.maicon@gmail.com

Fábio J. Parreira  
Departamento de Tecnologia da  
Informação  
Universidade Federal de Santa  
Maria (UFSM)  
Frederico Westphalen, RS, Brasil  
fabiojparreiral@gmail.com

Sidnei Renato Silveira  
Departamento de Tecnologia da  
Informação  
Universidade Federal de Santa  
Maria (UFSM)  
Frederico Westphalen, RS, Brasil  
sidneirenato.silveira@gmail.com

Renan Binda  
Laboratório de Mídias e Inclusão  
Social (EGC/LAMID)  
Universidade Federal de Santa  
Catarina (UFSC)  
Florianópolis, SC, Brasil  
renanbinda1@gmail.com

Vânia Ribas Ulbricht  
Laboratório de Mídias e Inclusão  
Social (EGC/LAMID)  
Universidade Federal de Santa  
Catarina (UFSC)  
Florianópolis, SC, Brasil  
vrulbricht@gmail.com

## ABSTRACT

This article presents the development of an API in *NodeJs* to automatically extract parameters in *EPUB*, *DOCX* and *PDF* digital learning objects. These parameters are used to check the degree of accessibility of Digital Learning Objects for people with hearing and visual impairments. The API uses refactoring techniques and the use of recursive functions, resulting in a *JSON* object that can be used by many programming languages. The article presents the results of preliminary tests, where it was found that *EPUB* files are better structured, leaving the parameters completely free to be extracted. The validation will be supported by the Development and Innovation group in Rehabilitation and Digital Assistive Technology (DIRTA-UFSM) and the accessibility team at the Social Media and Social Inclusion (LAMID) laboratory at the Federal University of Santa Catarina (UFSC).

## KEYWORDS

*Digital Accessibility, Digital Learning Objects, Parameter Extraction.*

## 1 INTRODUÇÃO

A educação igualitária é um direito garantido por lei, inalienável e universal, considerado como um meio para o desenvolvimento do indivíduo e da própria humanidade, o que torna ponto de partida para outros direitos, proporcionando a base para que a pessoa com deficiência possa ingressar no

mercado de trabalho, obter renda, viver com independência e dignidade [4].

Com o avanço tecnológico a educação on-line se popularizou, facilitando o aprendizado e rompendo barreiras temporais e espaciais, além de disponibilizar recursos educacionais, podendo estes serem utilizados em diversas plataformas, ou em documentos em formato digital [6].

Os Objetos de Aprendizagem (OAs) digital servem tanto para o ensino a distância, quanto ao apoio no ensino presencial [10] trazem informações destinadas à construção do conhecimento, e são estruturados de forma que possam ser reutilizados e combinados com outros OAs [3] tendo como principal objetivo promover a reflexão crítica do estudante sobre o conteúdo estudado, despertar a curiosidade, o questionamento, a inquietação, tendo a forma de apresentação do conteúdo como fundamental [9].

Segundo o Censo do IBGE de 2010, 61,1% das pessoas com algum tipo de deficiência não possuem instrução e ensino fundamental completo. Outras 14,2% têm ensino fundamental completo ou médio incompleto, 17,7% possuem ensino médio completo ou superior completo, apenas 6,7% concluíram o ensino superior e 0,4% proporção denominada “não determinada” [4].

Em se tratando de pessoas com necessidades especiais, a imagem faz parte do registro de memórias dos surdos, quanto mais ilustrativo, mais interesse pode ser despertado e quanto mais manipulado ou tecnológico mais as OAs serão significantes na aprendizagem [1]. Por outro lado, as pessoas

cegas ou com baixa visão, não conseguem acessar as informações contidas nas imagens, somente a descrição na mesma, por meio de leitores de tela.

Vários OAs que, teoricamente, são projetados de acordo com as diretrizes de acessibilidade apresentam incompatibilidade com algumas tecnologias assistivas, dentre elas o VLibras que traduz conteúdos digitais para a Língua Brasileira de Sinais (LIBRAS). Para diagnosticar tal compatibilidade, faz-se necessário analisar de forma criteriosa os parâmetros de acessibilidade dos OAs.

Utilizando do conhecimento das diretrizes de acessibilidade com base em Macedo [6] e Binda [2] e juntamente com o grupo de Acessibilidade das Universidades envolvidas no projeto, este trabalho propõe o desenvolvimento de uma API para extração dos atributos presentes nos documentos OAs digital, com objetivo final de auxiliar os criadores de conteúdos e professores a quantizar a acessibilidade de seus OAs digital, para que estes possam trazer a mesma informação para todos os alunos sem perder a qualidade do ensino e de aprendizagem.

O artigo apresenta de forma sucinta a análise, modelagem, funcionamento, resultado dos primeiros testes da API, e por fim as considerações finais.

## 2 ANÁLISE DA API

A API foi criada com a intenção de auxiliar outros desenvolvedores de software a projetar sistemas que utilizem os parâmetros de acessibilidade previamente extraídos. Usouse de conceitos de código limpo e refatoração apresentadas por Martin [8], neste sentido, define as variáveis e funções com nomes claros e objetivos, com intuito de possibilitar a continuidade, para outros formatos de documentos, disponibilizando o código fonte na plataforma *GitHub* [5].

No desenvolvimento da API foi utilizado *Node.js*, para a criação do motor de execução, a linguagem de programação *JavaScript*, usando funcionalidades apresentadas no *EcmaScript 2015*, o ambiente de desenvolvimento *Visual Studio Code*, e algumas bibliotecas (*cheerio*, *pdf-parser*, *pdfjs-extract* e *rimraf*) para otimização do processo de extração de parâmetros.

Para iniciar o processo de extração de parâmetros realizado pela API, é inserido um documento pelo método *GET* ou *POST*, as extensões *EPUB*, *DOCX* e *PDF* são identificadas automaticamente e caso não seja nenhuma delas é atribuído a extensão outros, retornando um objeto *JSON*, informando que o documento analisado possui a extensão inválida, conforme representado na Fig. 1.

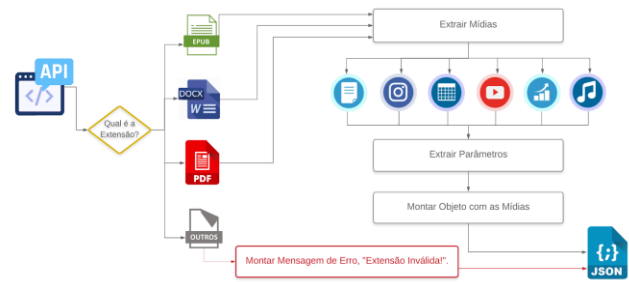


Figura 1. Fluxo do processo realizado pela API

Quando a extensão for válida, a API irá realizar a chamada de extração para o tipo de documento correspondente.

Ao trabalhar com arquivos *DOCX* e *EPUB* faz-se necessário renomeá-los para extensão *ZIP*, pois ao descompactá-los é fornecida toda estrutura do documento, sendo organizada em pastas e arquivos. Para realizar tal atividade foi criada uma função para executar esse processo de forma automatizada e recursiva.

Em documentos *EPUB*, foi utilizado a biblioteca *cheerio*, para efetuar a raspagem do *XHTML*, realizando uma varredura entre as *tags*, e os arquivos de folha de estilo, com a finalidade de carregar o conteúdo e obter os dados pertencentes a estilização.

Para os documentos *DOCX*, foi utilizada a biblioteca *pointer* para acessar o caminho dos atributos contidos nos arquivos *XML*. Em documentos *PDF* utilizou-se as bibliotecas *pdf-parser* e *pdfjs-extract*, que trabalham em conjunto extraindo informações distintas, destas são extraídas apenas conteúdo de texto.

A API trabalha com as mídias texto, imagem, tabela, vídeo, gráfico e áudio. Em cada uma destas mídias são extraídos os atributos indicados por Macedo [6]:

**Textos:** sequência, cor, tamanho, tipo da fonte, texto descritivo, quantidade de caracteres, alinhamento, tipo de texto, cor de fundo do texto; **Imagens:** sequência, título da descrição, descrição e legenda; **Tabelas:** sequência, título da descrição, descrição, legenda e estilo; **Vídeos:** sequência, título da descrição, descrição, link e legenda; **Gráficos:** sequência, título da descrição, descrição e legenda; **Áudios:** sequência, título da descrição e legenda.

Ao término do procedimento é retornado um objeto *JSON*, com todos os atributos separados por mídia, caso ocorra algum erro na extração destes, a aplicação retorna um objeto com erro, que pode ser facilmente interpretado.

## 2.2 Validação

A validação inicial foi realizada utilizando 3 documentos OAs de Teste, disponíveis no *GitHub* [5], sendo um 1 arquivo *DOCX*, 1 *EPUB* e 1 *PDF*, e também foram testados 38 documentos do repositório *MERLOT* [7], sendo 19 OAs *DOCX* e 19 do tipo *EPUB*,

os quais foram escolhidos aleatoriamente juntamente com o grupo de pesquisa de acessibilidade da Universidade Federal de Santa Maria (UFSM).

Ao final verificou-se a porcentagem de acerto na extração dos atributos, demonstradas na tabela 1.

	MERLOT		OAs TESTE		
	DOCX	EPUB	DOCX	EPUB	PDF
Qtd OAs	19	19	1	1	1
<b>MÍDIA</b>					
Imagem	-	-	100%	100%	0%
Tabela	-	-	85%	100%	0%
Texto	92%	100%	100%	100%	66%
Áudio	-	-	100%	100%	0%
Vídeo	-	-	100%	100%	0%
Gráfico	-	-	100%	100%	0%

Tabela 1. Resultados do testes para extração de cada mídia

Nos OAs Teste, o documento EPUB obteve 100% de acerto em todos as mídias. Na extração do arquivo DOCX, verificou-se que o atributo do tipo de tabela é extraído somente em tabela simples, ocasionando um acerto de 85%, e 100% nas demais mídias.

Já o documento PDF obteve 66% de acerto, somente na mídia texto, as demais mídias foram 0%. A API não consegue realizar a extração dos atributos das demais mídias, mesmo nos documentos formatados corretamente, em testes com o leitor de tela JAWS o mesmo não identificou o ALT da imagem. Diferentemente do DOCX e EPUB o PDF não é estruturado por arquivos, definindo-se como uma folha única.

Ao realizar a média geral dos acertos, por formato de documento, constatou-se que o EPUB obteve 100%, DOCX 97,5% e PDF 11%.

Já nos OAs, extraídos do repositório MERLOT [9], do tipo DOCX e EPUB apresentavam somente a mídia de Texto, a API extraiu 92% dos atributos nos OAs DOCX, sendo que este resultado se da ao fato de que em alguns parágrafos a API extraiu fonte inválida, já nos OAs do tipo EPUB, a API extraiu 100% dos atributos.

### 3 CONSIDERAÇÕES FINAIS

Ao analisar os resultados dos três formatos constatou-se que, a princípio, o pior formato para OAs acessível é o PDF, e o melhor formato é o EPUB, esse resultado se deve ao fato do documento ser estruturado em XHTML, formatado em CSS, o que deixa os parâmetros totalmente livres para serem extraídos.

Como sugestão de trabalhos futuros pretende-se adequar a API para que a mesma suporte o padrão SCORM, explorar melhor a extração dos parâmetros em documentos PDF e fazer novas validações com uma gama maior de documentos que contemplem todas as mídias e utilizar técnicas mais eficazes de extração de características, como as técnicas de processamento de imagens e inteligência artificial.

### AGRADECIMENTOS

Este projeto foi realizado em parceria com a equipe de acessibilidade do laboratório de Mídias Sociais e inclusão Social (LAMID) da Universidade Federal de Santa Catarina (UFSC) e do grupo de Desenvolvimento e Inovação em Reabilitação e Tecnologia Assistiva Digital (DIRTA-UFSM).

### REFERÊNCIAS

- [1] T. F. Almeida; B. B. Volpe; A. C. Frasson (2018). Materiais Didáticos Elaborados como Objetos de Aprendizagem: Produtos Educacionais para Estudantes Surdos no Ensino Regular. Revista de Produtos Educacionais e Pesquisas em Ensino. Universidade Estadual do Norte do Paraná Cornélio Procópio, Paraná, v. 2, n. 2, p. 135-148, 2018. Disponível em: <http://bit.ly/AlmeidaVolpeFrasson>. Acesso em: 04 jul. 2019.
- [2] R. P. Binda (2018). Artefato para representação interativa de diretrizes para produção de material educacional acessível. [Dissertação] Renan de Paula Binda; Orientadora, Vânia Ribas Ulbricht. Florianópolis, SC, 2018..
- [3] M. L. F. Carneiro; M. S. Silveira (2014). Objetos de Aprendizagem como elementos facilitadores na Educação a Distância. Educ. rev., Curitiba, n. spe4, p. 235-260, 2014. Disponível em: <http://bit.ly/CarneiroSilveira>. http://dx.doi.org/10.1590/0104-4060.38662. Acesso em: 22 jun. 2019
- [4] CENSO. (2012). Cartilha do Censo 2010: Pessoas com deficiência. Disponível em: <http://bit.ly/Censolbge>. Acesso em: 02 jul. 2019
- [5] GitHub.com. (2020). GITHUB. Disponível em: <github.com/julianaf/ExtracaoDeParametros> Acesso em 13 fev. 2020.
- [6] C. M. S. Macedo (2010). Diretrizes para Criação de Objetos de Aprendizagem Acessíveis. [Tese] Claudia Mara Scudelari de Macedo; Orientadora, Vânia Ribas Ulbricht. Florianópolis, SC, 2010.
- [7] Marlot.org. (2020). MERLOT. Disponível em: <https://www.merlot.org/merlot> Acesso em 05 jan. 2020.
- [8] C. M. Martin (2009). Livro - Código Limpo - Habilidades e Práticas do Agile Software. Alta Books
- [9] J. D. Omodei (2016). Acessibilidade em Objetos de Aprendizagem na EaD: Uma Análise em um Curso de Especialização. Revista Científica em Educação a Distância, v. 6, n. 1, abr. 2016. ISSN 2177-8310, doi: https://doi.org/10.18264/eadf.v6i1.374. Acesso em: 20 jun. 2019.
- [10] K. C. B. Pinto; R. P. Silva; F. G. Teixeira (2019). Requisitos de Projeto de Interfaces Gráficas de Objetos de Aprendizagem Acessíveis para Usuários com Baixa Visão. Revista Educação Gráfica, Brasil, Bauru. ISSN 2179-7374. V. 23, nº 1. Abril de 2019. p. 96 - 115. Disponível em: <http://bit.ly/PintoSilvaTeixeira> Acesso em: 20 jun. 2019