

**PROBOLSAS 2020**  
**EDITAL N' 5/2019/PROEX**

## **RELATÓRIO DE ATIVIDADES**

### **1. Identificação**

**Título do projeto:** MANUTENÇÃO E APRIMORAMENTO DA BIBLIOTECA DIGITAL DE LITERATURA BRASILEIRA

**Nome:** ALLAN SOARES SILVA - matrícula UFSC nº 19200410

**Período:** 01/03/2020 a 31/12/2020

### **2. Descrição das atividades efetivamente realizadas pelos bolsistas.**

O projeto de extensão Manutenção e Aprimoramento da Biblioteca Digital de Literatura Brasileira tem como objetivo aprimorar e desenvolver ferramentas utilizadas nesta biblioteca aumentando as formas de acesso e leitura às obras, implementando ferramentas de acessibilidade e corrigindo possíveis erros. Dada à adequação ao tipo de conteúdo disponibilizado, a biblioteca foi rebatizada recentemente como BPLP (Biblioteca Digital de Literatura de Países Lusófonos).

Tendo em mente principalmente a digitalização de obras ainda não cadastradas e a ampliação das formas de acesso às obras já disponíveis, os dez meses de projeto foram focados na implementação de obras codificadas em XML seguindo o padrão TEI (*Text Encoding Initiative*) como um novo formato de leitura. TEI é um consórcio que estabelece alguns padrões para a representação de obras literárias em formatos digitais.

A disponibilização de obras literárias TEI facilita possíveis correções textuais e permite uma maior portabilidade para diferentes dispositivos, uma vez que é o navegador web que fica responsável pela renderização das obras, adaptando-as de acordo com as especificações do respectivo aparelho.

A partir do projeto, três novas funcionalidades foram implementadas na BLPL: publicação de obras literárias codificadas em TEI; conversão das obras codificadas em TEI para o formato HTML; e conversão de obras codificadas em HTML para sua respectiva codificação em TEI.

#### **2.1 Pesquisa das tecnologias a serem utilizadas no projeto**

O primeiro mês de projeto foi destinado a pesquisa de ferramentas e linguagens que poderiam vir a ser utilizadas, ao estudo das que fossem selecionadas e também ao preparo do repositório do projeto.

Nesse período o bolsista adquiriu os conhecimentos iniciais das linguagens XML e XSL e na biblioteca lxml do python. O bolsista já tinha experiência prévia com as linguagens de programação e marcação Javascript, Python, HTML e CSS que eventualmente foram utilizadas no projeto.

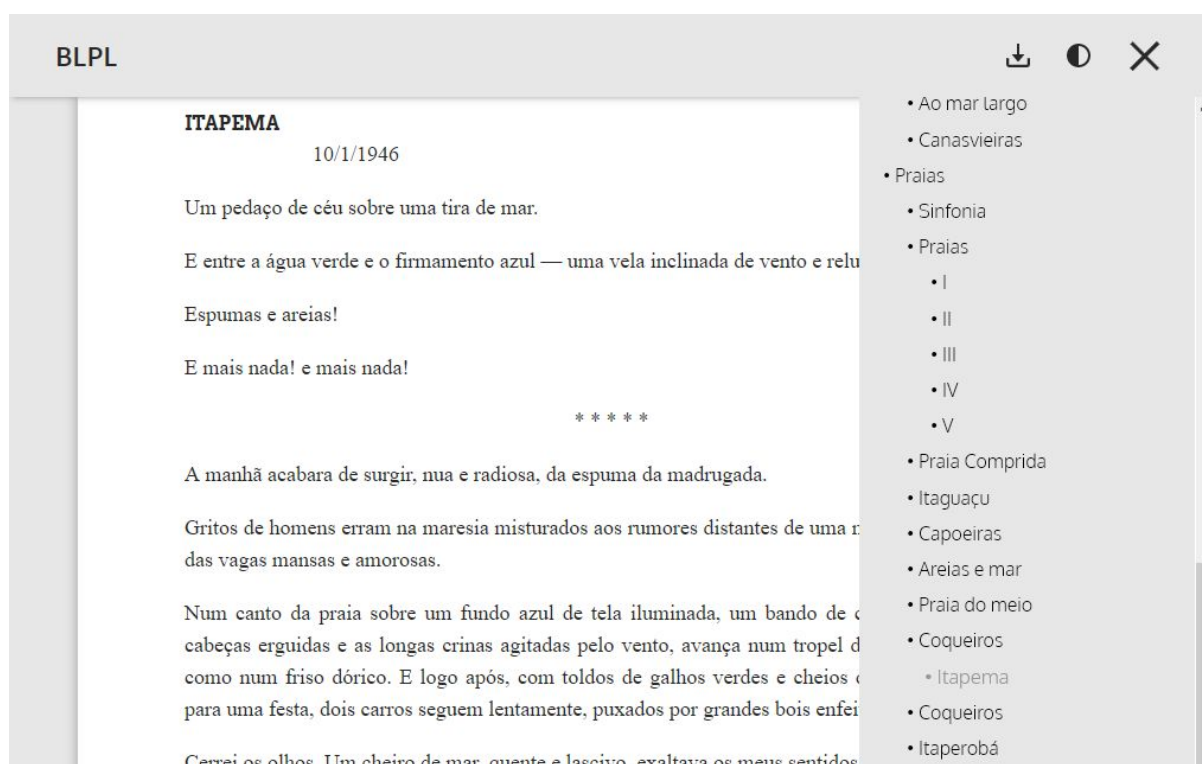
## 2.2 Desenvolvimento do leitor de obras no formato TEI

Após o período inicial de pesquisa, iniciou-se o desenvolvimento do leitor de obras XML normalizadas em TEI que seria utilizado na BLPL. O leitor de obras codificadas em TEI trata-se de uma customização da solução TEI Boilerplate<sup>1</sup>, que permite a publicação de conteúdos estilizados TEI diretamente em navegadores modernos sem que haja o processamento no servidor.

A partir do TEI Boilerplate, foram definidos novos estilos para a formatação de elementos textuais de obras literárias e foram feitas adaptações em algumas estruturas já disponibilizadas pelo boilerplate.

A Figura 1 apresenta a interface Web do leitor TEI desenvolvido no projeto. O desenvolvimento deste leitor foi o mais trabalhoso e demorado, e ocorreu durante praticamente todo o projeto, tendo outras atividades sendo feitas em paralelo. Durante todo o processo de desenvolvimento foram feitas reuniões semanais junto ao orientador para tomar decisões de design e usabilidade, reportar possíveis bugs e propor novas funcionalidades. Algumas dessas reuniões contavam com a participação de pesquisadores do NUPILL (Núcleo de Pesquisas em Informática, Literatura e Linguística) para averiguar se as estruturas renderizadas estavam formatadas da forma correta e, quando necessário, propor alternativas que se adaptassem melhor a diferentes dispositivos, mas que ainda fossem fiéis às versões físicas dos textos.

**Figura 1** - Modelo final do leitor TEI.



Fonte: BLPL.<sup>2</sup>

<sup>1</sup> <<http://teiboilerplate.org/>>

<sup>2</sup> Conto "Homens e Algas", do autor Othon Gama d'Eça. Disponível em: <<https://www.literaturabrasileira.ufsc.br/documentos/?id=130253>>. Acesso em: 29 dez. 2020.

## 2.3 Desenvolvimento do conversor de TEI para HTML

Outro dos grandes tópicos para desenvolvimento que surgiu durante o projeto foi o de criar um conversor de obras em XML para HTML. A necessidade de desenvolver esse conversor veio com a funcionalidade de download das obras no leitor TEI.

Visando manter o download da obra como um só arquivo, optou-se pelo formato HTML com os códigos CSS e JavaScript embutidos. A conversão de XML para HTML mostrou-se necessária pois embutir esses códigos nos arquivos de formato XML resultou em diversas falhas e exigiria várias adaptações.

## 2.4 Desenvolvimento do conversor de obras em HTML para XML

Buscando ampliar o acervo de obras XML normalizadas em TEI, foi sugerido a conversão de algumas obras já existentes no formato HTML para o formato TEI. O conversor já havia sido em grande parte feito pelo pesquisador do NUPILL, Adiel Mittmann, e este bolsista continuou o desenvolvimento organizando o código, corrigindo alguns erros e implementando novas funcionalidades.

Algumas das obras que viriam a ser convertidas tinham sido geradas em diferentes versões do software Microsoft Word e, conseqüentemente, apresentavam diferentes estruturas que acabariam mudando de forma inesperada o funcionamento do conversor. Sendo assim, foi necessário separar essas obras pela versão do software que as gerou, selecionando o elemento na estrutura do documento que contém essa informação, o elemento *meta*.

Outro obstáculo na conversão das obras de HTML para TEI foi a presença ou ausência de uma estrutura de índice. Uma vez que o leitor TEI gera o índice da obra automaticamente selecionando os títulos da mesma torna-se mais adequado a remoção de uma possível estrutura de índice já existente para que seja utilizada somente a do leitor TEI. Com isso, além da separação por versão do software que gerou a obra, foi feita a separação pela condição de a obra ter ou não uma estrutura de índice.

## 3. Produtos gerados

Podemos considerar três produtos gerados durante o projeto:

- O leitor TEI
- O conversor de obras XML em TEI para HTML
- O conversor de obras HTML para XML

Vale citar que todos os produtos desenvolvidos são de código aberto e estão disponíveis de forma pública no repositório *git* do GitHub do bolsista<sup>3</sup>, possibilitando um fácil acesso ao código fonte e envio de sugestões.

### 3.1 Leitor TEI

Atualmente disponível em algumas obras na BPLP, o Leitor TEI possui um design simples, com cores discretas e poucos elementos para que a obra fique em evidência e

---

<sup>3</sup> Disponível em: <<https://github.com/1allan/BLPL>>

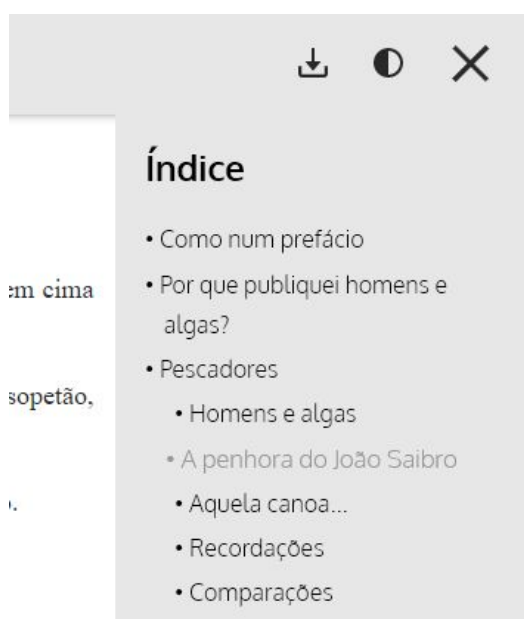
seja possível uma leitura mais clara e sem distrações. Além de buscar uma identidade visual simples, foi tido como de extrema importância que o Leitor fosse responsivo, para permitir a utilização nos mais diferentes tipos de dispositivos.

O Leitor conta também com um menu de índice integrado, possui modo noturno e permite o download da obra em HTML com um índice simples, possibilitando a abertura em qualquer navegador. A obra pode ser baixada clicando no ícone de download presente no cabeçalho do Leitor, como visto na Figura 2 a seguir.



**Figura 2** - Botão de download da obra

O menu de índices é gerado automaticamente pelo Leitor TEI, que busca pelos títulos no documento da obra e a hierarquia dos mesmos e os apresenta de forma ordenada com um espaçamento à esquerda de acordo com a sua hierarquia no texto. Em obras do gênero de conto, por exemplo, a hierarquia é feita pelas subdivisões capítulo, parte e seção e de acordo com elas, o espaçamento é aplicado, sendo: um espaçamento para os capítulos, dois para as partes e três para as seções.



**Figura 3** - Menu de índices

### 3.2 Conversor de obras XML codificadas em TEI para HTML

Para que seja possível a funcionalidade de download das obras no Leitor TEI, é utilizado o conversor de obras XML codificadas em TEI para HTML. O conversor foi desenvolvido em Python e utiliza a biblioteca lxml que facilita em diversos aspectos a leitura e a manipulação de documentos XML e HTML. O funcionamento pode ser descrito em 3 etapas:

- **Leitura dos arquivos:** é feita a leitura da obra no formato XML e dos arquivos de estilo CSS e JavaScript necessários para a renderização da obra no navegador.
- **Filtragem de elementos XML:** uma vez que o consórcio TEI utiliza diversos elementos incompatíveis com HTML, é necessário removê-los ou alterá-los para evitar erros. Nesta etapa o conversor verifica todos os elementos da obra em XML e, caso o elemento seja incompatível, é substituído por um elemento *div* com um atributo *class* com seu nome original. Visando atender as mudanças feitas nos elementos do arquivo XML citadas anteriormente, o script modifica também as especificações contidas no CSS, transformando as descrições dos elementos para classes.
- **Anexação dos arquivos de estilo:** após a filtragem dos elementos, ocorre a anexação dos arquivos de estilo no novo arquivo HTML. Aqui são criados os elementos *style* e *script* para comportar os códigos CSS e JavaScript, respectivamente.

### 3.3 Conversor de obras em HTML para XML

Para facilitar a ampliação do acervo de obras em formato TEI, foi criado o conversor de obras em HTML para XML, permitindo a disponibilização de obras já digitalizadas em HTML no formato XML normalizado de acordo com o TEI, um formato bem mais legível e flexível.

As obras em HTML utilizadas pelo conversor foram em grande parte geradas pelo *Microsoft Word* e os trechos textuais das mesmas são, em sua maioria, diferenciados apenas pelo valor dos atributos *style*. Sendo assim, o conversor busca por todos os elementos do documento que possuem texto, verifica o atributo *style* e classifica se o elemento representa uma linha, um título ou uma subdivisão do texto. Por fim, é adicionado o cabeçalho padrão das obras em TEI da BLPL.

O conversor de obras em HTML para XML pode ser utilizado pelo usuário da BLPL para gerar uma primeira codificação TEI de uma obra a partir do HTML clicando no botão "Converter TEI" no formulário de catalogação de obras literárias, como é mostrado na Figura 4.

Vale ressaltar que, por mais que o conversor consiga criar toda a estrutura principal da obra, ainda é necessário a verificação e alteração manual de alguns elementos gerados, uma vez que alguns textos possuem estruturas muito específicas que não podem ser detectadas de forma automática. Essas edições podem ser feitas diretamente na BLPL, acessando a interface de catalogação da obra e clicando no botão "Editar TEI", como é mostrado na Figura 5.

Idioma principal do documento

Selecione

**Arquivo**

Clique aqui para visualizar o documento 1.15 MB [Converter TEI](#) [Substituir arquivo](#)

Título

Dom Casmurro

Fonte

Descrição

Idioma principal do documento

Selecione

[Atualizar](#) [Remover](#)

\* Campos obrigatórios

**Figura 4 -** Pannel de catalogação de obra na BLPL

URL da mídia

Selecione

**Arquivo**

Clique aqui

Título

Obra em

Fonte

Descrição

Idioma prin

Selecio

[Atualizar](#)

**Editor TEI**

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-stylesheet type="text/xsl" href="../../teibp/xsl/teibp.xsl"?>
3 <TEI xmlns="http://www.tei-c.org/ns/1.0">
4   <teiHeader>
5     <fileDesc>
6       <titleStmt>
7         <title>Homens e Algas</title>
8         <author>Othon Gama D'Eça</author>
9       </titleStmt>
10      <publicationStmt>
11        <p>(Information about distribution of the resource)</p>
12      </publicationStmt>
13      <sourceDesc>
14        <bibl>
15          <title>Homens e algas</title> de Othon d'Eça. Florianópolis: Ed. da
16          UFSC, 2007. </bibl>
17          <bibl>
18            <title>Homens e algas</title> de Othon d'Eça. Florianópolis: Fundação
19            Banco do Brasil; Ed. da UFSC, 1992.</bibl>
20          </sourceDesc>
21        </fileDesc>
22      </teiHeader>
23      <text>
24        <front>
25          <div type="preface">
26            <head>Como num prefácio</head>
27          </div>
28        </front>
29      </text>
30    </TEI>

```

[Fechar](#) [Salvar](#)

**Figura 5 -** Editor de obras TEI na BLPL

## 4. Resultados alcançados

Com o término do projeto, temos como resultado o estabelecimento de uma base sólida para mais um método de ampliar o acervo de obras digitalizadas da Biblioteca Digital de Literatura de Países Lusófonos e a disponibilização de um formato novo de leitura.

Os produtos desenvolvidos durante o projeto já possuem todos os requisitos necessários para que ocorra uma ampla adesão do novo formato de obras em TEI e são abertos para a implementação de novas funcionalidades. Através do conversor de obras em XML para HTML, torna-se possível o download das obras literárias de forma simples. Já o conversor de HTML para XML torna possível a disponibilização de um novo formato para obras já digitalizadas. Por fim, com o Leitor TEI, as obras podem ser lidas num formato claro e em qualquer dispositivo.

Atualmente a BLPL conta com 50 obras disponíveis no formato TEI e essa quantidade tende a aumentar drasticamente assim que o novo formato for mais difundido.

## 5. Comentários do aluno

A partir do projeto eu pude ampliar ainda mais os meus conhecimentos sobre as ferramentas que eu já utilizava e adquiri experiência com XML e XSL, tecnologias que tive quase nenhum contato prévio.

Além de experiências novas com algumas ferramentas, tive, pela primeira vez, contato com o ambiente de desenvolvimento de projetos de extensão, o que fez aumentar ainda mais o meu desejo de desenvolver projetos e pesquisas e possivelmente continuar na academia após a graduação.

Pude também conhecer pessoas excelentes durante o projeto, como: o coordenador do NUPILL, Alckmar Luiz dos Santos, e o pesquisador do NUPILL, Adiel Mittmann.

Por fim, gostaria de agradecer o meu orientador, o professor Dr. Roberto Willrich, por ter me cedido essa oportunidade que, com certeza, agregou de forma positiva e terá bons reflexos no meu futuro como Cientista da Computação.

## 6. Avaliação do orientador sobre o desempenho do bolsista.

Devido a pandemia, as atividades realizadas pelo bolsista foram na sua maior parte em teletrabalho, mas foram mantidas as reuniões semanais síncronas utilizando ferramentas de webconferência. Apesar da distância física, o bolsista demonstrou ótimas habilidades e autonomia para resolver todos os desafios enfrentados durante o desenvolvimento das novas funcionalidades da BPLP. Todos os objetivos propostos para o bolsista foram plenamente atingidos.

Acredito que, com a participação no projeto, o aluno teve a possibilidade de aplicar seus conhecimentos adquiridos no curso de Ciência da Computação. Além disso, o bolsista teve uma oportunidade única para aprofundar seus conhecimentos nas áreas de desenvolvimento de software e tecnologias Web. Estes conhecimentos são fundamentais para a sua vida profissional, por ser a tecnologia na qual ele deverá utilizar em toda a sua carreira.

---

ALLAN SOARES SILVA

Bolsista

---

Prof. Roberto Willrich

Coordenador do Projeto