



UNIVERSIDADE ESTADUAL PAULISTA  
"JÚLIO DE MESQUITA FILHO"  
Campus de Marília



**CULTURA  
ACADÊMICA**  
*Editora*

# Planejamento da estrutura e operacionalização do tesouro Unesp no Tematres

Walter Moreira; Deise Maria Antonio Sabbag;  
Cibele Araújo C. M. dos Santos; José Carlos Francisco dos Santos;  
Luciana Beatriz Piovezan Rio Branco; Érica Fernanda Vitorini

**Como citar:** MOREIRA, W. *et al.* Planejamento da estrutura e operacionalização do tesouro Unesp no Tematres. In: FUJITA, M. S. L.; MOREIRA, W. **Manual do planejamento, construção e manutenção do Tesouro Unesp para bibliotecas: do conceitual a práxis.** Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2021. p. 193-218.

DOI: <https://doi.org/10.36311/2021.978-65-5954-069-3.p193-218>



All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution-Non Commercial-ShareAlike 3.0 Unported.

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição - Uso Não Comercial - Partilha nos Mesmos Termos 3.0 Não adaptada.

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-NoComercial-CompartirIgual 3.0 Unported.

# 6. PLANEJAMENTO DA ESTRUTURA E OPERACIONALIZAÇÃO DO TESAURO NO TEMATRES

*Walter Moreira*

*José Carlos Francisco dos Santos*

*Luciana Beatriz Piovezan Rio Branco*

*Cibele Araújo C. M. dos Santos*

*Deise Maria Antonio Sabbag*

*Érica Fernanda Vitorini*

- 6.1 Introdução
- 6.2 Estrutura do registro de autoridades
- 6.3 Apresentação e leiaute
- 6.4 A escolha do software de gerenciamento do tesauro
- 6.5 Formato de intercâmbio: registros de autoridade em formato MARC
- 6.6 Interoperabilidade com outras linguagens de indexação

Referências

Apêndice D - Manual de conversão registros de autoridades em formato MARC21  
- XML para texto etiquetado

## **6.1 INTRODUÇÃO**

Dentre os aspectos que caracterizam o tesauro, duas características podem ser destacadas como elementos que demandam o seu planejamento cuidadoso: o controle do vocabulário (pela identificação de termos preferidos, de relações de equivalência entre os termos, de padronização etc.) e a estruturação do vocabulário (pela identificação e representação dos conceitos e das relações conceituais no conjunto de termos que o compõe).

Considerando-se o caráter dinâmico inerente às linguagens, de modo geral, e os aspectos pragmáticos da linguagem documentária, de modo particular, a construção e a manutenção de um tesauro é, inevitavelmente, um processo contínuo, que irá requerer esforço e atenção permanentes.

O processo de construção e de manutenção do Tesauro Unesp vem sendo realizado desde 2013 pelo Grupo de Linguagem da Rede de Bibliotecas da Unesp. O Tesauro Unesp foi criado para uso na interface de busca integrada de todas as bases de dados da Universidade. O processo teve início com a utilização da base de registros de autoridades de assuntos da Lista de Cabeçalhos de Assuntos da Rede BIBLIODATA (LCARB) na base de dados Aleph do catálogo Athena. Esse movimento propiciou a compatibilização dos registros de autoridades de assuntos com três outras linguagens: Library Congress Subject Headings (LCSH), Terminologia da Biblioteca Nacional (TBN) e Medical Subject Headings (MeSH) para áreas da Saúde (SANTOS; FUJITA; MOREIRA, 2018).

O Grupo de Linguagem da Rede de Bibliotecas da Unesp, formado por catalogadores e pesquisadores ligados a essa instituição, foi criado e capacitado para lidar com as questões que envolviam a inclusão de termos novos no que ainda se chamava de “Linguagem Unesp”, de modo a promover a substituição gradativa da LCARB adotada anteriormente. Em suas diversas fases, o processo de construção do Tesauro UNESP foi subsidiado teórica e metodologicamente pelo desenvolvimento de variadas investigações, dentre as quais podem ser citadas: a compatibilização de registros de autoridades (FUJITA; PIOVEZAN; SANTOS, 2017; FUJITA; RUBI, 2015), a elaboração de estruturas hierárquicas (FUJITA; SANTOS, 2016; FUJITA et al., 2018) e a análise e seleção de software de gestão de tesouros (FUJITA et al., 2017).

Tendo em vista a complexidade dos procedimentos que envolvem a produção e manutenção de um tesouro, ainda mais quando se trata de um instrumento complexo e com um grande número de entradas, como é o caso do Tesouro Unesp, faz-se necessário organizar de modo sistemático os procedimentos adotados. Nesse sentido, o objetivo deste manual é orientar os procedimentos metodológicos de integração semiautomática dos registros de autoridades da Unesp com o TemaTres, software de gestão de vocabulários utilizado no processo.

## **6.2 ESTRUTURA DO REGISTRO DE AUTORIDADES**

Os pontos de acesso por assuntos são construídos a partir da criação de um registro de autoridade de assunto por meio de um arquivo de autoridade em formato Machine Readable Cataloging - MARC21 (FUJITA; PIOVEZAN; SANTOS, 2017).

Conforme estabelece o MARC21 (2005), o registro de autoridade consiste no estabelecimento de formas padronizadas de nomes (pessoais, corporativos, eventos, jurisdição, título uniforme, combinações nome/título) e assuntos (termos tópicos, nomes geográficos, nomes com subdivisão de assunto, termos e subdivisão de assunto), sendo ambos utilizados como pontos de acesso. Considerando-se isso, foram padronizadas as entradas principais e secundárias no registro bibliográfico com o termo “nome” para entrada principal e secundária e o termo “assunto” no catálogo online construído por meio de software de gestão e manutenção de catálogos (FUJITA; PIOVEZAN; SANTOS, 2017).

O registro de autoridade de assunto é composto por indicações de relações e, sempre que necessário à clareza da delimitação do termo, são complementados com notas de escopo. As principais funções do registro de autoridade de assunto são: a) fornecer o termo de assunto autorizado ou referir ao termo autorizado; b) indicar se um termo pode ou não ser subdividido geograficamente; c) fornecer uma nota de escopo, quando houver necessidade de explicar a cobertura conceitual do termo e d) indicar relações entre termos (FUJITA; SANTOS, 2016). Para fins de ilustração, apresenta-se exemplo de registro de autoridade de assunto de termo tópico do software Aleph em formato OPAC de visualização web e no registro MARC21, do módulo de catalogação (Figuras 1 e 2).



O registro de autoridade define a forma autorizada do cabeçalho no sistema de informação do assunto principal. Também chamada de “termo tópico” ou “assunto tópico”, esse registro compõe o campo MARC 150 do registro de autoridades de um documento. Registram-se também as formas não autorizadas, ou seja, as que não são estabelecidas como assunto principal, mas que podem ser considerados como termos equivalentes (remissivas) e que são dispostas no campo MARC 450 do registro de autoridades. A função da rede de remissivas é promover a conexão precisa entre a busca do usuário e a recuperação efetiva da informação, de modo que o usuário não tenha que pensar em uma sucessão de possíveis termos relacionados ao documento buscado. Com esse recurso, o usuário não precisa buscar necessariamente pelo termo principal, pois o sistema opera tanto com o termo principal quanto com as suas remissivas. Assim, se um usuário pesquisa um assunto que é uma remissiva, o sistema vai rastrear o termo tópico no qual a remissiva está registrada, sendo interligado diretamente com o registro bibliográfico, pois o campo 150 do registro de autoridade é o equivalente ao campo MARC 650 (Assunto – termo tópico) do registro bibliográfico, que contém o termo de indexação (FUJITA; PIOVEZAN; SANTOS, 2017).

Outro tipo de relação entre termos que é indicada no registro de autoridade de assunto é a que ocorre entre o termo genérico e o termo específico, por meio da qual é indicada a posição do termo na hierarquia de assuntos. Um termo genérico é indicado no campo MARC 550 do registro de autoridades, dispondo-se o código “g” no subcampo “w”. Como essa relação é recíproca e assimétrica, o sistema deve ser configurado para incluir a relação inversa automaticamente. Esse tipo de relação entre termos é útil ao catalogador como subsídio à decisão sobre qual termo de assunto indicar como ponto de acesso ao documento, bem como ao usuário, que poderá navegar pela hierarquia de termos para decidir em que nível de generalidade o assunto é mais relevante ao seu interesse (FUJITA; PIOVEZAN; SANTOS, 2017).

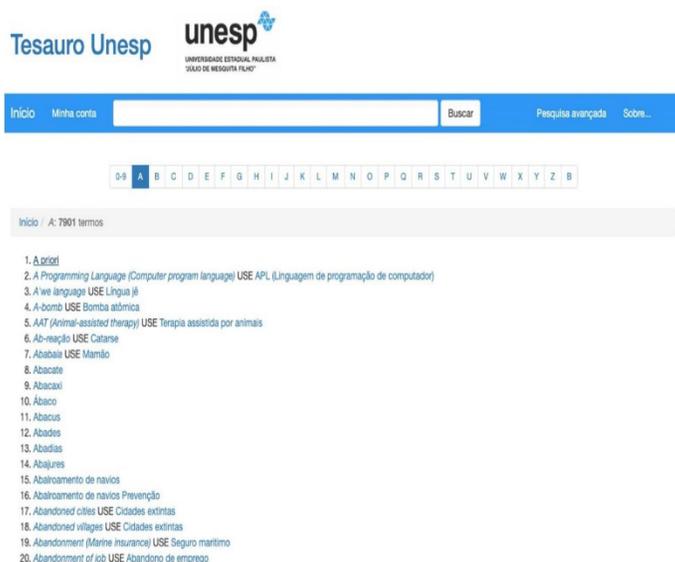
Indica-se também no registro de autoridade as conexões entre termos relacionados, ou termos associados. Nesse tipo de relação vinculam-se entre si os termos que possuem alguma conexão diferente das relações anteriores (de equivalência e de hierarquia). Essa relação também é indicada

no campo MARC 550 subcampo “a” do registro de autoridades (FUJITA; PIOVEZAN; SANTOS, 2017).

### 6.3 APRESENTAÇÃO E LEIAUTE

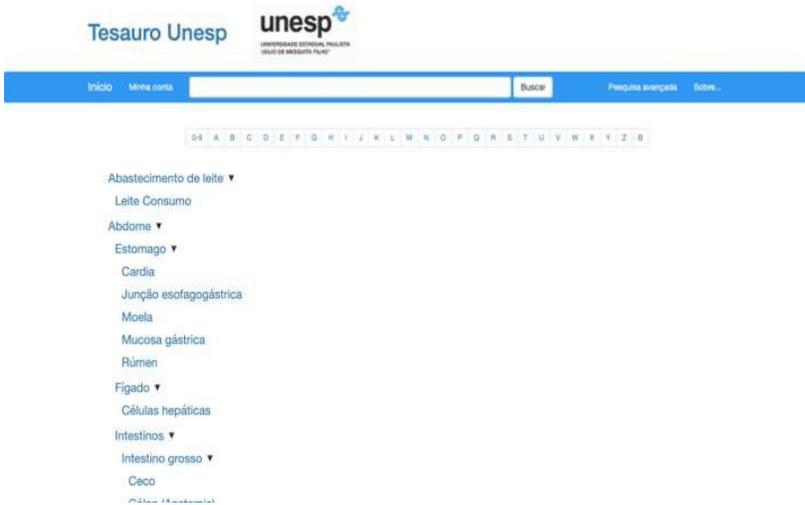
Há, tradicionalmente, dois modos básicos de exibição de um tesouro: um arranjo alfabético e um arranjo sistemático. O arranjo alfabético (Figura 3) permite localizar e acessar os conceitos a partir das palavras inicialmente expressas pelo usuário, com função semelhante à de um índice. O arranjo sistemático (Figura 4) permite visualizar os outros conceitos com os quais o conceito selecionado mantém relação na estrutura do tesouro.

Figura 3 – Apresentação alfabética do Tesouro Unesp



Fonte: Página web do Tesouro Unesp

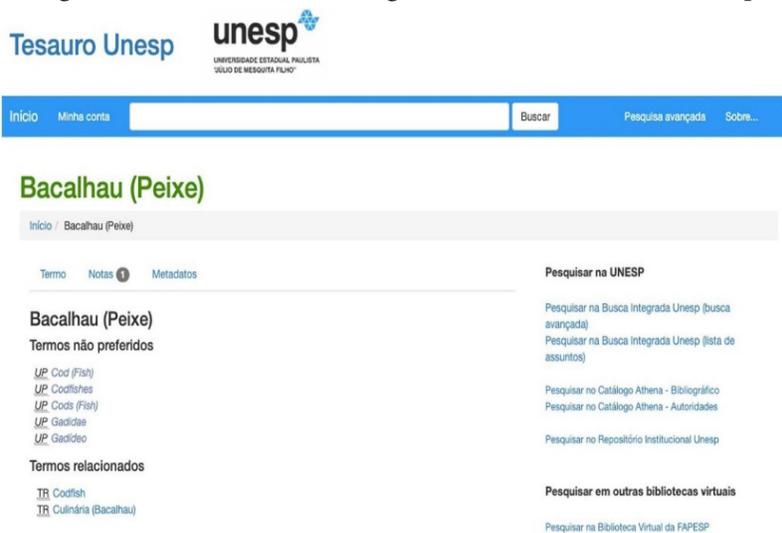
Figura 4– Apresentação sistemática do Tesauro Unesp



Fonte: Página web do Tesauro Unesp

A seleção de um determinado termo, seja a partir do arranjo alfabético seja a partir do arranjo sistemático, exibe a visualização do registro único, mostrando o termo preferido, o(s) termo(s) não preferidos e quaisquer outras relações ou notas acrescentadas ao termo em tela (Figura 5). Nesse modo de visualização apresenta-se também, do lado direito da tela, a integração do tesauro com os diversos sistemas de catálogos online da Unesp e de outras bases.

Figura 5 – Visualização de registro único no Tesouro Unesp



Fonte: Página web do Tesouro Unesp

Somam-se a todos os procedimentos de interoperabilidade dos registros de autoridades entre os softwares Aleph e o TemaTres os resultados quantitativos a partir da execução importação ao TemaTres, conforme o detalhamento apresentado na subseção 5.4. No Quadro 1 apresenta-se o total dos termos distribuídos por níveis, sendo os níveis caracterizados pelas relações hierárquicas que cada termo possui.

Quadro 1 – Quantidade de termos por níveis hierárquicos no Tesouro Unesp

Níveis	Quantidade de termos
1	16764
2	4877
3	2908
4	1356
5	693
6	397

7	393
8	240
9	131
10	49
11	10
12	2

Fonte: Elaborado pelos autores (2019).

No cômputo total, foram importados 89.581 termos, com 46.988 relações entre eles. Destes termos 40.456 são termos preferidos e 49.255 são termos não preferidos. Foram acrescentadas 16.657 notas de escopo aos registros de autoridade importados da base de dados do Aleph.

#### **6.4 A ESCOLHA DO SOFTWARE DE GERENCIAMENTO DO TESAURO**

O software TemaTres, em função das características apresentadas na sequência, foi escolhido para a gestão e manutenção do Tesauro Unesp com a importação dos registros de autoridade compatibilizados. O Tesauro Unesp está atualmente disponível junto à interface de pesquisa integrada das bases de dados da Unesp<sup>1</sup>.

A disponibilização do Tesauro Unesp para consulta é parte fundamental da Política de Indexação das bibliotecas Universitárias da Unesp<sup>2</sup> (FUJITA, 2016). Considera-se que a intermediação do tesauro nas ações de representação de informação pelo indexador e de busca de informação pelo usuário melhora o processo como um todo, com ganhos significativos na recuperação da informação, tornando-a mais precisa e mais eficaz.

O TemaTres é um software livre<sup>3</sup> desenvolvido por Diego Ferreyra, na Argentina, em 2004 (RODRÍGUEZ GAIRÍN; RUSSO GALLO; SULÉ DUESA, 2008). Trata-se de um aplicativo web que pode

<sup>1</sup> [http://parthenon.biblioteca.Unesp.br:1701/primo\\_library/libweb/action/search.do](http://parthenon.biblioteca.Unesp.br:1701/primo_library/libweb/action/search.do)

<sup>2</sup> <https://www.biblioteca.Unesp.br/portal/arquivos/manual-politica-indexacao-2017.pdf>

<sup>3</sup> Software livre é uma denominação que indica que depois de adquirido o software pode ser usado, copiado, estudado, modificado e redistribuído livremente.

ser empregado no gerenciamento de vocabulários, tesouros, taxonomias e representações formais de conhecimento. O TemaTres permite desenvolver estruturas de navegação web e pode funcionar como complemento de um sistema de gerenciamento de bibliotecas, para gerenciar a linguagem documentária adotada pela instituição.

Atualmente em sua versão 3.1, o TemaTres está disponível para download gratuito na página web do projeto<sup>4</sup>, na qual também é possível encontrar um tesouro de demonstração construído com o software. A interface de uso do TemaTres é multilíngue, incluindo os idiomas espanhol, alemão, francês, inglês, italiano e português.

De acordo com o manual de utilização do software, a instalação do TemaTres requer (TEMATRES, 2019): um servidor web que suporte PHP; uma instalação de PHP em funcionamento com versão 4.3.0 ou superior e um servidor de bases de dados. O processo de instalação é relativamente simples, com três passos básicos: a) descarregamento (download) e descompactação do arquivo de instalação; b) configuração da conexão com a base de dados MySQL no arquivo /include/db.tematres.php e c) cumprimento das instruções de instalação do arquivo.

O TemaTres possibilita a criação de linguagens orientadas para o controle de vocabulário e possibilita: a) descrever e controlar, em número ilimitado, diversas relações de equivalência (USE-UP), relações hierárquicas (TG-TE) e relações associativas (TR) entre os termos; permite ainda a criação de novos tipos de relações entre os termos; b) aplicar ferramentas de consistência que possibilitam o estabelecimento de relações entre termos equivalentes, não equivalentes e parcialmente equivalentes; c) atribuir múltiplas notas aos termos, tais como notas de aplicação, de escopo, notas explicativas, históricas, bibliográficas etc.; d) a identificação do status dos termos em relação à sua função no conjunto de termos (termo candidato, termo aceito, termo rejeitado); e) navegar entre termos, de modo sistemático ou alfabético, com sugestão de termos de busca; f) o desenvolvimento e a integração de vocabulários polihierárquicos; g) a exportação completa do vocabulário em formatos que visam à interoperabilidade entre linguagens controladas, por meio de padrões tais como: RDF SKOS-Core; Dublin Core (ISO 158362003) e BS 8723 (Structured Vocabularies for Information Retrieval), entre outros;

---

<sup>4</sup> <https://www.vocabularyserver.com/>

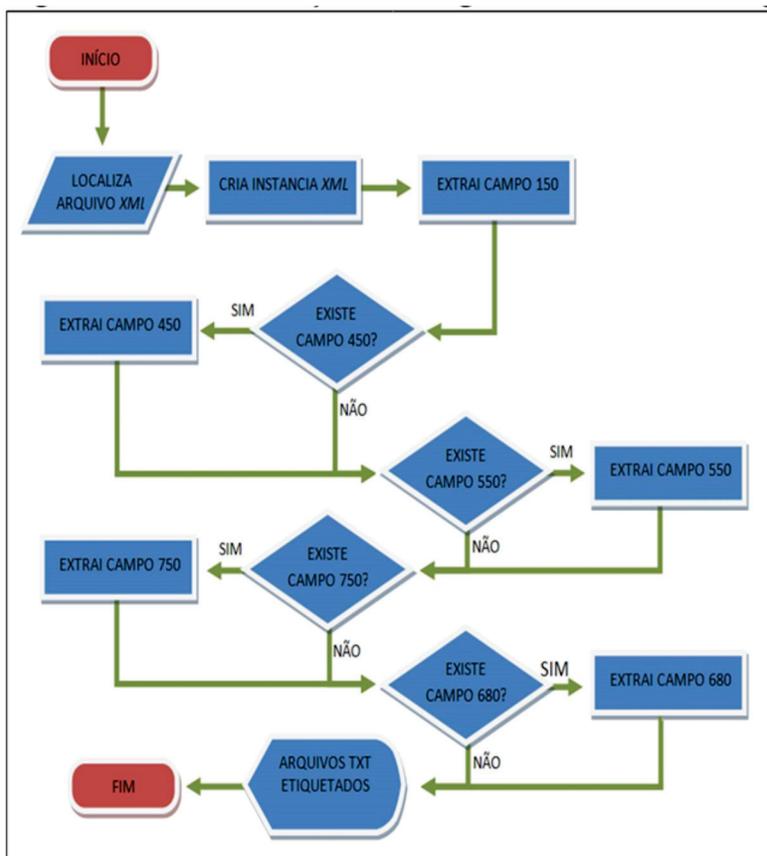
h) a interligação de dados entre vocabulários por meio do gerenciamento de suas relações; i) a auditoria de termos, isto é, a busca por termos livres e repetidos; j) a administração de usuários e editores, considerando-se os níveis de acesso ao sistema; k) a produção de relatórios estatísticos, tais como: média de palavras por termo, termos por número de termos genéricos, termos por número de termos específicos e termos por nível de profundidade.

## **6.5 FORMATO DE INTERCÂMBIO: REGISTROS DE AUTORIDADE EM FORMATO MARC**

A amostragem dos registros de autoridade foi realizada por meio de consulta ao banco de dados do software Aleph, regulados pela aplicação de filtros de registros compatibilizados em que constavam registros bibliográficos vinculados. A identificação dos padrões de tags exportadas no formato MARC21 em suporte XML foi realizada a partir de amostragens em números menores de registros.

O planejamento e desenvolvimento do algoritmo de interoperabilidade para estruturação dos registros resultou da identificação dos padrões, sendo necessária a análise dos registros na íntegra neste formato. Foi estruturado um diagrama de blocos para organizar a extração dos termos e suas relações. Na Figura 6 apresenta-se a sistematização do algoritmo. Este tratamento automático foi desenvolvido em linguagem PHP, para receber e carregar o arquivo XML. A partir dos dados carregados, passou-se a identificação dos termos tópicos e, sequencialmente, à extração das relações (SANTOS; FUJITA; MOREIRA, 2018).

Figura 6 - Sistematização em diagrama de blocos do algoritmo



Fonte: Santos; Fujita; Moreira (2018)

A estruturação dos registros de autoridades em linguagem de importação do TemaTres é realizada a partir da execução do algoritmo proposto na Figura 6. Optou-se pela linguagem etiquetada do TemaTres no suporte de arquivo .txt. Após este tratamento automático, foram gerados os arquivos em .txt com limitações de mil registros em cada arquivo (SANTOS; FUJITA; MOREIRA, 2018).

A linguagem de exportação do sistema de gerenciamento do acervo, neste caso o Aleph, é o formato MARC21 e em suporte de arquivo

.xml dos registros de autoridade. A Figura 7 ilustra um exemplo de um registro nesse formato (SANTOS; FUJITA; MOREIRA, 2018).

Figura 7 - XML em formato MARC dos registros de autoridade

```
1 <?xml version = "1.0" encoding = "UTF-8"?>
2 <collection xmlns="http://www.loc.gov/MARC21/slim"
3 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
4 xsi:schemaLocation="http://www.loc.gov/MARC21/slim
5 http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
6 <record xmlns="http://www.loc.gov/MARC21/slim"
7 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
8 xsi:schemaLocation="http://www.loc.gov/MARC21/slim
9 http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
10 <leader> nz 2200217n 4500</leader>
11 <controlfield tag="001">UNAU000204154</controlfield>
12 <controlfield tag="003">BBD</controlfield>
13 <controlfield tag="005">20170912154512.0</controlfield>
14 <controlfield tag="008">960517 n anznbnabn a ana d</controlfield>
15 <datafield tag="040" ind1=" " ind2=" ">
16 <subfield code="a">UNESP</subfield>
17 <subfield code="b">por</subfield>
18 </datafield>
19 <datafield tag="150" ind1=" " ind2=" ">
20 <subfield code="a">Projeto estrutural</subfield>
21 </datafield>
22 <datafield tag="360" ind1=" " ind2=" ">
23 <subfield code="i">Como subcabecalho usar Projetos e construção sob tipos de estruturas</subfield>
24 </datafield>
25 <datafield tag="550" ind1=" " ind2=" ">
26 <subfield code="w">g</subfield>
27 <subfield code="a">Projetos de engenharia</subfield>
28 </datafield>
29 <datafield tag="550" ind1=" " ind2=" ">
30 <subfield code="a">Projeto arquitetônico</subfield>
```

Fonte: Santos; Fujita; Moreira (2018).

O registro é marcado com a tag <record>, nas linhas com a tag <controlfield> descrevem-se as características do registro. As tags <datafield> apresentam a estrutura dos termos. No Quadro 2 apresentam-se as codificações dos registros de autoridades em MARC21 (SANTOS; FUJITA; MOREIRA, 2018).

Quadro 2 - Descrições dos termos em codificação MARC2

Código	Descrição	Exemplo tag XML	Etiqueta TemaTres
150	Termo tópico – termo utilizado	<datafield tag="150" ind1=" " ind2=" ">	-
450	Equivalência	<datafield tag="450" ind1=" " ind2=" ">	UF
550	Relacionados	<datafield tag="550" ind1=" " ind2=" ">	-
w	h – Específico	<subfield code="w">h</subfield>	NT
w	g – Genérico	<subfield code="w">g</subfield>	BT
750	Equivalente na LC	<datafield tag="750" ind1=" " ind2="0">	RT
680	Nota de escopo	<datafield tag="680" ind1=" " ind2=" ">	NA

Fonte: Santos; Fujita; Moreira (2018).

No Quadro 2 foram apresentados os códigos e as tags equivalentes no formato MARC21 exportado no suporte de arquivo .xml. Pode-se perceber que os atributos das tags são as indicações de referência do termo. No primeiro exemplo <datafield tag="150" ind1=" " ind2=" ">, a expressão datafield demonstra que os dados do campo serão descritos; logo na sequência o atributo tag indica a codificação a que se refere esse termo, neste caso seu valor é 150, indicando um termo tópico. Subordinadamente é apresentado a tag <subfield code="a"> Termo exemplo </subfield>, ilustrado na Figura 7. O subfield com o atributo code valorado em a indica o conjunto de caracteres que representam o termo, neste caso o termo é um Termo exemplo (SANTOS; FUJITA; MOREIRA, 2018).

Nos testes realizados para determinação do padrão de dados dos termos em .xml, percebeu-se a existência de termos compostos e representados em dois ou mais subfield, na Figura 8 é possível visualizar, na linha 209 e 210 do XML e no texto etiquetado está assinalado com # os dois subcampos (SANTOS; FUJITA; MOREIRA, 2018).

Figura 8 - Representação XML e texto etiquetado – termos compostos

<pre>195 &lt;record xmlns="http://www.loc.gov/MARC21/slim" 196 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" 197 xsi:schemaLocation="http://www.loc.gov/MARC21/slim 198 http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd"&gt; 199 &lt;leader&gt; nz 2200217n 4500&lt;/leader&gt; 200 &lt;controlfield tag="001"&gt;UNAU000208442&lt;/controlfield&gt; 201 &lt;controlfield tag="003"&gt;BBD&lt;/controlfield&gt; 202 &lt;controlfield tag="005"&gt;20170912100903.0&lt;/controlfield&gt; 203 &lt;controlfield tag="008"&gt;950327dn anznbabn a ana d&lt;/control 204 &lt;datafield tag="040" ind1=" " ind2=" "&gt; 205 &lt;subfield code="a"&gt;UNESP/BGI&lt;/subfield&gt; 206 &lt;subfield code="b"&gt;por&lt;/subfield&gt; 207 &lt;/datafield&gt; 208 &lt;datafield tag="150" ind1=" " ind2=" "&gt; 209 &lt;subfield code="a"&gt;Aço&lt;/subfield&gt; 210 &lt;subfield code="x"&gt;Estruturas&lt;/subfield&gt; 211 &lt;/datafield&gt; 212 &lt;datafield tag="450" ind1=" " ind2=" "&gt; 213 &lt;subfield code="a"&gt;Estruturas de aço&lt;/subfield&gt; 214 &lt;/datafield&gt; 215 &lt;datafield tag="450" ind1=" " ind2=" "&gt;</pre>	<p>Aço#Estruturas (UNAU000208442) UF: Estruturas de aço UF: Structural steel BT: Engenharia civil BT: Materiais de construção BT: Vigas RT: Construção metálica RT: Pontes metálicas RT: Steel,Structural</p>
--	---

Fonte: Santos; Fujita; Moreira (2018).

Em outros casos, as variações ocorrem nos atributos das tags indicando as referências conforme apontado no Quadro 2. Portanto, o resultado dessa atividade centrou-se na coleta dessas informações, as quais por sua vez, não são tão perceptíveis visualmente. Adicionalmente, alguns testes foram realizados para que se pudesse observar o comportamento dos registros. Na Figura 9, apresenta-se a interface de importação do TemaTres, referindo-se a um dentre os padrões de importação. O texto etiquetado foi selecionado para este estudo (SANTOS; FUJITA; MOREIRA, 2018).

Figura 9 - Padrão de importação no TemaTres

Tesouro Unesp unesp  
UNIVERSIDADE ESTADUAL PAULISTA  
"JULIO DE MESQUITA FILHO"

Início Minha conta  Buscar Pesquisa avançada

Importar tesouro de um arquivo txt tabulado

Selecionar formato

Arquivo  Nenhum arquivo selecionado

Texto tabulado:

```
South America
  Argentina
    Buenos Aires = Bs As
  Brazil
  Uruguay
```

Texto etiquetado:

```
South America
ET: America
NT: Argentina
UF: South-america
RT: Latin America
```

Fonte: Santos; Fujita; Moreira (2018).

Os subsídios dessa atividade de importação darão suporte para a programação do algoritmo de transição de linguagem, interoperabilidade, a partir da extração e conversão em texto etiquetado (SANTOS; FUJITA; MOREIRA, 2018).

A partir da execução da etapa anterior passa-se para a organização e implementação do código do algoritmo de interoperabilidade na linguagem PHP. Alguns requisitos são elementos necessários para a compilação do código fonte e análise dos testes. Para tanto, utilizou-se ferramenta de edição de programas em PHP (Komodo Edit 10) e servidor Apache (EasyPHP – Devserver 17). O algoritmo é executado em servidor local simulando a web. Na Figura 10, é apresentada parte do código do algoritmo, que por sua vez sistematiza a proposta ilustrada na Figura 6 (SANTOS; FUJITA; MOREIRA, 2018).

Figura 10 - Codificação em PHP do algoritmo

```
27 foreach ($records as $record) {
28     ++$cont;
29
30     // echo 'controledfield: '. $record->controlfield[0];
31     $id=$record->controlfield[0];
32     $datafields= $record->datafield;
33     foreach ($datafields as $datafield){
34     switch ($datafield->attributes()-> tag){
35         case '150':
36             if (isset ($datafield->subfield[0])) {
37                 $termot= $datafield->subfield[0];
38             }
39             if (isset ($datafield->subfield[1])) {
40                 $termot.= ' '.$datafield->subfield[1];
41             }
42             //$termot.= ' ('.$id.')';
43             $texto.="\\r\\n". $termot;
44             break;
45         case '450':
46             $termoUF= 'UF: ';
47             if (isset ($datafield->subfield[0])) {
48                 $termoUF.= $datafield->subfield[0];
49             }
50             if (isset ($datafield->subfield[1])) {
51                 $termoUF.= ' '.$datafield->subfield[1];
52             }
53             $texto.= "\\r\\n". $termoUF;
54             break;
55     }
```

Fonte: Santos; Fujita; Moreira (2018)

Nas linhas 27 e 33 da Figura 10, ocorrem as sistematizações dos registros do arquivo XML, de maneira a proporcionar a navegabilidade entre os registros para extração dos termos e suas relações. A verificação da qualificação do termo é realizada nas linhas 35 e 45 da mesma figura.

O trabalho de estruturação dos registros de linguagens de importação do TemaTres é o resultado da execução do desenvolvimento do algoritmo de interoperabilidade. Como resultado apresenta-se, na Figura 11, a codificação etiquetada, a partir dos resultados do algoritmo (SANTOS; FUJITA; MOREIRA, 2018).

Figura 11 - Registro no formato etiquetado

```
2 Placas (Engenharia)
3 UF: Disks (Mechanics)
4 UF: Panels
5 UF: Structural plates
6 BT: Placas e cascas elásticas
7 BT: Análise estrutural (Engenharia)
8 RT: Cascas (Engenharia)
9 RT: Plates (Engineering)
10 NA: (UNAU000201540) Usado para trabalhos sobre placas como estruturas de engenharia. Para trabalhos sobre cascas
11
12 Trajes História
13 UF: Indumentária medieval
14 UF: Trajes medievais
15 UF: Costume, Medieval
16 RT: Costume History
17 NA: (UNAU000201549)
18
19 Contos espíritas
20 BT: Ficção espírita
21 RT: Spiritual short stories
22 NA: (UNAU000201603)
23
24 Tabeua caraíba
25 UF: Caraibeira
26 UF: Caroba-do-campo
27 UF: Cinco-em-rama
28 UF: Craibeira
29 UF: Ipá-amarelo-do-cerrado
30 UF: Para-tudo
```

Fonte: Santos; Fujita; Moreira (2018).

A identificação dos registros aparece etiquetada como NA (nota de aplicação) e alguns exemplos de sua utilização podem ser visualizados nas linhas 10, 17 e 22 da Figura 11. Em cada arquivo .txt, são incluídos 1000 registros e conseqüentemente nomeados com a terminação de enumeração baseada no controle do último registro nele incluído, por exemplo: arquivo\_texto\_1000.txt, arquivo\_texto\_2000.txt, arquivo\_texto\_3000.txt e assim sucessivamente até o término dos registros. Essa medida foi utilizada para não congestionar a submissão do arquivo e para que a execução pudesse ser mais rapidamente efetuada no servidor local (SANTOS; FUJITA; MOREIRA, 2018).

O resultado da etapa final deste trabalho é a submissão, a partir dos resultados gerados pelo algoritmo de interoperabilidade, dos arquivos .txt etiquetados conforme os padrões estabelecidos para importação de termos apresentados na Figura 11. Isso foi realizado com o recurso da função de importação dos arquivos conforme a interface do Tema Três para tal procedimento, localizada no menu de administração do vocabulário controlado, ilustrado na Figura 9 (SANTOS; FUJITA; MOREIRA, 2018). A execução dessas atividades para obtenção da importação é descrita no Apêndice C, no Manual de conversão registros de autoridades em formato MARC21 - XML para texto etiquetado.

## **6.6 INTEROPERABILIDADE COM OUTRAS LINGUAGENS DE INDEXAÇÃO**

Zeng e Chan (2004) utilizam a definição elaborada pela American Library Association (2000) para conceituar a interoperabilidade como a capacidade de diferentes sistemas ou componentes de trocarem informações e de utilizarem essas informações sem esforço especial por parte dos sistemas. Assim entendida, a interoperabilidade é a capacidade de sistemas distintos se comunicarem com informações transparentes por meio da admissão de padrões comuns e protocolos que viabilizem o compartilhamento e o uso da informação (MOURA, 2009).

Na literatura podem-se encontrar vários termos conceitualmente próximos ao conceito de interoperabilidade, tais como: compatibilização, unificação, integralização, interoperabilidade, convertibilidade. Ainda que não seja objetivo deste capítulo esclarecer tais diferenças, é importante registrar que, na condição de termos pertencentes ao mesmo campo conceitual, tais expressões são utilizadas algumas vezes como sinônimas, o que pode causar prejuízos à compreensão do processo.

Para que a interoperabilidade ocorra, faz-se necessário o estabelecimento das camadas de representação estrutural, sintática, semântica e lógica: Unicode/URI; XML/Namespase/XML Schema; RDF/RDF Schema; Ontológica; Lógica; Prova e Confiança. A camada ontológica realiza a integralização terminológica articulando a terminologia visando minimizar os conflitos. Ou seja, é necessário que seja estabelecido um método de compatibilização de linguagens, no caso, um método de compatibilização de linguagens de indexação.

O método de compatibilização (ou medida de compatibilização) realiza a similaridade entre duas ou mais linguagens introduzindo o conceito de graus de compatibilidade e fazendo com que se estabeleça a distribuição entre compatibilidade no plano semântico e no plano linguístico.

Conforme Campos (2017), a compatibilidade é a qualidade de articulação entre vocabulários afins (direta ou indiretamente) definindo equivalências conceituais entre os termos por meio de relações de semelhanças.

Assim, para que a compatibilização aconteça é necessário: criar uma matriz de compatibilização (mapeamento das potencialidades semânticas e linguísticas das linguagens); identificar a taxa de coincidência

verbal (análise da medida de similaridade entre o símbolo linguístico e seu conteúdo conceitual); e estabelecer o grau de compatibilidade conceitual (coincidência conceitual, correspondência conceitual e correlação conceitual).

Neste sentido e orientado por essa abordagem, o processo de construção do Tesouro Unesp utilizou elementos de cinco linguagens existentes, a saber, Lista de Cabeçalhos de Assunto da Rede BIBLIODATA (LCARB); Terminologia da Biblioteca Nacional (TBN); Library of Congress Subject Headings (LCSH) e Medical Subject Headings (MeSH), Vocabulário Controlado da USP (VocaUSP).

A partir do recurso a essas quatro linguagens foi realizada a compatibilização dos registros conforme segue: a) registros de autoridade: a base de registros da LCARB foi internalizada no catálogo Athena possibilitando a construção da Linguagem Unesp a partir da compatibilização dos registros de autoridades existentes e atualização as informações pela TBN; b) para registros não traduzidos pela LCARB foram utilizadas as linguagens da LCSH e o MeSH.

A escolha da TBN foi realizada por meio de estudos comparativos que demonstraram a viabilidade desta linguagem em substituir a LCARB. A TBN possui maior atualidade sendo uma tradução da LCSH que tem como características uma estrutura hierárquica visível com relações associativas de tesouro (FUJITA et al., 2018).

Dessa forma, os bibliotecários catalogadores passaram a importar registros de autoridade com termos novos constantes das linguagens LCSH, TBN e MeSH. Algumas necessidades importantes foram levadas em consideração para a construção da linguagem, notadamente a atualização e a garantia de compatibilização para os registros importados da LCARB (FUJITA et al., 2018).

A linguagem desenvolvida e utilizada pela USP (VocaUSP) foi utilizada por sua semelhança de propósitos com a Linguagem Unesp e por ter sido desenvolvida com embasamento terminológico nas áreas específicas de conhecimento com as quais a USP trabalha, mesmo que esses fundamentos terminológicos sejam diferentes da LCSH (FUJITA et al. 2018).

## REFERÊNCIAS

- AMERICAN LIBRARY ASSOCIATION (ALA). *Committee on Cataloging: Description and Access (CC:DA) Task Force on Metadata*. Final report. 2000. Disponível em: <http://www.libraries.psu.edu/tas/jca/ccda/reports.html>. Acesso em: 05 nov. 2019.
- BRASILEIRO, I. B.; SANTOS, J. C. F.; MOREIRA, W. TemaTres e o Koha. In: SHINTAKU, M. (org.). *Guia do usuário do TemaTres*. Brasília: Ibict, 2019. p. 121146.
- CAMPOS, M. L. A. Interoperabilidade semântica e compatibilização de linguagens em ambientes heterogêneos: a questão do acesso aberto e as possibilidades de tratamentoterminológico. In: ENCONTROS PRÉ-CONFOA: GESTÃO DE CONTEÚDO E CRIS, 3., 2017. Rio de Janeiro. *Encontros [...]*. Rio de Janeiro: Fiocruz/ICICT, 2017. p. 1-47.
- FUJITA, M. S. L. *et al.* Avaliação das características do TemaTres e Multites para o controle de autoridades nas bibliotecas universitárias. *Scire*, Zaragoza, v. 23, n. 2, p. 7181, 2017.
- FUJITA, M. S. L. *et al.* Construction and evaluation of hierarchical structures of indexing languages for online catalogs of libraries: an experience of the São Paulo State University (UNESP). *Knowledge Organization*, Baden-Baden, v. 45, n. 3, p. 220-231, 2018.
- FUJITA, M. S. L. *Política de indexação para bibliotecas: elaboração, avaliação e implantação*. Marília; São Paulo: Oficina Universitária; Cultura Acadêmica, 2016.
- FUJITA, M. S. L.; PIOVEZAN, L. B.; SANTOS, N. S. A função do registro de autoridade de assunto na construção e uso de linguagens de indexação para catálogos online. In: SIMÕES, M. G.; BORGES, M. M. (coord.). *Tendências atuais e perspectivas futuras em organização do conhecimento: atas do III Congresso ISKO Espanha-Portugal; III Congresso ISKO Espanha*. Coimbra: Universidade de Coimbra, 2017. p. 577-586.
- FUJITA, M. S. L.; RUBI, M. Compatibilidade entre linguagens documentais para construção, atualização e adequação de vocabulário de bibliotecas universitárias. In: RODRÍGUEZ MUÑOZ, J. V. *et al.* (org.). *Organización del conocimiento: sistemas de información abiertos: actas del XII Congreso ISKO España y II Congreso ISKO España y Portugal*. Murcia: Universidad de Murcia, 2015 p. 345-356.
- FUJITA, M. S. L.; SANTOS, L. B. P. A estrutura lógico-hierárquica de linguagens de indexação utilizadas por bibliotecas universitárias. *Scire*, Zaragoza, v. 22, n. 2, p. 37-46, 2016.
- MARC21. *MARC 21: formato condensado para dados de autoridade*. Tradução e adaptação de Margarida M. Ferreira. Marília: FUNDEPE, 2005.

MOURA, M. A. Informação, ferramentas ontológicas e redes sociais ad hoc: a interoperabilidade na construção de tesouros e ontologias. *Informação & Sociedade*, João Pessoa, v. 19, n. 1, p. 59-73, jan. /abr. 2009.

RODRÍGUEZ GAIRÍN, J. M.; RUSSO GALLO, P.; SULÉ DUESA, A. A virtual exhibition of open source software for libraries, 2008. In: *BOBCATSSS SYMPOSIUM*, 16., 2008, Zadar (Croácia). Disponível em: <http://hdl.handle.net/10760/11151>. Acesso em: 13 dez. 2015.

SANTOS, J. C. F.; FUJITA, M. S. L.; MOREIRA, W. Tesouro Unesp: integração do registro de autoridade para o TemaTres. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Londrina. *Anais [...]*. Londrina: UEL; ANCIB, 2018.

TEMATRES. *TemaTres*: servidor de vocabularios controlados. 2019. Disponível em: <http://r020.com.ar/tematres/manual/>. Acesso em: 30 out. 2019.

ZENG, M. L.; CHAN, L. M. Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society or InformationScience and Technology*, New York, v. 55, n. 5, p. 377–395, 2004.

## APÊNDICE D

### MANUAL DE CONVERSÃO REGISTROS DE AUTORIDADES EM FORMATO MARC21 - XML PARA TEXTO ETIQUETADO

Este manual tem por finalidade orientar o usuário no processo de preparação dos dados e arquivos para importação no TemaTres. A apresentação deste manual foi publicada no capítulo 5 do livro Guia do Usuário do TemaTres (BRASILEIRO; SANTOS; MOREIRA,2019). Para efeito de facilidade para o usuário está representado o passo-a-passo. O sistema de conversão tem como principal interface apresentada na Figura 1.

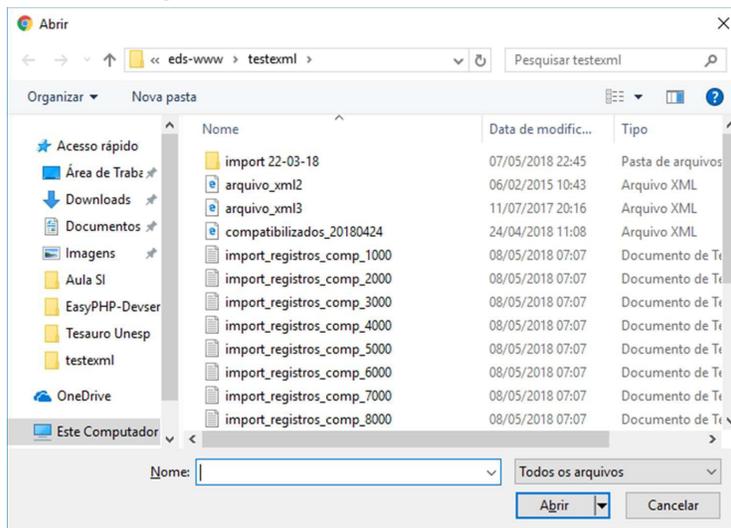
Figura 1 – Interface do sistema de conversão

Selecione o arquivo xml:  Nenhum arquivo selecionado

Fonte: Brasileiro; Santos; Moreira (2019, p. 142)

Clique em “Escolher arquivo”, conforme Figura 1. Após esse procedimento será aberta a janela para selecionar o arquivo para ser carregado (Figura 2).

Figura 2 – Janela de seleção de arquivos



Fonte: Brasileiro; Santos; Moreira (2019, p.143)

Ao selecionar o arquivo clique em abrir, logo o nome do arquivo será mostrado ao lado do botão “Escolher arquivo”, apresentado na Figura 3. Para executar os processos de conversão clique em “Enviar”.

Figura 3 – Interface do sistema com a seleção do arquivo

Selecione o arquivo xml:  compatibiliz...20180424.xml

Fonte: Brasileiro; Santos; Moreira (2019, p. 143)

Após o término da execução serão listados os arquivos em formato .txt, etiquetados no padrão de importação no software TemaTres, conforme Figura 4.

Figura 4 – Lista de arquivos convertidos para downloads

**Arquivo enviado com sucesso!**  
**Gerando os arquivos para o TemaTres...**  
Arquivo com registros até 1000: [BAIXAR](#)  
Arquivo com registros até 2000: [BAIXAR](#)  
Arquivo com registros até 3000: [BAIXAR](#)  
Arquivo com registros até 4000: [BAIXAR](#)  
Arquivo com registros até 5000: [BAIXAR](#)  
Arquivo com registros até 6000: [BAIXAR](#)  
Arquivo com registros até 7000: [BAIXAR](#)  
Arquivo com registros até 8000: [BAIXAR](#)  
Arquivo com registros até 9000: [BAIXAR](#)  
Arquivo com registros até 10000: [BAIXAR](#)  
Arquivo com registros até 11000: [BAIXAR](#)  
Arquivo com registros até 12000: [BAIXAR](#)  
Arquivo com registros até 13000: [BAIXAR](#)  
Arquivo com registros até 13534: [BAIXAR](#)

Fonte: Brasileiro; Santos; Moreira(2019,p.144)

Clique em “BAIXAR” para salvar no computador. Após acessar o tesouro no endereço <https://www.biblioteca.Unesp.br/tesauro>, faça o login, conforme disponibilizado pelo departamento de TI (Figura 5).

Figura 5 – Interface do TemaTres para efetuar login



Fonte: Elabora do pelos autores, 2019 .

Uma vez logado no sistema do Tesauro Unesp, acesse o item Menu>Administração>Manutenção da Base>Importar, conforme Figura 6.



Fonte: Elabora do pelos autores, 2019.

Conforme mostra a janela de importação, apresentada na Figura 7, selecione o formato a ser importado, neste caso “Texto etiquetado”, e depois clique em “Arquivo” para abrir a janela de seleção de arquivos conforme Figura 8.

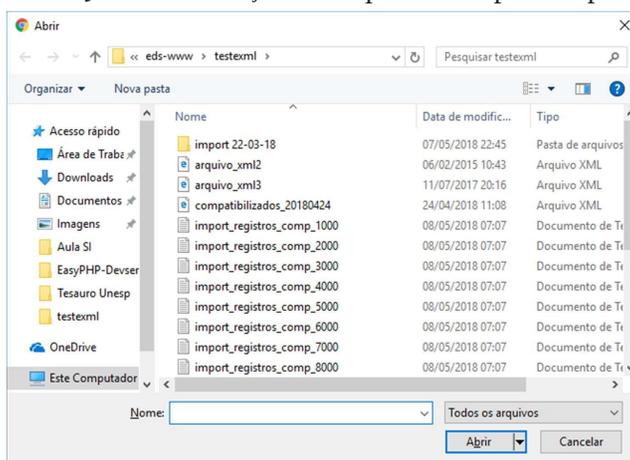
Figura 7 – Interface TemaTres para importação de registros em TXT



Fonte: Elaborada pelos autores, 2019.

Selecione o arquivo e clique em “Abrir” para executar a importação. Ao término do processo será exibida a mensagem de finalização. Fique atento para não clicar em qualquer outro botão ou clicar mais de uma vez para “Salvar”. O processo demora alguns instantes, pode ser percebido que está executando devido ao navegador representar o carregamento de uma página da web.

Figura 8 – Janela de seleção de arquivos .txt para importação



Fonte: Elaborado pelos autores, 2019.