RELATÓRIO DE ACOMPANHAMENTO – META 03

Estudos voltados ao desenvolvimento de vocabulário controlado nos temas de atuação, disponibilizado em sistema de informação

2023

ibict
Instituto Brasileiro de Informação em Ciência e Tecnologia
UNIDADE DE PESQUISA DO MCTI

PRESIDÊNCIA DA REPÚBLICA

Jair Messias Bolsonaro
Presidente da República

Hamilton Mourão
Vice-Presidente da República

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

Marcos Cesar Pontes
Ministro da Ciência, Tecnologia e Inovações

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA

Cecilia Leite Oliveira
Diretora

Reginaldo de Araújo Silva
Coordenação de Administração – COADM

Gustavo Saldanha
Coordenação de Ensino e Pesquisa, Ciência e Tecnologia da Informação – COEPE

José Luis dos Santos Nascimento
Coordenação de Planejamento, Acompanhamento e Avaliação – COPAV

Anderson Itaborahy
Coordenação-Geral de Pesquisa e Desenvolvimento de Novos Produtos – CGNP

Bianca Amaro de Melo
Coordenação-Geral de Pesquisa e Manutenção de Produtos Consolidados – CGPC

Tiago Emmanuel Nunes Braga
Coordenação-Geral de Tecnologias de Informação e Informática – CGTI

Milton Shintaku
Coordenação de Tecnologias para Informação – COTEC

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES
Instituto Brasileiro de Informação em Ciência e Tecnologia

RELATÓRIO DE ACOMPANHAMENTO – META 03

Estudos voltados ao desenvolvimento de vocabulário controlado nos temas de atuação, disponibilizado em sistema de informação

Coordenação de Tecnologias
para Informação (COTEC)
Brasília
2022

© Instituto Brasileiro de Informação em Ciência e Tecnologia – Ibict (2022)

EQUIPE TÉCNICA

Diretora do Instituto Brasileiro de Informação em Ciência e Tecnologia
Cecilia Leite Oliveira

Coordenador-Geral de Tecnologias de Informação e Informática – CGTI
Tiago Emmanuel Nunes Braga

Coordenador do Projeto
Milton Shintaku

Organizadores do Relatório
Milton Shintaku
Jaqueline Rodrigues de Jesus
Mirele Carolina Souza Ferreira Costa
Raíssa da Veiga de Menêses

Pesquisadores
Deise Maria Antonio Sabbag
Fernanda Maciel Rufino
Fernando de Jesus Pereira
Jaqueline Rodrigues de Jesus
Lucas Ângelo da Silveira
Mirele Carolina Souza Ferreira Costa
Priscila Rodrigues dos Santos
Raíssa da Veiga de Menêses
Renata Monteiro Rodrigues
Rosilene Paiva Marinho de Sousa
Victor Ramos Silva

Normalização
Jaqueline Rodrigues de Jesus

Revisão
Flavia Karla Ribeiro Santos
Rafael Teixeira de Souza

Capa e Projeto Gráfico
Victor Ramos Silva

Este Relatório é um produto do Projeto Estudos para Implementação de Repositório Institucional na Biblioteca Terezinha Ayres Costa.

Ref. Processo SEI Nº 01302.000479/2018-80 (Processo de Contratação)
Ref. Processo SEI Nº 01302.000437/2020-63 (Processo de Execução)
Ref. FUNDEP - 28139

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto Brasileiro de Informação em Ciência e Tecnologia ou do Ministério da Ciência, Tecnologia, Inovações e Comunicações.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

SUMÁRIO

1 Introdução 06

2 Objetivos 07

2.1 Objetivo geral 07

2.2 Objetivos específicos 07

3 Resultados 08

3.1 Estudos iniciais 08

3.2 Estudos para adaptação de vocabulários controlados existentes, como o DECS 09

3.2.1 Implementação do Sistema de Tesauros da Anvisa 11

3.2.2 Implementação do Layout no Sistema de Tesauros da Anvisa 13

3.2.3 Migração do tesauros do ambiente computacional do Ibict para a Anvisa 14

3.3 Proposição de adaptação do repositório para uso automático de vocabulário controlado 15

3.2.1 Alteração no TemaTres e DSpace para a integração 18

4 Considerações Finais 30

1 INTRODUÇÃO

Desde a criação das bibliotecas físicas, o processo de indexação foi um dos mais desafiadores, visto que precisava indicar as palavras-chave que representassem o conteúdo dos documentos na descrição temática. Com as bibliotecas digitais, esse processo teve a complexidade intensificada, tendo em vista a diversidade de documentos que passaram a figurar nesse sistema de informação, requerendo bases de termos autorizados maiores e com a dinamicidade dos sistemas informatizados.

Em bibliotecas digitais de governo, esse problema tornou-se mais evidente, na medida em que oferta informação a todos os cidadãos. Nesse sentido, requer termos que auxiliem a recuperação da informação, que representem a área de atuação do órgão. Por isso, uma das soluções pode ser encontrada em tesauros, que podem ser utilizados como bases de termos autorizados e disseminam a terminologia de uma área.

Nesse contexto, em setembro de 2020, a Agência Nacional de Vigilância Sanitária (Anvisa) firmou um projeto de pesquisa, por meio do Termo de Execução Descentralizada (TED) nº 06/2020, com o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), que tem como objeto "Estudos para Implementação de Repositório Institucional (RI) da Anvisa". Com esse projeto, o Ibict, enquanto instituto de pesquisa especializado em sistemas de informação, apoia a Anvisa, grande produtora de documentação técnica, a implementar a sua Biblioteca Digital.

Em meio às metas acordadas entre a Anvisa e o Ibict no projeto de pesquisa, estava a implementação do Tesauro da Anvisa, como meta descrita por "Estudos voltados ao desenvolvimento de vocabulário controlado nos temas de atuação, disponibilizado em sistema de informação". Com isso, estudos foram realizados com enfoque em termos do vocabulário que compõe o tesauro, assim como na ferramenta para registro e disseminação.

2 OBJETIVOS

2.1 Objetivo geral

Retratar o cumprimento da Meta 3 do projeto de pesquisa firmado entre a Anvisa e o Ibict por meio de seus resultados.

2.2 Objetivos específicos

- Apresentar os estudos iniciais;
- Relatar os resultados dos estudos sobre o vocabulário controlado;
- Relatar os resultados na implementação do tesauro.

3 RESULTADOS

Um projeto de pesquisa projeta resultados esperados, que são proposições, visto que se espera obtê-los. Entretanto, conforme se chega aos resultados dos primeiros estudos, esses resultados esperados podem sofrer ajustes, diante do novo cenário. Com isso, o projeto de pesquisa ajusta-se ao novo contexto, de forma a cumprir a meta estabelecida, dando prosseguimentos aos estudos.

3.1 Estudos iniciais

Conforme projetado, os primeiros estudos efetuados na Anvisa tiveram por finalidade conhecer a agência, sua produção técnica e científica, e contaram com o apoio da Gerência de Gestão Documental e Memória Cooperativa (Gedoc), por meio de consultas, especialmente à servidora Tatiana Conceição Caetano, bibliotecária da Anvisa, e à documentação levantada e repassada sobre a atuação da agência.

Um dos documentos orientadores dos estudos foi o “Cadeia de Valor”, desenvolvido pelo Escritório de Processos, em que foram apresentados todos os grandes processos da instituição, seus insumos e produtos, incluindo a documentação gerada. Explicitamente neste documento, em macroprocesso de “Gestão”, processo de “Gestão da Informação, Pesquisa e Conhecimento”, surge a atividade de “Gerir Vocabulário Controlado”, com as tarefas a ela ajustadas, como levantar, cadastrar, publicar e atualizar vocabulário controlado.

Nesse contexto, nota-se a importância dada ao vocabulário controlado, mesmo que não tenha sido definido o tipo - taxonomia, tesauro, ontologia etc. Dessa forma, com base em interações com a Anvisa, verificou-se que o melhor tipo de vocabulário controlado a ser utilizado na pesquisa seria o tesauro, por apresentar os termos hierarquicamente dispostos com a sua definição, ajudando o seu uso.

Em reunião promovida pela Gedoc com outras unidades da Anvisa que tinham interesses no projeto, foi verificado que havia uma equipe externa ao Gedoc

responsável pela gestão de um tesauro da Anvisa. Em vista disso, a equipe do Gedoc ficou responsável por articular as atividades do projeto relacionadas à construção do tesauro com as outras unidades da Anvisa, tendo em vista o cumprimento da Meta 3. Desse modo, entendeu-se que o cumprimento da meta se daria pela disponibilização do sistema de tesauros e criação de documentação técnica e tecnológica para manter esse sistema, que diz respeito à Meta 4, voltada ao repasse de tecnologia.

3.2 Estudos para adaptação de vocabulários controlados existentes, como o DECS

Originalmente, no projeto, a Meta 3 tinha visava à realização de estudos sobre adaptação de vocabulários controlados existentes, como o Descritor da Ciência da Saúde (DeCS). Entretanto, com os estudos iniciais, verificou-se que outra unidade da Anvisa estava com a missão de criação do tesauro da agência, impactando o cumprimento da meta. Em vista disso, o cumprimento dessa meta foi ajustado de forma a não se sobrepor às ações internas da agência.

Coube ao projeto, a implementação de um sistema de tesauros constituído de um acervo inicial de termos, mas com a possibilidade de serem registrados novos termos. Além disso, deveria ser seguida a mesma orientação dada na organização do tesauro, que teve como base a cadeia de valor. Sendo assim, tendo em vista que, dentro do tesauro, os meta-termos são categorias utilizadas para a organização dos termos, considerando a Cadeia de Valor, os meta-termos inicialmente definidos foram:

<ARTICULAÇÃO>
<GESTÃO>
<GOVERNANÇA>
<REGULAMENTAÇÃO>

A partir disso, a Gedoc forneceu uma lita inicial, produzida por outra unidade, constituída de aproximadamente 300 termos candidatos a serem inseridos no sistema de tesauros. Todos os termos passaram pelo processo de análise e aqueles que cumpriam requisitos básicos foram inseridos como termos descritores. Desse modo, cada termo foi inserido dentro de seu respectivo

meta-termo e com suas respectivas relações.

Após essa etapa, a GEDOC apresentou a necessidade de inserir, no tesauro, os termos sobre Vigilância Sanitária existentes dentro do Descritor da Ciência da Saúde (DeCS), visto que esses termos são utilizados pela GEDOC na indexação de documentos. Assim, foi criado o meta-termo <VIGILÂNCIA SANITÁRIA> e dentro dele foram inseridos todos os termos sobre o assunto encontrados no DeCS. Nessa etapa, foram incluídos aproximadamente 700 termos.

Figura 1 - Meta-termos.

Fonte: Elaboração dos autores, 2022.

Depois de inserir os termos do DeCS, iniciou-se a busca e curadoria por novos termos relacionados à temática Vigilância Sanitária em outros materiais, como:

a) Tesauro do Ministério da Saúde;
b) Documento da Cadeia de Valor;
c) Bibliografia sobre Vigilância Sanitária.

Durante essas pesquisas, foram inseridos aproximadamente 100 termos sobre vigilância sanitária. Atualmente, o tesauro possui 979 termos, distribuídos entre os cinco meta-termos e com até dez níveis hierárquicos.

Ao final do projeto, a GEDOC comunicou que as demais gerências, que inicialmente seriam também responsáveis, não estão mais participando do projeto do tesauro, o que fez com que passe a ser responsabilidade apenas da GEDOC. Diante disso, torna-se necessário realizar algumas alterações no projeto para adequação do tesauro às necessidades apenas da GEDOC, alterações essas que vão desde a organização do tesauro (meta-termos) até o escopo dos termos inseridos. Considerando que essa alteração de Gestão aconteceu no final do projeto, não foi possível realizar as modificações no

tesauro.

3.2.1 Implementação do Sistema de Tesauros da Anvisa

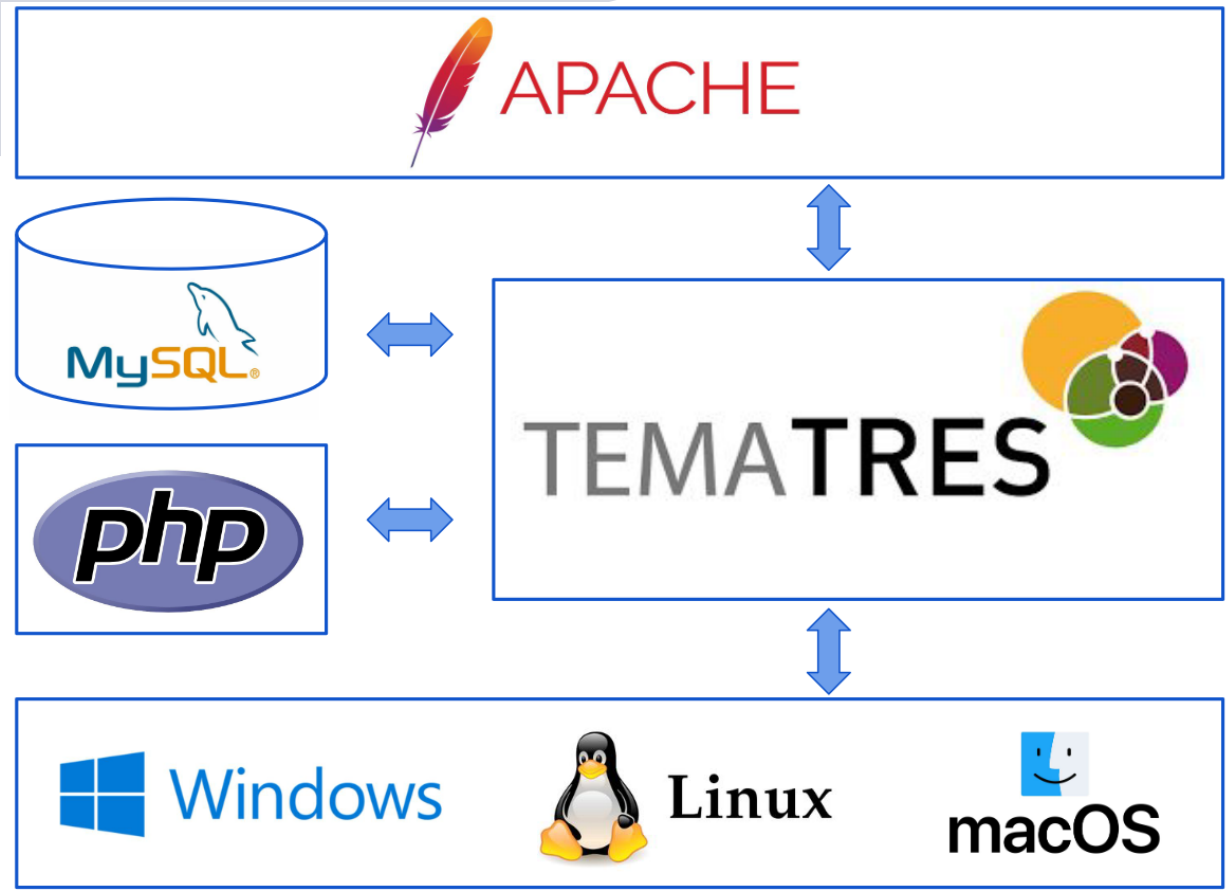
Para a construção do Sistema de Tesauros da Anvisa, o *software* escolhido foi o TemaTres, que é uma ferramenta livre com a finalidade de gerir vocabulários controlados via interface Web. Esse *software* foi escolhido, tendo em vista a facilidade de uso e a forma intuitiva de trabalho. O uso dessa ferramenta tem o propósito de oferecer um sistema informatizado à Anvisa, bem como de apoiar a uniformização da produção e a disseminação de terminologia direcionada às atividades da agência. Assim, busca-se ter um sistema que apoie outras atividades da Anvisa e dissemine a sua terminologia, justificando a escolha do TemaTres, um *software* livre, de código aberto[1] amplamente utilizado no mundo para atuar como vocabulário controlado.

O TemaTres pode ser implantado na maioria dos sistemas operacionais modernos, como Windows, Mac OS e distribuições Linux. Implementado no Sistema de Tesauros da Anvisa, segue a arquitetura LAMP, composto por Sistema Operacional Linux, Servidor de Aplicação Apache, Banco de Dados MySQL e linguagem de programação PHP. Esta arquitetura é amplamente utilizada no desenvolvimento de sistemas web, o que torna a sua adoção uma vantagem competitiva. Além disso, possui certa simplicidade no desenvolvimento.

Também é importante ressaltar que o TemaTres oferece uma interface de web services que permite e facilita a sua integração ou articulação com outras plataformas de gestão ou o desenvolvimento de serviços derivados a partir da exploração de vocabulários controlados. A sua estrutura é composta de uma aplicação desenvolvida em PHP, um servidor Apache, e um banco de dados MySQL. Além disso, trata-se de um sistema web que pode ser acessado por um navegador de internet com suporte HTML5, CSS e JavaScript. A Figura 2 apresenta a arquitetura de sistema padrão do TemaTres:

[1] Código aberto significa que os programas fontes (editáveis) ficam disponíveis para serem baixados e alterados se preciso, contrapondo-se aos softwares de código fechado que disponibilizam apenas os programas executáveis.

Figura 2 - Arquitetura de sistema padrão do TemaTres.



Fonte: Elaboração dos autores, 2022.

Os requisitos utilizados para implementação do TemaTres foram:

- Sistema operacional Linux de 64 bits distribuição Debian GNU/Linux 9;
- Servidor web Apache version: Apache/2.4.25;
- PHP 7.4.16 ou posterior;
- Sistema de gerenciamento de banco de dados (SGBD) MySQL versão 4.1.1 ou posterior.

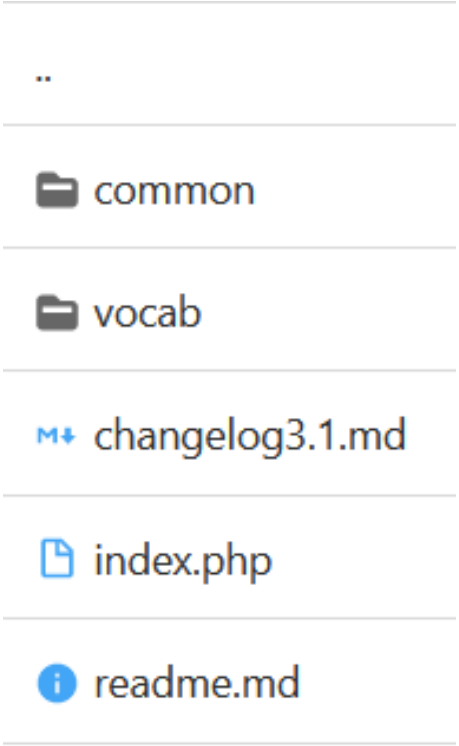
O Sistema de Tesauros da Anvisa implementado com o *software* TemaTres disponível no ambiente do Ibict pode ser acessado através do link http://tesauros.anvisa.ibict.br/. Para informações técnicas da implementação do TemaTres, pode ser consultado o Guia de instalação, configuração e manutenção do TemaTres[2].

[2] COSTA, Mirele Caroline Souza Ferreira; SHINTAKU, Milton. **Guia de instalação, configuração e manutenção TemaTres**: o passo a passo para a instalação, configuração e manutenção do software TemaTres. Brasília: Ibict, 2021.

3.2.2 Implementação do Layout no Sistema de Tesauros

O layout foi implementado para obter a identidade visual do Sistema de Tesauros da Anvisa, todas as modificações foram realizadas na pasta raiz do TemaTres, em HTML e CSS.

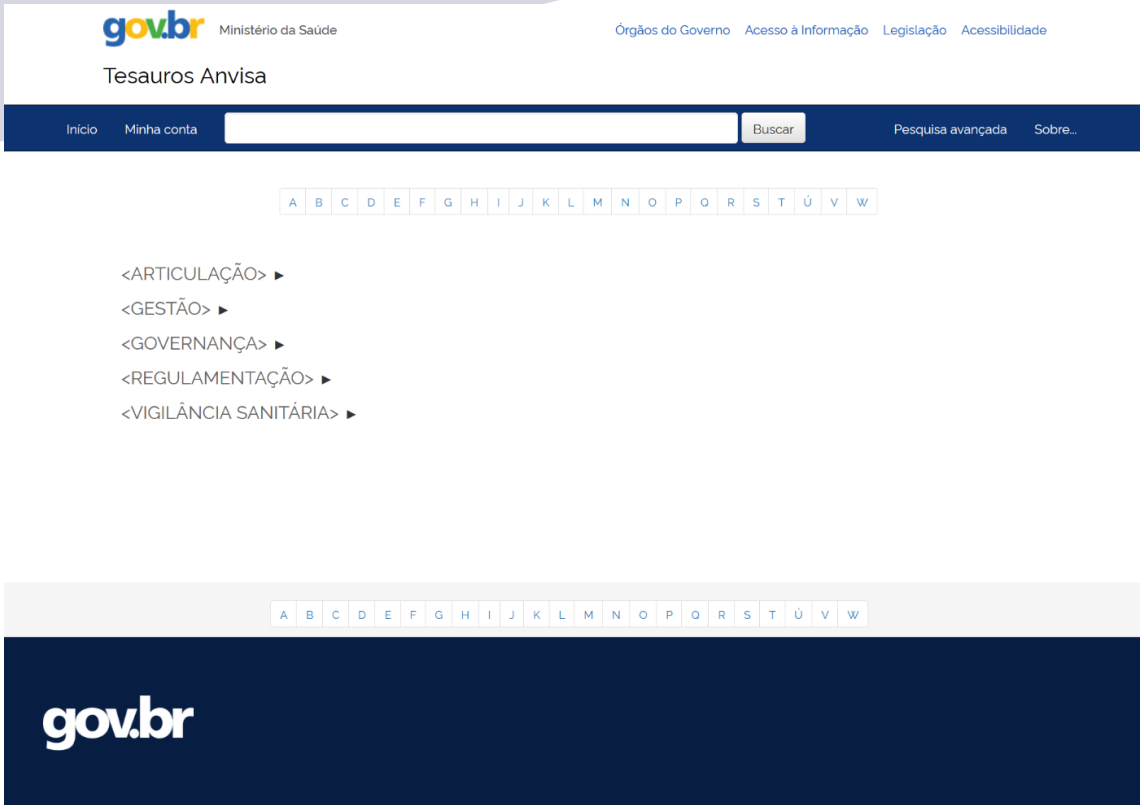
Figura 3 - Pasta raiz do TemaTres.



Fonte: Elaboração dos autores, 2022.

Para a customização do tema no TemaTres, foram inseridos os arquivos de folhas de estilo: “style.css” e “dsgov.min.css”, em CSS no diretório “vocab”. Para a customização do cabeçalho e do rodapé do Tesauros da Anvisa, foram modificados os arquivos “fun.gral.html” e “index.php”, localizados em tematres/common/include/fun.gral.html e tematres/vocab/index.php, respectivamente.

Figura 4 - Tela inicial do Sistema de Tesauros da Anvisa.



Fonte: Elaboração dos autores, 2022.

3.2.3 Migração do tesauros do ambiente computacional do Ibict para a Anvisa

A migração do Sistema de Tesauros para o ambiente de homologação da Anvisa ocorreu a partir da disponibilização dos backups da aplicação e do banco de dados do TemaTres para a equipe de informática da Anvisa no ambiente GIT do Ibict. Além do suporte remoto para apoiar na implantação, foram desenvolvidas documentações técnicas para auxiliar a implantação do software TemaTres:

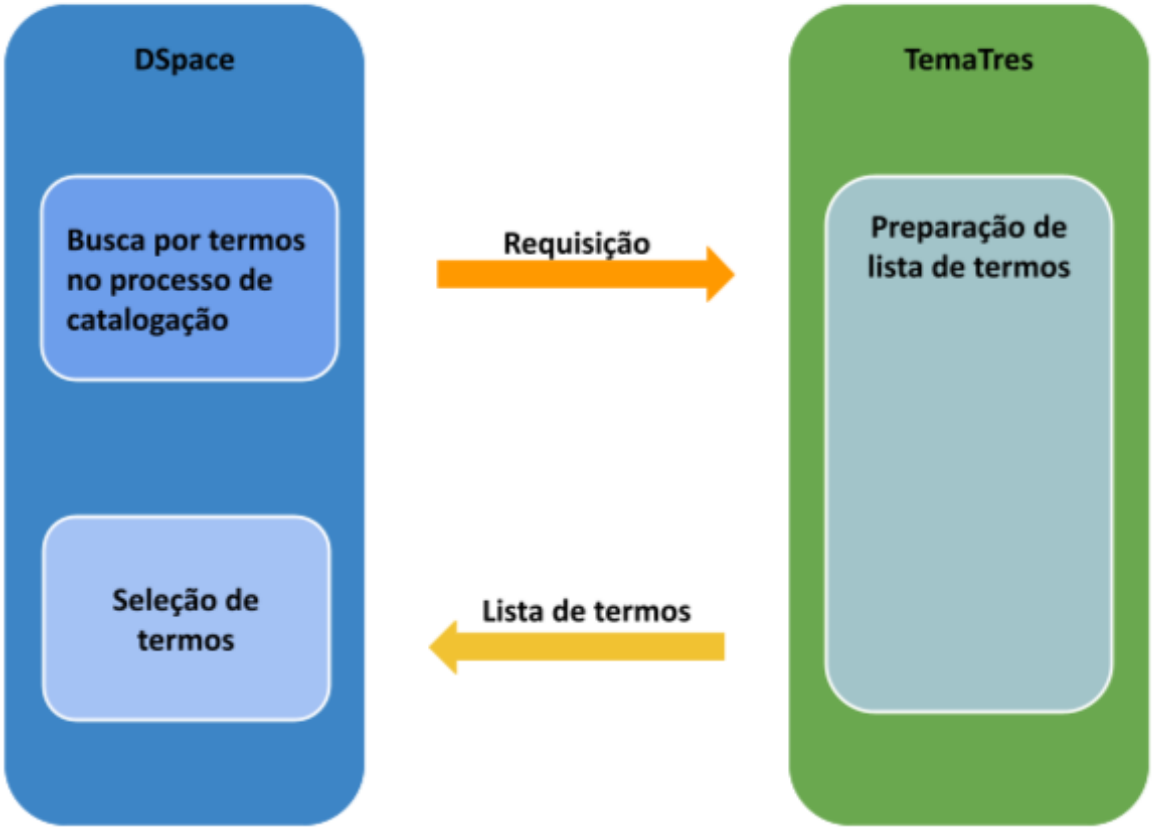
- Guia Tecnológico do software TemaTres: o passo a passo para a instalação, configuração e manutenção[3];
- Relatório Técnico: Passo a Passo para implantação do TemaTres a partir do GIT[4].

[3] COSTA, Mirele Carolina Souza Ferreira; SHINTAKU, Milton. **Guia Tecnológico do software TemaTres**: o passo a passo para a instalação, configuração e manutenção. Brasília: Ibict, 2021.
[4] COSTA, Mirele Carolina Souza Ferreira; SHINTAKU, Milton. **Relatório Técnico**: Passo a Passo para implantação do TemaTres a partir do GIT. Brasília: Ibict, 2021.

3.3 Proposição de adaptação do repositório para uso automático de vocabulário controlado

A integração entre o Sistema de Tesauros (TemaTres) e a Biblioteca Digital da Anvisa (DSpace) tem relação com a padronização dos termos inseridos no metadado Tesauro da Anvisa (*dc.subject.keyword*), por meio de um vocabulário controlado com os termos autorizados. Assim, no processo de submissão de itens – principalmente os pertencentes à memória técnica – o usuário poderá selecionar termos do tesauros de forma automática, diretamente na biblioteca digital, conforme esquema apresentado na Figura 5.

Figura 5 - Estrutura de funcionamento da integração do DSpace com TemaTres.



Fonte: Elaboração dos autores, 2022.

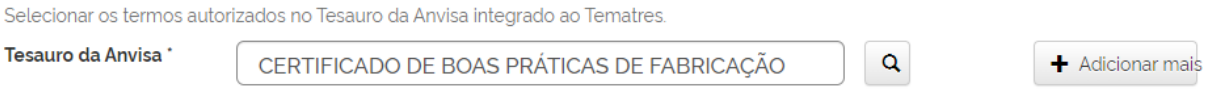
Nesse sentido, os dois sistemas se complementam, na medida em que o TemaTres oferta os termos autorizados para o uso no DSpace, normalizando as informações inseridas no processo de submissão de itens. Assim, o uso da integração é vista apenas no DSpace, no processo de catalogação, mesmo que tenham sido efetuadas alterações nos dois sistemas.

A inserção de termos autorizados apresenta impactos positivos em vários pontos, como na melhoria do processo de busca e descoberta, além de assegurar que a Biblioteca Digital utilize apenas os termos autorizados na

descrição temática dos itens. O uso de base de autoridades é comum em sistemas de bibliotecas tradicionais de objetos físicos, mas ainda com menor uso em bibliotecas digitais.

Com o integrador implantado, todos os termos inseridos no TemaTres estão disponíveis para serem utilizados na Biblioteca Digital da Anvisa (DSpace). Para tanto, o DSpace tem que estar configurado no formulário de entrada. A forma pela qual se deve utilizar o integrador inicia-se com um novo depósito, mais especificamente no processo de descrição, com o preenchimento do formulário de entrada, no campo de palavras-chave, como mostra a Figura 6. O campo de palavras-chave ou assunto geralmente é classificado como obrigatório por muitas bibliotecas digitais, mas isso depende da configuração do formulário de entrada.

Figura 6 - Campo palavras-chave na Biblioteca Digital da Anvisa.

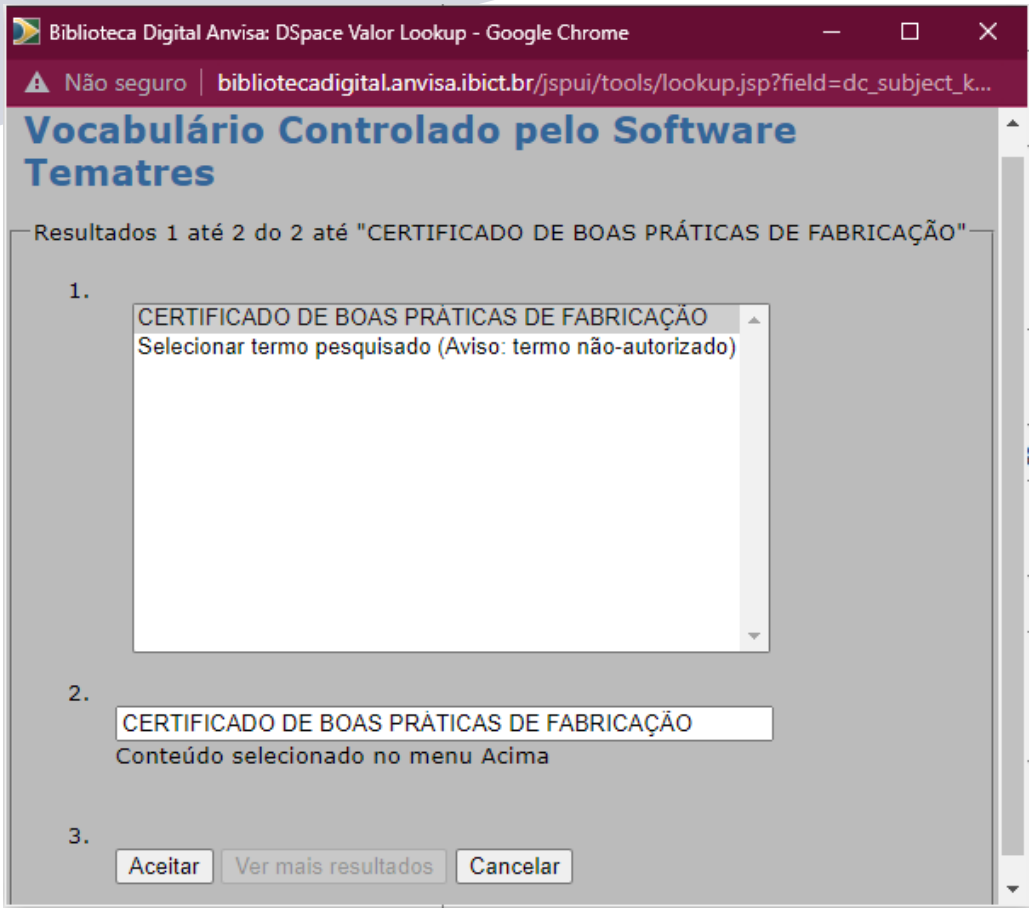


Fonte: Elaboração dos autores, 2022.

Como mostra a Figura 6, a presença do integrador pode ser notada pela existência da lupa ao lado do campo de palavras-chave, inexistente nos formulários-padrão do Dspace. Assim, basta preencher o campo e clicar na lupa para buscar o termo inserido no TemaTres. No exemplo, utilizou-se o termo 'certificado de boas práticas de fabricação', para o integrador buscar na base os termos contidos no Tesauro da Anvisa.

A resposta é exibida em forma de uma janela PopUp, como mostra a Figura 7. A palavra inserida no campo será buscada no TemaTres. Como mostra o exemplo, a busca foi realizada com o termo 'certificado de boas práticas de fabricação' e, assim, na base de termos do TemaTres, foi encontrada. Logo, o usuário pode utilizar o termo clicando em 'Aceitar'. Caso o termo não seja exibido, pode estar cadastrado como não autorizado ou ainda não cadastrado. Porém, na Biblioteca Digital da Anvisa é possível utilizar o termo mesmo não sendo um termo autorizado.

Figura 7 - Resultado da busca no TemaTres, via PopUp da Biblioteca Digital da Anvisa.



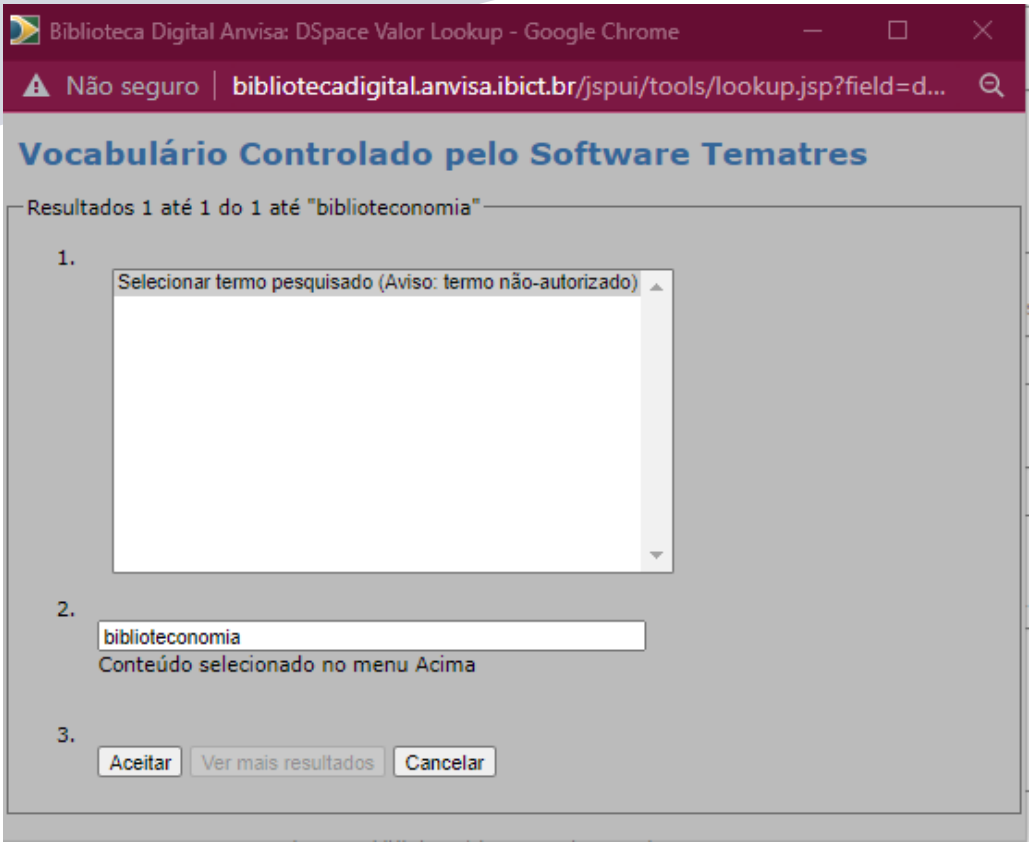
Fonte: Elaboração dos autores, 2022.

Ao clicar em Aceitar, a janela do Tesauro da Anvisa (*Pop-Up*) fechará automaticamente, e o termo selecionado aparecerá no campo do formulário de entrada (Figura 6). Com isso, o termo autorizado preenche o campo. Para adicionar outras palavras-chave na mesma catalogação, basta clicar em Adicionar mais e seguir os passos supracitados. Caso queira desistir do termo, basta limpar o campo e iniciar uma nova busca.

O DSpace permite que sejam inseridos termos não autorizados no processo de submissão, ou seja, termos que não estão presentes no TemaTres. Assim, se oferece maior liberdade aos depositantes na seleção e escolha dos termos mais adequados.

Quando uma busca não resultar em nenhum registro, a tela aparecerá como na Figura 8. Quando isso acontecer, devem ser feitas outras buscas, a fim de que se possam encontrar outras palavras-chave que representem tematicamente o documento, ou então utilizar um termo não autorizado, clicando na opção “Selecionar o termo pesquisado”.

Figura 8 - Termo não-autorizado.



Fonte: Elaboração dos autores, 2022.

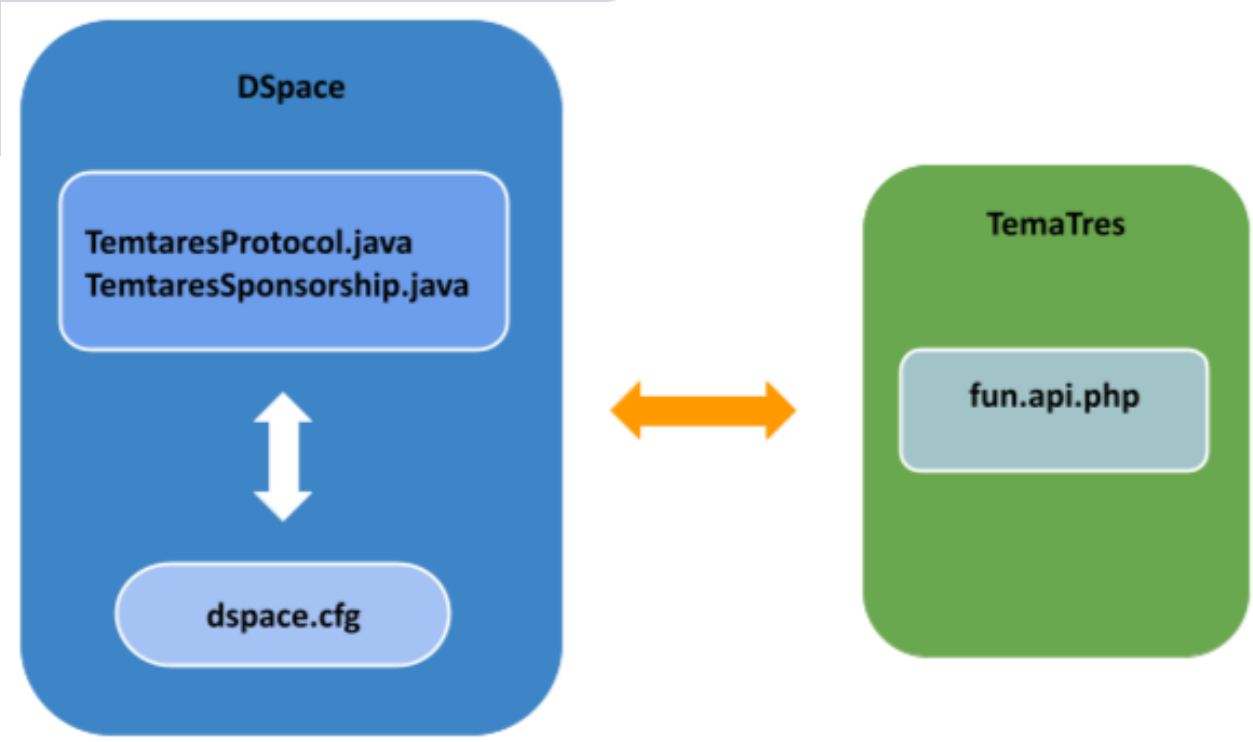
A integração entre o TemaTres e o DSpace visa manter o controle na inserção de termos utilizados na descrição temática dos itens. Dessa forma, o TemaTres cumpre o objetivo de ser um gestor de termos autorizados, o qual pode ser utilizado como banco de autoridade pela Biblioteca Digital, possibilitando o controle de autoridade pessoal e de assunto.

3.3.1 Alteração no TemaTres e DSpace para a Integração

A integração entre ferramentas requer a criação de meios que permitam uma comunicação contínua ou por demanda, executadas principalmente por meio de protocolos de comunicação. No caso, é preciso adicionar um protocolo de comunicação no DSpace e ajustar outro no TemaTres, com vistas a atender à necessidade de comunicação entre os softwares.

De forma simplificada, a Figura 9 apresenta a integração entre as duas ferramentas. Da parte do DSpace, é preciso criar todos os protocolos de comunicação a serem utilizados para a integração, enquanto o TemaTres apenas ajusta o existente. Com isso, torna-se possível ao DSpace a requisição de informações do TemaTres e obter respostas.

Figura 9 - Estrutura do conjunto de classes envolvidas.



Fonte: Elaboração dos autores, 2022.

A integração foi efetuada com o DSpace na versão 6, e o TemaTres, na versão 3. Para tal procedimento, foi necessário intervenção de um informático, e os testes com java 8 – tanto em ambiente de homologação quanto de produção – devem optar por trabalhar com java oracle. Note que a integração é de mão única e o DSpace requisita informações, mas não possibilita que o TemaTres possa requisitar informações do DSpace. Por isso, a maior alteração do DSpace foi a criação dos protocolos de comunicação.

A maior parte das alterações foram efetuadas no DSpace, com a criação de protocolos e sua ativação por meio do arquivo de configuração principal dspace.cfg. Nesse projeto, utilizou-se a versão 6, mais precisamente a última versão estável (6.3), que pode ser baixada no site o Git[5] do DSpace, optando pelas versões src-release. Assim, ao se descompactar por convenção, deverá ser denotado o DSpace-6.3.

Inicialmente, é preciso criar o protocolo responsável por controlar as requisições entre DSpace e TemaTres (TemaTresProtocol.java) no diretório que agrega os outros programas voltados ao controle de autoridade, a fim de manter o padrão do software. Os comandos apresentados seguem o padrão do

[5] Disponível em: https://github.com/DSpace/DSpace/releases/tag/dspace-6.3.

sistema operacional debian/ubuntu.

```
$ cd DSpace-6.3/dspace-api/src/main/java/org/dspace/content/authority
$ echo > TematresProtocol.java
```

Depois de criado o arquivo TemaTresProtocol.java, utilize um editor de preferência e adicione o código do protocolo, como apresentado a seguir. O código, em linguagem java, implementa a comunicação entre os sistemas.

```
package org.dspace.content.authority;

import java.io.IOException;
import java.util.List;
import javax.xml.parsers.SAXParser;
import javax.xml.parsers.SAXParserFactory;
import javax.xml.parsers.ParserConfigurationException;
import org.xml.sax.XMLReader;
import org.xml.sax.InputSource;
import org.xml.sax.helpers.DefaultHandler;
import org.xml.sax.Attributes;
import org.xml.sax.SAXException;
import org.xml.sax.SAXParseException;
import org.apache.log4j.Logger;
import org.dspace.content.Collection;
import org.dspace.core.ConfigurationManager;
import org.apache.http.HttpResponse;
import org.apache.http.client.HttpClient;
import org.apache.http.client.methods.HttpGet;
import org.apache.http.client.utils.URLEncodedUtils;
import org.apache.http.impl.client.DefaultHttpClient;
import org.apache.http.message.BasicNameValuePair;

public abstract class TemaTresProtocol implements ChoiceAuthority
{
   private static Logger log = Logger.getLogger(TemaTresProtocol.class);

   private static String url = null;
```

```
public TemaTresProtocol()
  {
    if (url == null)
    {
      url = ConfigurationManager.getProperty("TemaTres.url");
      if (url == null)
      {
            throw new IllegalStateException("Missing DSpace configuration
keys for TemaTres Query");
      }
    }
  }

  public abstract Choices getMatches(String text, Collection collection, int
start, int limit, String locale);

  @Override
      public Choices getBestMatch(String field, String text, Collection
collection, String locale)
  {
    return getMatches(field, text, collection, 0, 2, locale);
  }

  @Override
  public String getLabel(String field, String key, String locale)
  {
    return key;
  }

  protected Choices query(String result, String label, String authority,
              List<BasicNameValuePair> args, int start, int limit)
  {
    HttpClient hc = new DefaultHttpClient();
    String srUrl = url + "?" + URLEncodedUtils.format(args, "UTF8");
    HttpGet get = new HttpGet(srUrl);
```

```
      log.debug("Trying TemaTres Query, URL="+srUrl);
      try
      {
        HttpResponse response = hc.execute(get);
        if (response.getStatusLine().getStatusCode() == 200)
        {
          SAXParserFactory spf = SAXParserFactory.newInstance();
          SAXParser sp = spf.newSAXParser();
          XMLReader xr = sp.getXMLReader();
               TemaTresHandler handler = new TemaTresHandler(result, label,
authority);
          xr.setFeature("http://xml.org/sax/features/namespaces", true);
          xr.setContentHandler(handler);
          xr.setErrorHandler(handler);
 xr.parse(new InputSource(response.getEntity().getContent()));
          int confidence;
          if (handler.total == 0)
          {
            confidence = Choices.CF_NOTFOUND;
          }
          else if (handler.total == 1)
          {
            confidence = Choices.CF_UNCERTAIN;
          }
          else
          {
            confidence = Choices.CF_AMBIGUOUS;
          }
           return new Choices(handler.result, start, handler.total, confidence,
false);
        }
      }
      catch (IOException e)
      {
        log.error("TemaTres query failed: ", e);
        return null;
```

```
        }
        catch (ParserConfigurationException e)
        {
            log.warn("Failed parsing TemaTres result: ", e);
            return null;
        }
        catch (SAXException e)
        {
            log.warn("Failed parsing TemaTres result: ", e);
            return null;
        }
        finally
        {
            get.releaseConnection();
        }
        return null;
    }

    private static class TemaTresHandler
        extends DefaultHandler
    {
        private Choice result[] = null;
        int rindex = 0;
        int total = 0;
        private String resultElement = null;
        private String labelElement = null;
private String authorityElement = null;
        protected String textValue = null;
        public TemaTresHandler(String result, String label, String authority)
        {
            super();
            resultElement = result;
            labelElement = label;
            authorityElement = authority;
        }
```

```
    @Override
    public void characters(char[] ch, int start, int length)
      throws SAXException
    {
      String newValue = new String(ch, start, length);
      if (newValue.length() > 0)
      {
        if (textValue == null)
        {
          textValue = newValue;
        }
        else
        {
          textValue += newValue;
        }
      }
    }

    @Override
     public void endElement(String namespaceURI, String localName,String
qName)
      throws SAXException
    {
      if (localName.equals("cant_result"))
      {
        String stotal = textValue.trim();
        if (stotal.length() > 0)
        {
          total = Integer.parseInt(stotal);
          result = new Choice[total];
          if (total > 0)
          {
            result[0] = new Choice();
            log.debug("Got "+total+" records in results.");
          }
        }
```

```
        }
        else if (localName.equals(resultElement))
{
          // after start of result element, get next hit ready
          if (++rindex < result.length)
          {
            result[rindex] = new Choice();
          }
        }
        else if (localName.equals(labelElement) && textValue != null)
        {
          // plug in label value
          result[rindex].value = textValue.trim();
          result[rindex].label = result[rindex].value;
        }
                                    else if (authorityElement != null &&
localName.equals(authorityElement) && textValue != null)
        {
          result[rindex].authority = textValue.trim();
        }
        else if (localName.equals("message") && textValue != null)
        {
          log.warn("TemaTres response error message: " + textValue.trim());
        }
      }

      @Override
      public void startElement(String namespaceURI, String localName,
                  String qName, Attributes atts)
        throws SAXException
      {
        textValue = null;
      }

      @Override
      public void error(SAXParseException exception)
```

```
        throws SAXException
    {
        throw new SAXException(exception);
    }

    @Override
    public void fatalError(SAXParseException exception)
        throws SAXException
    {
        throw new SAXException(exception);
    }
  }
}
```

Após a criação do arquivo TemaTresProtocol.java, será preciso criar o arquivo TemaTresSponsorship.java, que será responsável pelo processamento das requisições no DSpace. Para tanto, execute os seguintes comandos:

```
$ cd DSpace-6.3/dspace-api/src/main/java/org/dspace/content/authority
$ echo > TemaTresSponsorship.java
```

Utilizando um editor de preferência, adicione o seguinte código java:

```
package org.dspace.content.authority;
import java.util.ArrayList;
import java.util.List;
import org.dspace.content.Collection;
import org.apache.http.message.BasicNameValuePair;
public class TemaTresSponsorship extends TemaTresProtocol{
  private static final String RESULT = "term";
  private static final String LABEL = "string";
  private static final String AUTHORITY = "term_id";

  public TemaTresSponsorship()
  {
    super();
  }
```

```
@Override
  public Choices getMatches(String text, Collection collection, int start, int
limit, String locale)
  {
    if (text == null || text.trim().length() == 0)
    {
      return new Choices(true);
    }
                                  List<BasicNameValuePair>  args  =  new
ArrayList<BasicNameValuePair>();
    args.add(new BasicNameValuePair("arg", text));
    args.add(new BasicNameValuePair("task","search"));

    Choices result = query(RESULT, LABEL, AUTHORITY, args, start, limit);
    if (result == null)
    {
      result = new Choices(true);
    }
    return result;
  }
@Override
  public Choices getMatches(String field, String text, Collection collection,
int start, int limit, String locale) {
    return getMatches(text, collection, start, limit, locale);
  }
}
```

Após a criação dos arquivos java que processam a comunicação com o TemaTres, é necessário configurar o dspace.cfg para utilizar o serviço.

```
$ cd DSpace-6.3/dspace/target/dspace-installer/config
```

Então, no arquivo dspace.cfg, deve-se descomentar a linha 1429, que ativa o uso de controle de autoridade.

```
plugin.named.org.dspace.content.authority.ChoiceAuthority = \
```

Na sequência, deve-se adicionar o seguinte trecho após a linha 1429, para adicionar a opção do uso do TemaTres.

```
org.dspace.content.authority.TemaTresSponsorship = TemaTresSponsorship
```

Por fim, tem-se a seção de controle de autoridade, com alguns PlugIns comentados, como o do Library of Congress (LCNameAuthority), mas com o TemaTres ativo.

```
##### Authority Control Settings #####
plugin.named.org.dspace.content.authority.ChoiceAuthority = \
 org.dspace.content.authority.SampleAuthority = Sample, \
 org.dspace.content.authority.LCNameAuthority = LCNameAuthority, \
 org.dspace.content.authority.SHERPARoMEOPublisher = SRPublisher, \
 org.dspace.content.authority.SHERPARoMEOJournalTitle = SRJournalTitle, \
        org.dspace.content.authority.TemaTresSponsorship = TemaTresSponsorship
```

Após finalizar as customizações efetuadas no TemaTres, é necessário configurar o dspace.cfg com o propósito de o DSpace consultar, de forma assíncrona, a base de dados do TemaTres. Por questão de padrão, após as linhas que fazem referência ao plugin TemaTres, adicione as seguintes linhas:

```
TemaTres.url = https://IPTemaTres/vocab/services.php
choices.plugin.dc.subject.keyword = TemaTresSponsorship
choices.presentation.dc.subject.keyword = lookup
authority.controlled.dc.subject.keyword = true
```

Observação: “dc.subject.keyword” está informando o metadado no input-forms.xml, que mapeia as palavras-chaves.

O DSpace já possui alguns PlugIns para implementação de controle de autoridade, com o uso de vocabulários controlados, conforme visto no arquivo de configuração. Com isso, adiciona-se o TemaTres, seguindo o padrão utilizado na ferramenta, facilitando ajustes em caso de atualização de versão do DSpace.

As alterações no TemaTres são em menor número, visto que ele já possui protocolos de comunicação que serão utilizados nesta integração (services.php e fun.api.php). Entretanto, pela necessidade apresentada no projeto, se requer a quantidade de registros em cada requisição efetuada pelo DSpace. Assim, considera-se que o TemaTres esteja instalado em /var/www/html/TemaTres. Ao acessar o arquivo common/include/fun.api.php, na linha 1055, deve-se adicionar o trecho de código abaixo:

```
$arrayResume[“cant_result”] = count($response["result"]);
```

Em seguida, deve-se salvar e verificar se os termos estão sendo renderizados no navegador. Nesses casos, a resposta à requisição deve ser a seguinte: https://IPTemaTres/vocab/services.php?task=letter&arg=a. Note que o argumento de busca é a letra “a”, indicada pela “task=letter” e “arg=a”. Como resposta, o sistema responde quantos termos da base iniciam com a letra “a” nas primeiras linhas, segundo é mostrado abaixo:

```
<vocabularyservices>
<cant_result>X</cant_result>
```

4 CONSIDERAÇÕES FINAIS

Como apresentado pelo relatório, grande parte das atividades relacionadas ao Vocabulário Controlado da Anvisa foram executadas. Porém, no decorrer do projeto, a GEDOC informou que as demais gerências não participam mais na construção, desenvolvimento e uso do Tesauro. Diante disso, o Tesauro precisará passar por alterações na sua estrutura e nos termos de forma a se adequar às necessidades da GEDOC.

Nesse sentido, o Ibict requer aditamento, a fim de prosseguir com as pesquisas, gerando novas atividades. De forma genérica, tem-se duas grandes atividades, voltadas para a criação das políticas de tesauros e da capacitação. Tais demandas foram resultados dos estudos que verificaram as necessidades específicas, tanto teóricas quanto aplicadas.

MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÕES

PÁTRIA AMADA
BRASIL
GOVERNO FEDERAL