

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

FELIPE DE MELLO FLORES

**ACERPI-Query: Consultas sobre uma base
de dados de documentos de portarias
públicas**

Monografia apresentada como requisito parcial
para a obtenção do grau de Bacharel em Ciência
da Computação

Orientador: Prof^a. Dr^a. Renata Galante
Co-orientador: Christian Schmitz

Porto Alegre
2022

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos André Bulhões Mendes

Vice-Reitora: Prof^ª. Patricia Helena Lucas Pranke

Pró-Reitor de Graduação: Prof^ª. Cíntia Inês Boll

Diretora do Instituto de Informática: Prof^ª. Carla Maria Dal Sasso Freitas

Coordenador do Curso de Ciência de Computação: Prof. Rodrigo Machado

Bibliotecária-chefe do Instituto de Informática: Alexsander Borges Ribeiro

AGRADECIMENTOS

Gostaria de começar agradecendo à professora Renata de Matos Galante, orientadora deste trabalho, com sua ajuda e orientação para encontrar uma proposta de trabalho interessante e também durante todo o seu desenvolvimento, também ao Christian Schmitz e ao professor Edimar Manica, com suas sugestões, dicas e apoio na implementação do trabalho. Estendo o agradecimento também a todos os professores do Instituto de Informática que fizeram parte da minha formação e que fizeram eu me apaixonar pela computação.

Agradeço também a todos os meus amigos de Porto Alegre e Montenegro que durante todo este ano estiveram comigo celebrando a vida sempre que possível. Estes momentos de descontração e alegria foram essenciais para aliviar o estresse de fim de curso e são o que fazem a vida ter mais sentido. A vida é muito melhor quando compartilhada com vocês.

Agradeço de coração ao meu irmão, cunhada, e a todos os meus primos, primas, tias e tios. É impossível colocar em palavras o quanto vocês são importantes para mim, mesmo com a distância e poucos encontros ao ano o carinho e o amor por vocês nunca diminui. Cada um de vocês contribuiu na minha jornada de alguma maneira e levo sempre um pedaço de vocês no meu peito. Vocês são minha família, amigos, companheiros, professores, e a base da minha vida.

Por último e mais importante, agradeço a minha mãe Rossana e meu pai Fernando. A minha vida toda vocês me deram carinho, amor, apoio e educação para eu me tornar quem eu sou hoje. Vocês me ensinaram valores de amor e compaixão que me moldaram e que sigo todos os dias, e que sempre busco passar adiante para o mundo. Vocês me prepararam, me encorajaram, me orientaram e me deram um empurrão para que eu pudesse trilhar o meu próprio caminho, vocês serão sempre minhas maiores referências e meus maiores professores, e por isso vocês têm minha eterna Gratidão e meu eterno Amor.

Muito obrigado.

RESUMO

Instituições Federais emitem portarias que contém ordens ou instruções para aplicação de regulamentos, disseminação de informações, aposentadorias, promoções, substituições, entre outras ações de caráter administrativo. Estas portarias estão disponíveis publicamente; entretanto, geralmente estes documentos estão disponíveis apenas em PDF e sem permitir consultas, filtragem ou buscas mais avançadas. A abordagem ACERPI (Abordagem para Coleta de documentos, Extração de informação e Resolução de entidades em Portarias Institucionais) coleta as portarias e constrói um banco de dados orientado a documentos buscando resolver este problema de identificar entidades (nomes de pessoas) em portarias públicas. Este trabalho propõe o **ACERPI-Query** que tem o objetivo principal de propor consultas sobre a base de dados de portarias públicas, com o intuito de expandir os metadados e conseqüentemente a base de dados original. Inicialmente, este trabalho propõe uma série de extrações de portarias públicas. Usando expressões regulares, este trabalho expande a etapa de extração identificando portarias de progressão funcional, aposentadoria e a entidade que assina as portarias. Por fim, o trabalho propõe diversas consultas possíveis sobre o banco de dados resultante, facilitando consultas e filtros mais avançados sobre a base de dados de portarias.

Palavras-chave: Coleta de documentos. extração de informações. consultas NoSQL.

ACERPI-Query: Queries on a document oriented database of public ordinances

ABSTRACT

Federal institutions issue ordinances that contain orders or instructions for the application of regulations, dissemination of information, retirements, promotions, substitutions, among other administrative actions. These ordinances are publicly available; however, these documents are generally available only in PDF and do not allow for more advanced queries, filtering or searches. The ACERPI approach (Approach to Collecting Documents, Extraction of Information and Resolution of Entities in Institutional Ordinances) collects the ordinances and builds a document-oriented database seeking to solve this problem of identifying entities (names of persons) in public ordinances. This work proposes the **ACERPI-Query** which has the main objective of proposing queries on the database of public ordinances, in order to expand the metadata and consequently the original database, this work, initially, proposes a series of extractions of public ordinances. Using regular expressions, this work expands the extraction step by identifying ordinances for functional progression, retirement and the entity that signs the ordinances. Finally, the work proposes several possible queries on the resulting database, facilitating more advanced queries and filters in the ordinances.

Keywords: Documents retrieval, information extraction, NoSQL queries.

LISTA DE FIGURAS

Figura 3.1 Portaria Exemplo	16
Figura 3.2 Visão Geral ACERPI	17
Figura 3.3 Portaria Exemplo 2	18
Figura 4.1 Portaria de Progressão	26
Figura 4.2 Portaria de Promoção.....	27
Figura 4.3 Portaria de Progressão Servidor	31
Figura 4.4 Portaria de Aposentadoria	34

LISTA DE TABELAS

Tabela 2.1	Tabela de símbolos da sintaxe de expressões regulares.....	14
Tabela 4.1	Tabela de Progressões e Promoções - Magistério Superior.....	25
Tabela 5.1	Tabela de Progressões por Mês e Ano.....	47
Tabela 5.2	Tabela de Aposentadorias por Mês e Ano	48

LISTA DE ABREVIATURAS E SIGLAS

- PDF *Portable Document Format* - Formato Portátil de Documento
- XML *Extensible Markup Language* - Linguagem de Marcação Extensível
- ACERPI Abordagem para Coleta de documentos, Extração de informação e Resolução de entidades em Portarias Institucionais

SUMÁRIO

1 INTRODUÇÃO	10
2 CONCEITOS E TECNOLOGIAS UTILIZADAS	11
2.1 NoSQL.....	11
2.1.1 MongoDB	11
2.2 XML	12
2.3 Expressões Regulares.....	13
2.4 Considerações Finais	14
3 TRABALHOS PRÉVIOS E TRABALHOS RELACIONADOS.....	15
3.1 Trabalhos Prévios.....	15
3.1.1 ACERPI	15
3.1.1.1 Coleta	15
3.1.1.2 Extração	17
3.1.1.3 Resolução de Entidades	18
3.1.2 Considerações	19
3.2 Trabalhos Relacionados.....	19
3.3 Considerações Finais	20
4 EXTRAÇÕES.....	21
4.1 Modelo de dados.....	21
4.2 Extração da Assinatura	23
4.3 Extração da Progressão e Promoção de Professores.....	24
4.4 Extração da Progressão de Servidores Técnico-Administrativos.....	29
4.5 Extração da Aposentadoria.....	33
4.6 Considerações Finais	36
5 CONSULTAS.....	37
5.1 Consultas Básicas de Entidades e Portarias	37
5.2 Consultas Por Assinaturas	40
5.3 Consultas Por Progressões e Aposentadorias	42
5.4 Histórico de Professores e Servidores	48
5.5 Considerações Finais	50
6 CONCLUSÃO	51
REFERÊNCIAS.....	52

1 INTRODUÇÃO

Instituições Federais emitem Portarias que contém ordens ou instruções para aplicação de regulamentos, disseminação de informações, nomeações, exonerações, aposentadorias, promoções, substituições, entre outras ações de caráter administrativo em relação aos servidores. Estas Portarias, por serem de interesse público, estão disponíveis publicamente em sites e repositórios das Instituições. Entretanto, geralmente estes documentos estão disponíveis apenas em PDF e sem permitir consultas, filtragem ou buscas mais avançadas. Além disso, cada Instituição tem o seu formato próprio para disponibilizar as portarias.

A abordagem ACERPI (**Abordagem** para **Coleta** de documentos, **Extração** de informação e **Resolução** de entidades em **Portarias Institucionais**)(SCHMITZ et al., 2021) coleta as portarias e constrói um banco de dados orientado a documentos buscando resolver o problema de identificar entidades do tipo pessoa. A abordagem coleta as portarias e extrai de seus textos as entidades mencionadas, construindo por fim um banco de dados com as entidades e em quais portarias são mencionadas, junto com alguns outros metadados extraídos.

Neste trabalho, é proposto o ACERPI-Query que tem como objetivo explorar o banco de dados resultante da abordagem ACERPI, propondo consultas possíveis a fim de extrair valor sobre os dados. Além disso, a fim de permitir consultas mais avançadas, foi expandida a quantidade de informações extraídas das portarias, incluindo no banco de dados as próprias portarias junto com metadados do que a portaria se refere. Por exemplo, identificando que a portaria se refere a aposentadoria e qual o nome do servidor se aposentando e a data.

O restante do trabalho tem a seguinte organização: o Capítulo 2 apresenta os conceitos e tecnologias utilizadas ao longo do trabalho; o Capítulo 3, descreve a abordagem ACERPI e outros trabalhos prévios, e em seguida são apresentados trabalhos relacionados; no Capítulo 4 são apresentadas as expansões feitas na extração de informações das portarias; o Capítulo 5, demonstra diversas consultas possíveis sobre o banco de dados com o objetivo de extrair valor dos dados armazenados; e, por fim, o Capítulo 6 conclui e revisa as contribuições do trabalho, além de propor possíveis futuros trabalhos.

2 CONCEITOS E TECNOLOGIAS UTILIZADAS

Neste capítulo, são introduzidos os conceitos e tecnologias utilizados ao longo do trabalho. São mostradas as tecnologias que foram usadas para extração e armazenamentos dos dados.

2.1 NoSQL

Bancos de dados *NoSQL* são bases de dados não tabulares e armazenam diferente de tabelas relacionais. Possuem esquemas flexíveis e escalam facilmente com grandes quantidades de dados. Existem 4 principais tipos (MONGODB, 2020c):

- Orientado a documento: Documentos contêm pares de campos e valores. Estes valores podem ser uma variedade de tipos, incluindo *strings*, números, *booleanos*, *arrays*, ou objetos.
- Chave-valor: Um tipo mais simples de banco de dados onde cada item contém chaves e valores.
- Orientado a coluna: Armazena dados em tabelas, linhas e colunas dinâmicas.
- Orientado a Grafos: Armazenam dados em nodos e arestas. Nodos normalmente representando entidades e as arestas representando suas relações.

Neste trabalho, foi utilizado *NoSQL* orientado a documentos para o armazenamento de todos os dados.

2.1.1 MongoDB

MongoDB (MONGODB, 2020a) é um banco de dados *NoSQL* baseado em documentos (MONGODB, 2020b) estruturados em uma maneira similar à estrutura JSON (*JavaScript Object Notation*). Neste trabalho, MongoDB é utilizado para o armazenamento de todos os dados, e também para realizar todas as consultas em cima desses dados

MongoDB armazena registros como documentos, que são agrupados em coleções. Uma base de dados é formada de uma ou mais coleções (MONGODB, 2022b). Estes conceitos foram usadas no trabalho para armazenar os diferentes tipos de documentos em cada coleção, como a coleção de Portarias e a coleção de Entidades, formando a

base de dados a ser trabalhada. Documentos são armazenados em formato BSON (uma representação binária de JSON) e são compostos por pares de chave-valor (MONGODB, 2022c).

Consultas no MongoDB são feitas através de *pipelines* de agregação, compostas de estágios que processam os documentos. Cada estágio possui uma operação, alguns estágios utilizados nesse trabalho são: *\$match*, que filtra os documentos baseado em alguma condição; *\$lookup*, que performa uma junção externa esquerda com outra coleção; *\$project*, que remodela os documentos, adicionando ou removendo campos; *\$group*, que agrupa os documentos baseado em uma expressão e o *\$sortByCount*, que além de agrupar, também ordena todos os documentos (MONGODB, 2022a). Um exemplo de agregação pode ser vista na Listagem 2.1 em uma base de dados de pedidos de pizza.

Listagem 2.1 – Exemplo de Consulta no MongoDB

```
1 [
2   // Stage 1: Filtra os pedidos de pizza pelo tamanho.
3   {
4     $match: { size: "medium" }
5   },
6   // Stage 2: Agrupa os documentos resultantes pelo
7   // nome da pizza e calcula a quantidade total.
8   {
9     $group: { _id: "$name", totalQuantity: { $sum: "
10    $quantity" } }
11 }
```

2.2 XML

XML é uma linguagem de marcação para criação de documentos organizados de forma hierárquica. É um dos subtipos da SGML (*Standard Generalized Markup Language*) capaz de descrever diversos tipos de dados (W3C, 2008).

XML foi utilizado nesse trabalho como formato intermediário de armazenamento dos dados antes de carregá-los para o banco de dados. Na Listagem 2.2, é apresentado um exemplo de um XML para descrever uma receita de pão.

Listagem 2.2 – Exemplo de um XML para descrever uma receita de pão

```

1 <receita nome="pão" tempo_de_preparo="5 minutos" tempo_de_cozimento
  ="1 hora">
2 <titulo>Pão simples</titulo>
3 <ingredientes>
4 <ingrediente quantidade="3" unidade="xícaras">Farinha de Trigo<
  /ingrediente>
5 <ingrediente quantidade="7" unidade="gramas">Fermento</
  ingrediente>
6 <ingrediente quantidade="1.5" unidade="xícaras" estado="morna">
  Água</ingrediente>
7 <ingrediente quantidade="1" unidade="colheres de chá">Sal</
  ingrediente>
8 </ingredientes>
9 <instrucoes>
10 <passo>Misture todos os ingredientes , e dissolva bem.</passo>
11 <passo>Cubra com um pano e deixe por uma hora em um local morno
  .</passo>
12 <passo>Misture novamente , coloque numa bandeja e asse num forno
  .</passo>
13 </instrucoes>
14 </receita>

```

2.3 Expressões Regulares

Expressão Regular é uma sequência de caracteres que especifica um padrão de busca em um texto, permitindo também operações de captura e substituição. Existem diferentes sintaxes para escrever expressões regulares, as mais conhecidas são *POSIX* e *Perl*.

Neste trabalho, utilizamos a biblioteca *re* da linguagem Python, baseada na sintaxe *Perl*, para, utilizando expressões regulares, buscar e extrair dados dos textos das portarias (PYTHON, 2022). A Tabela 2.1 descreve todos os operadores usados neste trabalho.

Tabela 2.1 – Tabela de símbolos da sintaxe de expressões regulares

Meta-caractere	Descrição
\s	Espaço em branco
.	Qualquer carácter
[]	Conjunto de caracteres
+	Uma ou mais repetições
*	Zero ou mais repetições
?	Zero ou uma vez
	Casa ou com a expressão anterior ou com o posterior
()	Captura caracteres

2.4 Considerações Finais

Neste capítulo apresentamos conceitos e tecnologias utilizados ao longo do trabalho que facilitaram e possibilitaram sua implementação. Foi apresentado o conceito de *NoSQL* e o banco MongoDB, que foram utilizados para armazenamento de todos os dados extraídos. Também foi introduzido o XML, uma linguagem de marcação que foi utilizada como intermediária e facilitação no momento de extrair os dados e carregá-los no banco. Por último, detalhamos o que são expressões regulares e como podemos utilizá-las para buscar e extrair padrões das portarias.

3 TRABALHOS PRÉVIOS E TRABALHOS RELACIONADOS

Este capítulo apresenta os trabalhos prévios desenvolvidos a respeito da abordagem ACERPI Schmitz et al. (2021). Logo após, são descritos alguns trabalhos relacionados.

3.1 Trabalhos Prévios

Esta seção descreve o ACERPI e suas extensões, estes trabalhos serviram de base para a extensão proposta neste trabalho.

3.1.1 ACERPI

A abordagem ACERPI (SCHMITZ et al., 2021) busca criar um banco de dados não relacional orientado a documentos a partir de documentos publicados por Instituições Federais. O objetivo é permitir uma maior facilidade na busca e filtragem destes documentos, que geralmente não permitem nenhuma consulta avançada.

O ACERPI tem como entrada um conjunto de documentos de Portarias, como o repositório da UFRGS (UFRGS, 2016), que contém documentos em PDF como o ilustrado na Figura 3.1. Como saída, é construído um banco de dados com as entidades mencionadas nas Portarias e detalhes dos documentos.

A abordagem ACERPI é dividida em 3 etapas: coleta, extração e resolução de entidades. A base de dados final pode ser utilizada para obter informações de um servidor, os documentos que o mencionam e os principais metadados extraídos. A Figura 3.2 demonstra a visão geral de todas as etapas.

3.1.1.1 Coleta

A etapa de coleta inclui a aquisição das portarias das Instituições a partir de técnicas de *Web Scraping*. As portarias são baixadas dos repositórios da Instituições e, após isso, são transformadas em formato de texto. Em seguida, estes textos são interpretados

Figura 3.1 – Portaria Exemplo



SERVIÇO PÚBLICO FEDERAL

PORTARIA Nº 10403 de 13/11/2017

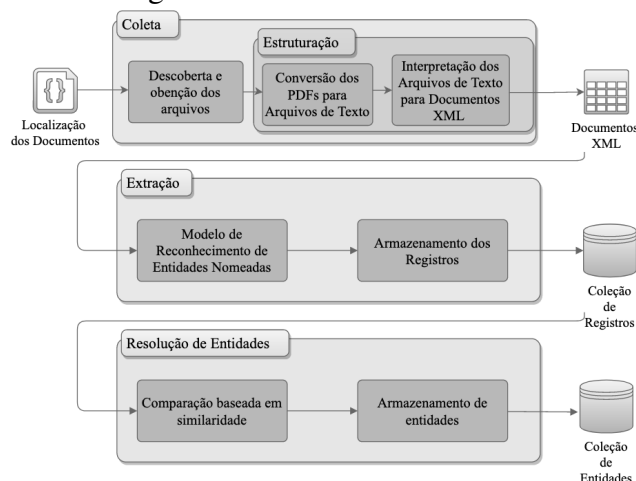
A PRÓ-REITORA DE GESTÃO DE PESSOAS DA UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL EM EXERCÍCIO, no uso de suas atribuições que lhe foram conferidas pela Portaria nº. 8117, de 10 de outubro de 2016, e conforme a Solicitação de Afastamento nº32907,

RESOLVE

Designar, temporariamente, nos termos da Lei nº. 8.112, de 11 de dezembro de 1990, com redação dada pela Lei nº.9.527, de 10 de dezembro de 1997, a ocupante do cargo de PROFESSOR DO MAGISTÉRIO SUPERIOR, do Quadro de Pessoal desta Universidade, **RENATA DE MATOS GALANTE** (Siape: 1488770), para substituir **CARLA MARIA DAL SASSO FREITAS** (Siape: 0351477), Diretor do Instituto de Informática, Código CD-3, em seu afastamento no país, no período de 14/11/2017 a 15/11/2017, com o decorrente pagamento das vantagens por 2 dias.

VÂNIA CRISTINA SANTOS PEREIRA
Pró-Reitora

Figura 3.2 – Visão Geral ACERPI



usando expressões regulares e convertidas para um formato XML, não apenas o texto como também algumas informações são estruturadas, como data e id da portaria.

3.1.1.2 Extração

O objetivo desta etapa é identificar os servidores mencionados na portaria e os metadados ligados a eles. Isto é alcançado através de técnicas de *Named Entity Recognition* e *Transfer Learning*.

Primeiro acontece a extração de nomes de servidores e suas matrículas SIAPE utilizando expressões regulares e processamento de linguagem natural. Logo após, essas informações são armazenadas no MongoDB em uma coleção de registros. Cada documento tem o formato como ilustrado na Listagem 3.1.

Listagem 3.1 – Exemplo de um registro no banco de dados

```

1 {
2   "id": 4630,
3   "name": 'RENATA DE MATOS GALANTE',
4   "siape": ["1488770"],
5   "document": { name: '50216' }
6 }
```

3.1.1.3 Resolução de Entidades

A última etapa se refere a resolução de entidades, onde o objetivo é identificar quais registros se referem à mesma entidade do mundo real. Isto é atingido comparando os registros da etapa anterior e agrupando por nome ou siape.

Por exemplo, ambas as portarias da Figura 3.1 e da Figura 3.3 fazem referência a servidora Renata de Matos Galante, ao final da etapa, haverá um documento do banco de dados referente a servidora Renata e ambas as portarias serão mencionadas como por exemplo a Listagem 3.2. Este processo vai se repetir a todas as entidades mencionadas em todas as portarias.

Figura 3.3 – Portaria Exemplo 2



SERVIÇO PÚBLICO FEDERAL

PORTARIA Nº 900 de 31/01/2018

O PRÓ-REITOR DE GESTÃO DE PESSOAS DA UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL, no uso de suas atribuições que lhe foram conferidas pela Portaria nº.7684, de 03 de outubro de 2016, do Magnífico Reitor, e conforme a Solicitação de Férias nº34471,

RESOLVE

Designar, temporariamente, nos termos da Lei nº. 8.112, de 11 de dezembro de 1990, com redação dada pela Lei nº.9.527, de 10 de dezembro de 1997, o ocupante do cargo de PROFESSOR DO MAGISTÉRIO SUPERIOR, do Quadro de Pessoal desta Universidade, **MARCELO WALTER** (Siape: 1550584), para substituir **RENATA DE MATOS GALANTE** (Siape: 1488770), Chefe do Depto de Informática Aplicada do Instituto de Informática, Código FG-1, em seu afastamento por motivo de férias, no período de 05/02/2018 a 14/02/2018, com o decorrente pagamento das vantagens por 10 dias.

MAURÍCIO VIÉGAS DA SILVA
Pró-Reitor de Gestão de Pessoas

Listagem 3.2 – Exemplo de uma entidade no banco de dados

```
1 {  
2   "records": [4630, 50216],  
3   "names": ["RENATA DE MATOS GALANTE"],  
4   "siapes": ["1488770"]  
5 }
```

A abordagem ACERPI-Link (EICH, 2021) buscou estender a abordagem ACERPI, melhorando a etapa de resolução de entidades ao melhorar a comparação de strings e também comparar o nome das entidades identificadas com entidades reais já conhecidas.

A abordagem ACERPI-Block (MARTINS, 2021) visou otimizar a resolução de entidades, melhorando sua performance. Foram exploradas diferentes técnicas de bloqueio diminuindo o tempo total para a geração de entidades e também diminuindo o ruído nos algoritmos de comparação de strings.

3.1.2 Considerações

Nesta seção apresentamos a abordagem ACERPI em mais detalhes, passando por todas suas etapas: coleta, extração e resolução de entidades. A abordagem ainda apresenta diversas limitações permitindo novas contribuições. A base de dados só possui informações sobre as entidades e sem muitos metadados sobre as portarias. O trabalho aqui desenvolvido busca cobrir essas limitações carregando as portarias para a base de dados e extraindo metadados sobre as portarias e por fim, apresentando consultas sobre os documentos.

3.2 Trabalhos Relacionados

Dois trabalhos relacionados foram usados como referência para o desenvolvimento deste trabalho. O primeiro intitulado *Regular expression-based reference metadata extraction from the web* (TANG et al., 2010) explora o uso de expressões regulares para extrair metadados após usar técnicas de *Web Scraping* para adquirir dados de sites. O

trabalho apresenta todos os estágios desde captura dos sites até a extração dos metadados e suas validações. O objetivo do trabalho é capturar os metadados e referências de sites de artigos científicos, muito similar a um dos objetivos do ACERPI-Query de extrair metadados depois de extrair os textos de PDFs.

O outro trabalho relacionado é intitulado *Modeling and Querying Data in MongoDB* (ARORA; AGGARWAL, 2013), onde os autores exploram a fundo o banco de dados MongoDB, apresentando exemplos de modelo de dados e da representação dos documentos. Além disso, o trabalho apresenta uma série de consultas exemplos no MongoDB, sempre comparando ao banco de dados relacional MySQL e apontando as diferenças, vantagens e desvantagens. O trabalho é similar à segunda parte do ACERPI-Query são executadas diversas consultas ao banco de dados ACERPI. Como o MongoDB é usado ao longo de todo o ACERPI-Query, esse trabalho possibilita um maior entendimento do banco de dados.

3.3 Considerações Finais

Este capítulo apresentou todos os trabalhos prévios relacionados a abordagem ACERPI, que são usados como base e foram fundamentais para o desenvolvimento deste trabalho. Além disso, são descritos dois trabalhos relacionados que foram usados como referência ao longo do trabalho, um relacionado a extração usando expressões regulares e outro relacionado a consultas usando o MongoDB.

4 EXTRAÇÕES

Este capítulo apresenta as extensões feitas no banco de dados da abordagem ACERPI. Primeiro, são apresentadas algumas mudanças nas coleções existentes e a adição da coleção de Portarias. Nas seções seguintes é apresentada uma série de novas extrações de informações que são carregadas para o banco de dados como metadados junto com as portarias. O objetivo é aumentar a quantidade de informações extraídas das Portarias para permitir as consultas mais avançadas descritas no Capítulo 5. As extrações são referentes às portarias publicadas pela UFRGS.

4.1 Modelo de dados

Como foi visto no capítulo anterior, a abordagem ACERPI tem como saída uma coleção de registros e de entidades, como pode ser observado nas Listagens 3.1 e 3.2 respectivamente. Apesar de possibilitar algumas consultas, é preciso estender as informações do banco de dados para permitir consultas mais avançadas.

O primeiro passo foi carregar também as próprias Portarias em uma coleção a partir dos XML resultantes da etapa de coleta. Além do texto, o documento também contém o nome do arquivo, o id da portaria e a data que foi emitida, informações que já estavam sendo extraídas. A Listagem 4.1 mostra um exemplo do formato dos arquivos XML, e a Listagem 4.2 demonstra como se refletiu para o documento na coleção das portarias.

Listagem 4.1 – Exemplo de uma portaria no formato XML

```

1 <Document id="8" nome_arquivo="47048" site="https://www1.ufrgs.br/
  sistemas/sde/gerencia-documentos/index.php/publico/ExibirPDF?
  documento=47048">
2   <portaria ID="10403" data="13/11/2017">
3     <text>PORTARIA N          10403          de
      13/11/2017, A PRÓ-REITORA DE GESTÃO DE PESSOAS DA
      UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL EM, EXERCÍCIO,
      no uso de suas atribuições que lhe foram conferidas
      pela Portaria n. 8117, de 10 de outubro de, 2016, e
      conforme a Solicitação de Afastamento n. 32907,, RESOLVE
      , Designar, temporariamente, nos termos da Lei n. 8.112,
      de 11 de dezembro de 1990, com redação, dada pela Lei n
      .9.527, de 10 de dezembro de 1997, a ocupante do cargo

```

de PROFESSOR DO MAGISTÉRIO, SUPERIOR, do Quadro de Pessoal desta Universidade, RENATA DE MATOS GALANTE (Siape: 1488770), para, substituir CARLA MARIA DAL SASSO FREITAS (Siape: 0351477), Diretor do Instituto de Informática, Código, CD-3, em seu afastamento no país, no período de 14/11/2017 a 15/11/2017, com o decorrente pagamento, das vantagens por 2 dias., VÂNIA CRISTINA SANTOS PEREIRA, Pró-Reitora</text>

4 </portaria>

5 </Document>

Listagem 4.2 – Exemplo de um documento de Portaria no banco de dados

```

1 {
2   "_id": {
3     "$oid": "62fc4ca71d992852e55a448d"
4   },
5   "file_name": "47048",
6   "ordinance_id": "10403",
7   "date": {
8     "$date": {
9       "$numberLong": "1510531200000"
10    }
11  },
12  "text": "PORTARIA N          10403
          de 13/11/2017, A PRÓ-REITORA DE
          GESTÃO DE PESSOAS DA UNIVERSIDADE FEDERAL DO RIO
          GRANDE DO SUL EM, EXERCÍCIO, no uso de suas
          atribuições que lhe foram conferidas pela Portaria
          n. 8117, de 10 de outubro de, 2016, e conforme a
          Solicitação de Afastamento n. 32907,, RESOLVE,
          Designar, temporariamente, nos termos da Lei n.
          8.112, de 11 de dezembro de 1990, com redação, dada
          pela Lei n.9.527, de 10 de dezembro de 1997, a
          ocupante do cargo de PROFESSOR DO MAGISTÉRIO,
          SUPERIOR, do Quadro de Pessoal desta Universidade,
          RENATA DE MATOS GALANTE (Siape: 1488770 ), para,

```

substituir CARLA MARIA DAL SASSO FREITAS (Siape: 0351477), Diretor do Instituto de Informática, Código, CD-3, em seu afastamento no país, no período de 14/11/2017 a 15/11/2017, com o decorrente pagamento, das vantagens por 2 dias., VÂNIA CRISTINA SANTOS PEREIRA, Pró-Reitora"

13 }

A partir disso já é possível uma união entre as três coleções, relacionando o id do campo *records* da Entidade com o id do Registro que possui o nome do documento, que será o nome do arquivo na coleção das Portarias. Entretanto, para evitar uma união muito complexa e custosa, a resolução de entidades foi alterada, incluindo o nome do arquivo na Entidade, possibilitando uma união direta entre as Entidades e as Portarias. Com estas extensões, já são possíveis múltiplas consultas relacionando as entidades e as portarias e também filtrando por suas datas. Além disso, já é possível buscas no texto de todas as portarias, como veremos em maiores detalhes no Capítulo 5.

Nas próximas seções, são apresentadas às novas extrações realizadas sobre as portarias, com o intuito de agregar mais metadados possibilitando consultas mais avançadas.

4.2 Extração da Assinatura

Todas as Portarias publicadas são assinadas por uma autoridade competente. Por exemplo, no caso da UFRGS, as portarias normalmente são assinadas pelo reitor(a), vice-reitor(a) ou pró-reitor(a). Analisando as portarias das Figuras 3.1 e 3.3, podemos identificar que as portarias sempre terminam com o nome da autoridade que está assinando, junto com sua função.

Como este padrão de assinatura se repete em todas as portarias, é possível extrair de cada portaria qual a autoridade que assinou e carregar este metadado junto no banco de dados. A extração é feita a partir dos arquivos de texto do passo anterior à estruturação dos arquivos XML. Identifica-se a penúltima linha do arquivo como o nome da autoridade, e a última linha como a função que a autoridade está exercendo. Ambas *strings* passam por um processo de higienização onde são removidos qualquer vírgula, ponto ou espaço extra. Um novo campo *signature* foi adicionado ao arquivo XML e ao documento resultante no banco de dados, como exemplificado na Listagem 4.3, documento formado a partir da

portaria da Figura 3.1.

Listagem 4.3 – Exemplo de um documento de Portaria com assinatura

```

1 {
2   "_id": {
3     "$oid": "62fc4ca71d992852e55a448d"
4   },
5   "file_name": "47048",
6   "ordinance_id": "10403",
7   "date": {
8     "$date": {
9       "$numberLong": "1510531200000"
10    }
11  },
12  "signature": {
13    "name": "VANIA CRISTINA SANTOS PEREIRA",
14    "role": "PRO-REITORA"
15  },
16  "text": "PORTARIA N 10403 de 13/11/2017, A PRÓ-
          REITORA DE GESTÃO DE PESSOAS [...] das vantagens
          por 2 dias. VÂNIA CRISTINA SANTOS PEREIRA, Pró-
          Reitora"
17 }

```

4.3 Extração da Progressão e Promoção de Professores

Portarias também são emitidas para informar a concessão de progressão de carreiras aos servidores. Nesta seção, é apresentada a extração destas progressões a respeito de professores.

O plano de carreira do professor do Magistério Superior contém 5 classes, cada classe possuindo níveis internos (BRASIL, 2012), como representado na Tabela 4.1. É determinado como **Progressão** a passagem para o nível seguinte dentro da mesma classe, e de **Promoção** a passagem de uma classe para a outra. A última classe, denominada Titular, possui apenas um nível. A classe A pode ser denominada de três diferentes for-

mas, dependendo do título do professor: Adjunto-A, se doutor; Assistente-A, se mestre; e Auxiliar, se graduado ou especialista.

Tabela 4.1 – Tabela de Progressões e Promoções - Magistério Superior

Classe	Denominação	Nível
E	Titular	Único
D	Associado	4
		3
		2
		1
C	Adjunto	4
		3
		2
		1
B	Assistente	2 1
A	Adjunto-A (se doutor)	2
	Assistente-A (se mestre)	1
	Auxiliar (se graduado ou especialista)	

Portarias de progressão de carreira do Magistério Superior possuem um padrão na sua estrutura de texto, o que possibilita o uso de expressões regulares a fim de reconhecer quais portarias se referem a progressões e capturar informações do professor, a progressão sendo concedida e a data de início. Na Figura 4.1, é possível ver um exemplo de uma portaria sobre a progressão da professora, e na Figura 4.2 um exemplo de promoção. É possível analisar o padrão e os pontos interessantes a serem capturados, as portarias começam com a frase "Conceder progressão funcional" ou "Conceder promoção funcional", assim é possível identificar as portarias de sobre progressão/promoção e capturar qual o tipo. Outro ponto de interesse é o trecho "Professora TANIA RAMOS FORTUNA," que se encontra logo depois, podendo assim capturar o nome do professor ou professora, o sexo irá influenciar entre "Professora" e "Professor". O próximo ponto é a denominação da classe de origem e destino junto com seus níveis, levando em conta as 7 denominações possíveis e 4 níveis. Outro cuidado é que na classe Titular, não possui nenhum nível. Neste exemplo as capturas de origem são "Adjunto, nível 03" e de destino "Adjunto, nível

04". A última captura é referente a data de início, logo que o início no novo cargo não necessariamente será a partir da portaria. A data é extraída a partir do trecho "com vigência financeira a partir de 07/10/2017,". Outra possibilidade no lugar da data é a frase "com vigência financeira a partir da data de publicação da portaria", neste caso, é usado o metadado da data de publicação da portaria que já estava sendo extraída. A expressão regular resultante que segue esses padrões captura as informações necessárias e lida com as possíveis alternativas, conforme apresentadas na Listagem 4.4.

Figura 4.1 – Portaria de Progressão

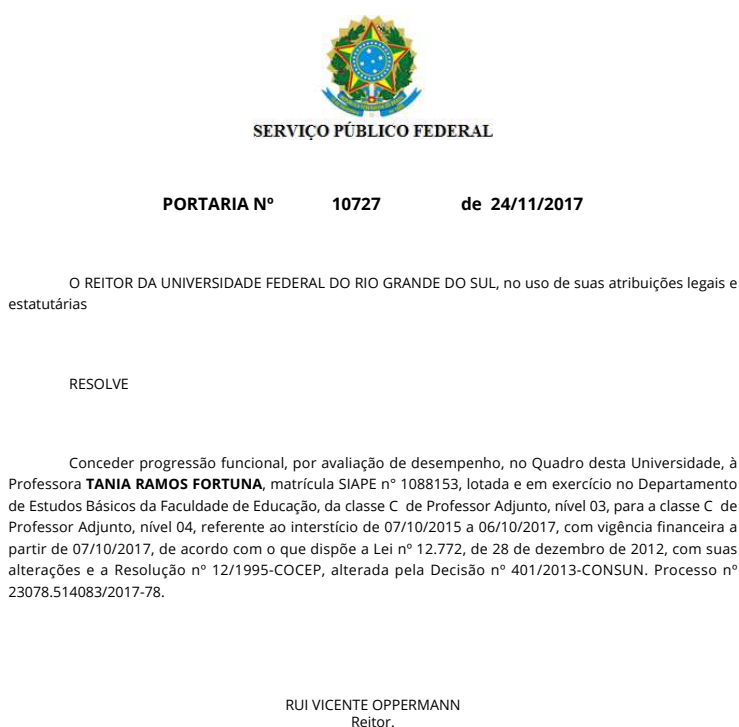


Figura 4.2 – Portaria de Promoção



SERVIÇO PÚBLICO FEDERAL

PORTARIA Nº 9826 de 31/10/2019

A VICE-REITORA DA UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL, no uso de suas atribuições, considerando o disposto na Portaria nº 7624, de 29 de setembro de 2016

RESOLVE

Conceder promoção funcional, por avaliação de desempenho, no Quadro desta Universidade, à Professora **TANIA RAMOS FORTUNA**, matrícula SIAPE nº 1088153, lotada e em exercício no Departamento de Estudos Básicos da Faculdade de Educação, da classe C de Professor Adjunto, nível 04, para a classe D de Professor Associado, nível 01, referente ao interstício de 07/10/2017 a 06/10/2019, com vigência financeira a partir de 07/10/2019, conforme **decisão judicial proferida no processo nº 5054491-30.2018.4.04.7100**, da 1ª Vara Federal de Porto Alegre, de acordo com o que dispõe a Lei 12.772 de 28 de dezembro de 2012, com suas alterações e a Decisão nº 331/2017 do CONSUN. Processo nº 23078.524478/2019-41.

JANE FRAGA TUTIKIAN
Vice-Reitora.

Listagem 4.4 – Expressão Regular Progressão de Professores

1 Conceder\s*(promoção|promocao|progressão)\s*funcional.*(?:Professor|Professora)\s*([A-ZÁÀÃÄÊËÈÍÓÔÕŨ\s]+),.*(Adjunto|Al Auxiliar|Assistente|Adjunto|Associado|Titular).*nível,?\s*([0-5]+).* (Adjunto|Al Auxiliar|Assistente|Adjunto|Associado|Titular)(?:.*nível,?\s*([0-5]+))?.*partir.*\s*([0-9]{1,2}\|[0-9]{2}\|[0-9]{4}|publicação)

Portanto, as informações capturadas são se progressão ou promoção, o nome do professor, a classe e o nível de origem, a classe e o nível de destino, e data de início. Todos estes metadados são carregados junto com a portaria no XML resultante, como ilustrado no exemplo da Listagem 4.5. Ao carregar os dados no banco de dados, os campos de progressão também são carregados no documento, como no exemplo da Listagem 4.6, um campo *type* foi adicionado ao documento para facilitar a identificação do tipo da portaria, os tipos de progressão e promoção receberam o *type* igual a *progression*. Com a adição destes metadados ao banco, é possível construir consultas mais avançadas como veremos no capítulo 5.

Listagem 4.5 – Exemplo de uma promoção no formato XML

```

1 <Document id="3" nome_arquivo="100170" site="https://www1.ufrgs.br/
  sistemas/sde/gerencia-documentos/index.php/publico/ExibirPDF?
  documento=100170">
2   <portaria ID="9826" data="31/10/2019" type="progression">
3     <text>PORTARIA N 9826 de 31/10/2019, A VICE-REITORA DA
4       [...] </text>
5     <progression type="PROMOCAO">
6       <name>TANIA RAMOS FORTUNA</name>
7       <originclass>ADJUNTO</originclass>
8       <originlevel>04</originlevel>
9       <destinyclass>ASSOCIADO</destinyclass>
10      <destinylevel>01</destinylevel>
11      <dateeffective>07/10/2019</dateeffective>
12    </progression>
13  </portaria>
14 </Document>

```

Listagem 4.6 – Exemplo de um documento de Portaria com promoção

```

1 {
2   "_id": {
3     "$oid": "62fc4ca21d992852e55a3b5e"
4   },
5   "file_name": "100170",
6   "ordinance_id": "9826",
7   "date": {
8     "$date": {
9       "$numberLong": "1572480000000"

```

```
10     }
11   },
12   "type": "progression",
13   "signature": {
14     "name": "JANE FRAGA TUTIKIAN",
15     "role": "VICE-REITORA"
16   },
17   "text": "PORTARIA N 9826 de 31/10/2019, A VICE-
18     REITORA [...]",
19   "progression": {
20     "progression_type": "PROMOCAO",
21     "name": "TANIA RAMOS FORTUNA",
22     "origin_class": "ADJUNTO",
23     "origin_level": "4",
24     "destiny_class": "ASSOCIADO",
25     "destiny_level": "1",
26     "date_effective": {
27       "$date": {
28         "$numberLong": "1570406400000"
29       }
30     }
31   }
```

4.4 Extração da Progressão de Servidores Técnico-Administrativos

Assim como a progressão do Magistério Superior, a progressão dos outros servidores da UFRGS é informada através de Portarias. Entretanto, a progressão de carreira de servidores de cargos Técnico-Administrativos segue classes e níveis diferentes e, portanto, um padrão de portaria diferente. O Plano de Carreira de cargos Técnico-Administrativos está estruturado em cinco níveis de classificação: A, B, C, D e E, com quatro níveis de capacitação cada. A classificação é determinada pelo tipo do cargo, seguindo a Lei nº 11.091/2005 (BRASIL, 2005). Diferente dos professores, não existe pro-

Listagem 4.7 – Expressão Regular Progressão de Servidores

```

1 Conceder\s*Progressão\s*por\s*Capacitação.*a\s*contar\s*de\s
  *([0-9]{1,2}\/[0-9]{2}\/[0-9]{4}).*(?:servidor|servidora)\s*([A-
  ZÁÀÃÄËÊËÍÓÔÕŨ\s]+).*cargo,?\s*de\s*([A-ZÁÀÃÄËÊËÍÓÔÕŨñáâã\s
  ,-\s]+).*Capacitação\s*([A-Z]\s*[IV]+).*Capacitação\s*([A-Z]\s*[IV
  ]+)

```

gressão entre as classes denominadas por letras, pois cada classificação é referente a cargos diferentes, a progressão de carreira dos servidores de cargo Técnico-Administrativos se refere apenas a progressão entre os níveis de capacitação.

As portarias referentes a progressão por capacitação dos servidores, seguem sempre o mesmo padrão, começando pela frase "Conceder Progressão por Capacitação", como exemplificado na Figura 4.3. As informações de interesse para captura são o nome do servidor, a data de início do novo nível, o cargo e o nível de origem e destino. A data é capturada por vir logo após a frase "a contar de"; o próximo trecho que capturamos é o nome do servidor que precedido pela palavra "servidor" ou "servidora"; o cargo se encontra após o trecho "cargo de"; e por fim, os níveis de origem e destino se encontram depois da palavra "Capacitação", como por exemplo, "Capacitação C II" e após "Capacitação C III", a ordem que elas aparecem determina a origem e o destino; a expressão regular resultante que captura todos esses trechos e ainda lida com possíveis diferenças é a apresentada na Listagem 4.7.

As informações capturadas são todas inseridas juntas nos arquivos XML, como na Listagem 4.8, e após carregadas no banco de dados, o documento resultante é exemplificado na Listagem 4.9. O campo *type*, assim como na progressão de professores, é *progression*, e o objeto com as informações também se chama *progression*, entretanto, dentro desse objeto é especificado que o tipo de progressão é de servidor.

Assim, conseguimos capturar quais portarias se referem a progressão e quais seus metadados de todos os funcionários da UFRGS, tanto professores, quanto servidores de cargos Técnico-Administrativos.

Listagem 4.8 – Exemplo de uma progressão de servidor no formato XML

```

1 <Document id="0" nome_arquivo="102233" site="https://www1.ufrgs.br/
  sistemas/sde/gerencia-documentos/index.php/publico/ExibirPDF?
  documento=102233">
2 <portaria ID="11125" data="13/12/2019" type="progression">

```

Figura 4.3 – Portaria de Progressão Servidor



SERVIÇO PÚBLICO FEDERAL

PORTARIA Nº 11125 de 13/12/2019

O PRÓ-REITOR DE GESTÃO DE PESSOAS DA UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL, no uso de suas atribuições, considerando o disposto na Portaria nº 7684, de 03 de outubro de 2016

RESOLVE

Conceder Progressão por Capacitação, a contar de 10/12/2019, referente ao interstício de 09/11/2017 a 09/12/2019, para o servidor **ALTERMIR MOREIRA DE ABREU**, ocupante do cargo de **Porteiro - 701458**, matrícula SIAPE 0358438, lotado na Faculdade de Odontologia, passando do Nível de Classificação/Nível de Capacitação C II, para o Nível de Classificação/Nível de Capacitação C III, em virtude de ter realizado o(s) seguinte(s) curso(s), conforme o Processo nº 23078.534226/2019-20:

Formação Integral de Servidores da UFRGS I CH: 30 (08/08/2017 a 31/08/2017)
ENAP - Um por Todos e Todos por Um - Pela Ética e Cidadania CH: 40 Carga horária utilizada: 20 hora(s) / Carga horária excedente: 20 hora(s) (10/12/2019 a 10/12/2019)
ENAP - Gestão Estratégica de Pessoas e Planos de Carreira CH: 20 (10/12/2019 a 10/12/2019)
ENAP - Gestão da Informação e Documentação - Conceitos Básicos em Gestão Documental CH: 20 (05/12/2019 a 06/12/2019)

MAURÍCIO VIÉGAS DA SILVA
Pró-Reitor de Gestão de Pessoas

```

3      <text>PORTARIA N 11125 de 13/12/2019, O PRÓ-REITOR [...]</
      text>
4      <progression type="SERVIDOR">
5          <name>ALTERMIR MOREIRA DE ABREU</name>
6          <originlevel>C II</originlevel>
7          <destinylevel>C III</destinylevel>
8          <dateeffective>10/12/2019</dateeffective>
9          <role>PORTEIRO</role>
10     </progression>
11 </portaria>
12 </Document>

```

Listagem 4.9 – Exemplo de um Documento de Promoção de um Servidor

```

1 {
2   "_id": {
3     "$oid": "62fc4c6e1d992852e559e55c"
4   },
5   "file_name": "102233",
6   "ordinance_id": "11125",
7   "date": {
8     "$date": {
9       "$numberLong": "1576195200000"
10    }
11  },
12  "type": "progression",
13  "signature": {
14    "name": "MAURICIO VIEGAS DA SILVA",
15    "role": "PRO-REITOR DE GESTAO DE PESSOAS"
16  },
17  "text": "PORTARIA N 11125 de 13/12/2019, O PRÓ-REITOR
      [...]",
18  "progression": {
19    "progression_type": "SERVIDOR",
20    "name": "ALTERMIR MOREIRA DE ABREU",
21    "origin_level": "C II",
22    "destiny_level": "C III",

```



```

23     "date_effective": {
24         "$date": {
25             "$numberLong": "1575936000000"
26         }
27     },
28     "role": "PORTEIRO"
29 }
30 }

```

4.5 Extração da Aposentadoria

Outro tipo de portaria é referente a aposentadoria de um servidor, a aposentadoria significando o fim da carreira deste servidor na Instituição. Assim como as progressões, as portarias de aposentadorias seguem um padrão, possibilitando a extração usando expressões regulares novamente. Neste caso, a portaria segue o mesmo padrão para todos os servidores, incluindo os professores do Magistério Superior.

A Figura 4.4 exemplifica uma portaria referente à aposentadoria. O padrão começa com "Conceder aposentadoria", possibilitando identificar que a portaria se refere a aposentadoria; após, identificamos o trecho "a TANIA RAMOS FORTUNA", referente ao nome do servidor, neste caso as letra maiúsculas e a palavra "matrícula" são essenciais para identificar o trecho que se refere ao nome. Nenhuma outra informação do texto da portaria é necessária, porque a data de aposentadoria sempre é referente a data de publicação, e o cargo em que o servidor está se aposentando é possível inferir de portarias de progressões anteriores. A expressão regular resultante pode ser vista na Listagem 4.10.

Listagem 4.10 – Expressão Regular de Aposentadorias

```

1 Conceder aposentadoria.*\sa\s*([A-ZÀÃÄÅËÊÈÉÍÓÔÕÛÑ\s,]+).*
   matrícula

```

Assim como nas progressões, as informações capturadas são armazenadas como um metadado no arquivo XML, como na Listagem 4.11. Por fim, estes dados são carregados junto no documento armazenado no banco de dados; entretanto, agora o campo *type*

Figura 4.4 – Portaria de Aposentadoria



SERVIÇO PÚBLICO FEDERAL

PORTARIA Nº 10490 de 21/11/2019

O REITOR DA UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL, no uso de suas atribuições legais e estatutárias

RESOLVE

Conceder aposentadoria, nos termos do artigo 3º, incisos I, II e III, da Emenda Constitucional nº 47, de 5 de julho de 2005, publicada no Diário Oficial da União do dia 6 subsequente, e de acordo com o artigo 3º da Emenda Constitucional nº 103, de 12 de novembro de 2019, a **TANIA RAMOS FORTUNA**, matrícula SIAPE nº 1088153, no cargo de Professor Associado, nível 1, da Carreira do Magistério Superior, do Quadro desta Universidade, no regime de dedicação exclusiva, com exercício no Departamento de Estudos Básicos da Faculdade de Educação, com proventos integrais. Processo 23078.530263/2019-69.

RUI VICENTE OPPERMANN
Reitor.

é *retirement*, para se referir ao tipo daquela portaria, como é possível ver no exemplo na Listagem 4.12. Logo, além de capturar os metadados referentes a todas as progressões, como apresentado nas seções anteriores, também é possível identificar as aposentadorias que podemos considerar o fim da carreira do servidor na instituição.

Listagem 4.11 – Exemplo de uma Aposentadoria no formato XML

```

1 <Document id="9" nome_arquivo="101217" site="https://www1.ufrgs.br/
  sistemas/sde/gerencia-documentos/index.php/publico/ExibirPDF?
  documento=101217">
2   <portaria ID="10490" data="21/11/2019" type="retirement">
3     <text>PORTARIA N 10490 de 21/11/2019, O REITOR [...]</text
4     >
5     <retirement type="retirement">
6       <name>TANIA RAMOS FORTUNA</name>
7       <date>21/11/2019</date>
8     </retirement>
9   </portaria>
10 </Document>

```

Listagem 4.12 – Exemplo de um Documento de Aposentadoria

```

1 {
2   "_id": {
3     "$oid": "62fc4c881d992852e55a1205"
4   },
5   "file_name": "101217",
6   "ordinance_id": "10490",
7   "date": {
8     "$date": {
9       "$numberLong": "1574294400000"
10    }
11  },
12  "type": "retirement",
13  "signature": {
14    "name": "RUI VICENTE OPPERMANN",
15    "role": "REITOR"
16  },
17  "text": "PORTARIA N 10490 de 21/11/2019, O REITOR

```

```
    [...].",  
18  "retirement": {  
19    "name": "TANIA RAMOS FORTUNA",  
20    "date": {  
21      "$date": {  
22        "$numberLong": "1574294400000"  
23      }  
24    }  
25  }  
26 }
```

4.6 Considerações Finais

Ao longo deste capítulo, foram apresentadas algumas alterações referente ao banco de dados da abordagem ACERPI. Uma nova coleção foi carregada referente às portarias, a coleção de *entities* foi alterada para facilitar consultas que necessitam união entre as duas coleções; e além disso, diversas extrações de informações das portarias foram feitas e adicionadas à coleção de portarias, como a extração de assinatura, a identificação e extração de informações referente a progressão e promoção de professore e de servidores Técnico-Administrativos, e também sobre aposentadorias.

Todas essas extensões ao banco de dados original tiveram o objetivo de aumentar a quantidade e qualidade de informações disponíveis. Logo, o que antes era um banco de dados com possibilidades limitadas de consultas, ao fim deste capítulo se tornou um banco de dados com muito mais metadados, possibilitando consultas mais avançadas e significativas, possibilitando extrair maior valor dos dados armazenados e cumprir o objetivo de facilitar para o usuário o acesso a informações de portarias. Estas consultas são discutidas e apresentadas no próximo capítulo.

5 CONSULTAS

O capítulo anterior descreveu diversas extrações e extensões ao banco de dados da abordagem ACERPI, todas com o objetivo de aumentar a quantidade de metadados para possibilitar a usuários consultas mais avançadas. Este capítulo apresenta consultas com o intuito de extrair conhecimento e valor do banco de dados. As primeiras seções começam com consultas mais básicas e em seguida, consultas cada vez mais avançadas são apresentadas.

A fonte dos dados usada foi o repositório da UFRGS (UFRGS, 2016), que contém todas portarias desde 2016, para estas consultas foram usados dados coletados até 3 de Março de 2020, contendo 44865 arquivos. Os dados de 2019 de Janeiro a Outubro sofreram problemas no processamento e não estão completos, entretanto isto não prejudica o objetivo desta etapa do trabalho que tem como foco apenas apresentar as consultas possíveis.

5.1 Consultas Básicas de Entidades e Portarias

A Seção 4.1 apresentou a nova coleção das portarias e também modificações a coleção *entities* para facilitar a união entre as duas. Isto possibilitou consultas que relacionam entidades com as portarias e também com suas datas, como demonstrado nesta seção.

Consulta 1. Quais portarias o servidor X aparece na janela de tempo Y?

Listagem 5.1 – Consulta 1

```
1  [{
2    "$match": {
3      "names": "RENATA DE MATOS GALANTE"
4    }},
5  {
6    "$lookup": {
7      "from": "ordinances_xml",
8      "localField": "ordinances",
9      "foreignField": "file_name",
10     "as": "result",
```

```

11     "pipeline": [
12         {
13             "$match": {
14                 "date": {
15                     "$gt": ISODate('2017-01-01'),
16                     "$lt": ISODate('2018-12-31')
17                 }
18             }
19         }
20 ]

```

A Listagem 5.1 demonstra a *pipeline* de agregação construída para esta consulta. A *query* é feita na coleção *entities*, no primeiro estágio filtra as entidades pelo nome desejado, neste exemplo "RENATA DE MATOS GALANTE"; no segundo, a operação *lookup* executa uma união com a coleção *ordinances* das portarias, unindo pela lista de portarias na entidade e o nome do arquivo presente na portaria; e por fim, filtra por apenas as portarias com datas maiores e menores que as datas providas, neste exemplo 28/10/2017 e 28/10/2018, respectivamente. A consulta é útil para usuários buscando portarias em que ele ou um servidor específico aparecem, podendo também filtrar por um período.

Consulta 2. Qual entidade com mais portarias na janela de tempo Y?

Listagem 5.2 – Consulta 2

```

1  [{
2    $lookup: {
3      from: "ordinances_xml",
4      localField: "ordinances",
5      foreignField: "file_name",
6      as: "portarias",
7      pipeline: [
8        {
9          $match: {
10             date: {
11                 $gt: ISODate("2018-10-28"),
12                 $lt: ISODate("2020-10-28")
13             }
14         }
15     }
16 }

```

```

14   }
15 }, {
16   $match: {
17     "portarias.0": {
18       $exists: true
19     }}
20 }, {
21   $project: {
22     count: {
23       $size: {
24         $ifNull: ["$portarias", []]
25       }},
26     names: 1,
27     siapes: 1,
28     portarias: 1
29   }
30 }, {
31   $sort: {
32     count: -1
33   }
34 }, {
35   $limit: 3
36  }]

```

A operação de *lookup* é a mesma vista na consulta anterior; após isso, a operação *match* filtra pelas entidades que possuem pelo menos uma portaria; a operação de *project* adiciona um campo *count* baseado no número de portarias presente para que nos estágios seguintes seja possível ordenar por esse campo e filtrar pelas três primeiras entidade.

Consulta 3. Quais portarias aparecem na janela de tempo Y?

A consulta é executada na coleção de portarias, retornando todas as portarias entre as datas. Útil para consultar as portarias de um determinado mês ou ano, por exemplo.

Listagem 5.3 – Consulta 3

```

1  [{
2    "$match": {

```

```

3     "date": {
4         "$gt": ISODate("2017-10-01"),
5         "$lt": ISODate("2017-10-31")
6     }
7 ]

```

Consulta 4. Buscas nos textos

Outra consulta importante é permitir fazer pesquisas diretamente nos textos de todas as portarias, pesquisando por trechos, palavras chaves, ou nomes de pessoas. Além disso, é possível inclusive permitir buscas usando expressões regulares. No exemplo da Listagem 5.4, filtramos pelas portarias com a expressão regular referente a progressão de carreira, como discutido no Capítulo 4.

Listagem 5.4 – Consulta 4

```

1  [{
2     "$match": {
3         "text": {
4             "$regex": "Conceder\\s*(promocao|progressao)\\s*
                    funcional"
5         }
6     }
7 ]

```

5.2 Consultas Por Assinaturas

A Seção 4.2 apresentou a extração do nome e função da pessoa que assinou a portaria. Nesta seção, algumas consultas que se tornaram possíveis serão descritas.

Consulta 5. Qual o servidor que mais assinou portarias?

Listagem 5.5 – Consulta 5

```

1  [{
2     "$sortByCount": "$signature.name"
3  }, {
4     "$limit": 3
5  }
6 ]

```


A operação *sortByCount* pode ser usada para agrupar todas as portarias pelo campo do nome e ordenar pelo número de aparições. A Listagem 5.6 apresenta a saída da consulta, nota-se que as três pessoas com mais assinaturas exerceram o cargo de, em ordem, Pró-Reitor de Gestão de Pessoas, Vice-Reitora e Reitor.

Listagem 5.6 – Resultado consulta 5

```
1 {"_id": "mauricio viegas da silva", "count": 15248},
2 {"_id": "jane fraga tutikian", "count": 5947},
3 {"_id": "rui vicente oppermann", "count": 5216}
```

Consulta 6. Qual a função que mais assinou portarias?

Listagem 5.7 – Consulta 6

```
1 [{
2   "$sortByCount": "$signature.role"
3 }, {
4   "$limit": 3
5 }]
6 ]
```

A consulta é similar à anterior, porém filtrando apenas pela função na assinatura. A Listagem 5.8 apresenta o resultado da consulta, nota-se algumas semelhanças com as funções dos servidores da consulta anterior, porém com algumas diferenças, uma vez que a cada mandato o servidor exercendo a função trocará.

Listagem 5.8 – Resultado consulta 6

```
1 {'_id': 'pro-reitor de gestao de pessoas', 'count': 1526
2   4},
3 {'_id': 'reitor', 'count': 6315},
4 {'_id': 'vice-reitora no exercicio da reitoria', 'count':
5   2970}
```

Consulta 7. Quais portarias assinou o servidor X na janela de tempo Y?

Listagem 5.9 – Consulta 7

```
1 [{
2   "$match": {
```

```

3     "signature.name": "mauricio viegas da silva",
4     "date": {
5         "$gt": datetime(2018, 1, 1, 0, 0, 0),
6         "$lt": datetime(2018, 12, 31, 23, 58, 59)
7     }
8 ]

```

A consulta filtra por um servidor e limita em um período de tempo, útil para quando é preciso averiguar quais portarias foram assinadas por um servidor.

Consulta 8. Quais portarias foram assinadas pela função X no tempo Y?

Listagem 5.10 – Consulta 8

```

1  [{
2     "$match": {
3         "signature.role": "reitor",
4         "date": {
5             "$gt": datetime(2018, 1, 1, 0, 0, 0),
6             "$lt": datetime(2018, 12, 31, 23, 58, 59)
7         }
8     }
9 ]

```

Similarmente a consulta anterior, nesta consulta é possível filtrar por um cargo e a data. Útil para consultarmos todas as portarias que o reitor assinou em um determinado ano, por exemplo.

5.3 Consultas Por Progressões e Aposentadorias

As últimas seções do Capítulo 4 descreviam extensões referentes à progressão de carreira dos servidores da UFRGS, tanto dos professores do Magistério Superior quanto de servidores Técnico-Administrativos. Nesta seção, algumas consultas são apresentadas; porém, estas consultas são melhor utilizadas para possibilitar as consultas na seção 5.4.

Consulta 9. Quais as portarias se referem a progressão/promoção?

Como visto na Seção 4.3, 4.4 e 4.5, um novo campo *type* foi adicionado as portarias, possibilitando consultas pela categoria das portarias. Esse tipo de consulta é útil para categorizar todas as portarias, facilitando consultas sobre elas. Também é possível

consultas a portarias referentes a aposentadoria, usando o *type* como *retirement*

Listagem 5.11 – Consulta 9

```

1  [{
2    "$match": {
3      "type": "progression"
4    }}
5  ]

```

Consulta 10. Qual o professor/servidor com mais progressões?

Listagem 5.12 – Consulta 10

```

1  [{
2    "$match": {
3      "type": "progression"
4    }
5  }, {
6    "$group": {
7      "_id": "$progression.name",
8      "progressions": {
9        "$addToSet": "$progression"
10     }}
11  }, {
12    "$project": {
13      "progressions": {
14        "$sortBy": {
15          "input": "$progressions",
16          "sortBy": {
17            "date_effective": 1
18          }}
19     },
20    "progressions_size": {
21      "$size": {
22        "$ifNull": ["$progressions", []]
23      }}
24  }, {

```

```

25     "$sort": {
26         "progressions_size": -1
27     }
28 }
29 ]

```

Nos primeiros estágios, a consulta agrupa as portarias pelo nome do servidor; após, a operação *project* ordena as progressões pela data e adiciona um campo com o número de progressões, para que no estágio seguinte seja possível ordenar por ele.

Consulta 11. Quais foram todas as progressões do professor/servidor X?

Listagem 5.13 – Consulta 11

```

1  [{
2     "$match": {
3         "type": "progression",
4         "progression.name": "ELIANE BANDINELLI"
5     }},
6  {
7     "$group": {
8         "_id": "$progression.name",
9         "progressions": {
10            "$addToSet": "$progression"
11        }
12    }}, {
13     "$project": {
14         "progressions": {
15             "$sortBy": {
16                 "input": "$progressions",
17                 "sortBy": {
18                     "date": 1
19                 }
20             }
21         }
22     }
23 }

```

A consulta agrupa todas as portarias de progressões pelo servidor sendo promovido e retorna todas as progressões. Esta consulta pode ser utilizada para identificarmos

todas as vezes que um servidor foi promovido, junto com os cargos e a data, como demonstrado na Listagem 5.14.

Listagem 5.14 – Resultado Consulta 11

```

1 { "_id": "ELIANE BANDINELLI",
2   "progressions": [
3     {
4       "progression_type": "PROGRESSAO",
5       "origin_class": "ADJUNTO",
6       "origin_level": "2",
7       "destiny_class": "ADJUNTO",
8       "destiny_level": "3",
9       "date_effective":
10        {"$date": {"$numberLong": "1167609600000"}}
11     },
12    {
13      "progression_type": "PROGRESSAO",
14      "origin_class": "ADJUNTO",
15      "origin_level": "3",
16      "destiny_class": "ADJUNTO",
17      "destiny_level": "4",
18      "date_effective":
19       {"$date": {"$numberLong": "1230768000000"}}
20     },
21    {
22      "progression_type": "PROMOCAO",
23      "origin_class": "ADJUNTO",
24      "origin_level": "4",
25      "destiny_class": "ASSOCIADO",
26      "destiny_level": "1",
27      "date_effective": {
28        "$date": {"$numberLong": "1293840000000"}
29      }
30    }
31  ]

```

32]}

Consulta 12. Quantas progressões aconteceram em cada mês?

Listagem 5.15 – Consulta 12

```

1  [{
2      "$project": {
3          "month": {
4              "$month": "$progression.date_effective"
5          },
6          "year": {
7              "$year": "$progression.date_effective"
8          }
9      }
10 }, {
11     "$group": {
12         "_id": {
13             "month": "$month",
14             "year": "$year"
15         },
16         "count": {
17             "$count": {}
18         }
19     }
20 }, {
21     "$sort": {
22         "_id.year": 1,
23         "_id.month": 1
24     }
25 ]

```

O estágio *project* adiciona os campos de mês e ano; após, a consulta agrupa todas as progressões pelo mês e ano que elas aconteceram e retorna a quantidade em cada mês, podendo ser utilizada para fazer um levantamento de quantos servidores foram promovidos ao longo dos meses e anos. A Tabela 5.1 apresenta o resultado.

Consulta 13. Quantas aposentadorias aconteceram a cada mês?

Tabela 5.1 – Tabela de Progressões por Mês e Ano

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2016	9	47	40	47	50	46	75	460	97	95	65	130
2017	111	138	115	99	108	99	113	132	79	88	100	92
2018	64	70	112	133	99	132	90	155	105	124	83	83
2019	6	6	6	3	4	4	12	22	15	37	60	74
2020	68	51	1									

Listagem 5.16 – Consulta 13

```

1  [{
2    "$project": {
3      "month": {
4        "$month": "$retirement.date"
5      },
6      "year": {
7        "$year": "$retirement.date"
8      }
9    }
10 }, {
11   "$group": {
12     "_id": {
13       "month": "$month",
14       "year": "$year"
15     },
16     "count": {
17       "$count": {}
18     }
19   }
20 }, {
21   "$sort": {
22     "_id.year": 1,
23     "_id.month": 1
24   }
25 }
26 ]

```

Similarmente, podemos fazer um levantamento das aposentadorias dos servidores e gerar relatórios sobre o número de aposentadorias. A Tabela 5.2 apresenta os resultados.

Tabela 5.2 – Tabela de Aposentadorias por Mês e Ano

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2016	0	12	29	19	15	17	10	17	9	15	6	4
2017	11	26	43	30	47	39	27	16	26	13	12	9
2018	3	24	37	18	18	13	18	21	7	12	13	9
2019										2	13	5
2020	5	13	2									

5.4 Histórico de Professores e Servidores

As últimas seções apresentaram algumas consultas a respeito das progressões e aposentadoria dos servidores. Com essas informações é possível rastrear todo o histórico da carreira do servidor com todos os cargos, as data de início e a datas de sua aposentadoria, como demonstrado nas próxima consultas.

Consulta 14. Histórico do Servidor.

Para auxiliar esta consulta, foi utilizado um *script* em Python, que conecta com o MongoDB e realiza as consultas anteriores agregando as progressões e aposentadorias dos servidores; em seguida, ordena cronologicamente e constrói o histórico utilizando sempre o nível de destino da progressão e a data efetiva. Após, é considerada a aposentadoria, caso exista, como o último nível e é concatenada junto com sua data. Por fim, é possível extrair da primeira progressão, que se tem registro, o nível de origem, pois esse é o cargo de entrada do servidor; entretanto, a data é desconhecida e é registrada como "data desconhecida". O *script* é apresentado no Algoritmo 1 e seu resultado é demonstrado na Listagem 5.17.

Algoritmo 1: Algoritmo para construção do histórico de servidores.

Entrada: conjunto de progressoes, aposentadoria
Saída : historico

```

1 primeiro ← conjunto de progressoes[0];
2 historico ← primeiro;

3 para cada progressao p em conjunto de progressoes faça
4   cargo ← novo_cargo(p["destiny_class"], p["destiny_level"],
5     p["date_effective"]);
6   historico ← historico ∪ cargo;
7 fim
8 se aposentadoria então
9   historico ← historico ∪ aposentadoria;
10 fim

```

 Listagem 5.17 – Resultado Consulta 14

```

1 TANIA RAMOS FORTUNA
2 ["ADJUNTO", "3", "data desconhecida"]
3 ["ADJUNTO", "4", "07/10/2017"]
4 ["ASSOCIADO", "1", "07/10/2019"]
5 ["Aposentadoria", "21/11/2019"]
6
7 ARI PAULO ZIMMER GASPAR
8 ["ASSISTENTE EM ADMINISTRACAO", "D I", "data
   desconhecida"]
9 ["ASSISTENTE EM ADMINISTRACAO", "D II", "29/04/2016"]
10 ["ASSISTENTE EM ADMINISTRACAO", "D III", "08/12/2017"]
11 ["Aposentadoria", "07/08/2018"]

```

Com essa consulta, é possível montar a carreira de todos os servidores a partir das portarias, possibilitando que usuários consultem o histórico dos servidores.

Consulta 15. Qual o cargo do servidor X no instante de tempo Y?

Após possuir o histórico da carreira do servidor, outra consulta possível é consultar o cargo do servidor em um determinado dia. Por exemplo, se o professor virou Adjunto Nível 4 em 2017 e foi promovido a Associado em 2019, ao consultar qual era o cargo em um dia de 2018, o cargo de Adjunto Nível 4 é inferido. Isto é útil em situações onde é preciso saber um retrato dos cargos de todos os servidores em determinado instante de tempo. Similar a anterior, esta consulta foi desenvolvida utilizando o mesmo *script* em

Python, ao passar uma data específica como parâmetro, é verificado qual era o cargo do servidor na data. O pseudocódigo é apresentado no Algoritmo 2 e um exemplo de seu resultado é demonstrado na Listagem 5.18 ao consultarmos os cargos no dia 20/09/2018.

Algoritmo 2: Algoritmo para o cargo do servidor em uma data.

Entrada: historico, data

Saída : cargo

```

1 cargo ← historico[0];

2 para cada cargo c em historico faça
3   se c["date_effective"] == "data desconhecida" então
4     continue;
5   fim
6   se c["date_effective"] > data então
7     break;
8   fim
9   cargo ← h;
10 fim

```

Listagem 5.18 – Resultado Consulta 15

```

1 TANIA RAMOS FORTUNA
2 ["ADJUNTO", "4", "07/10/2017"]
3
4 ARI PAULO ZIMMER GASPAR
5 ["Aposentadoria", "07/08/2018"]

```

5.5 Considerações Finais

Ao longo deste capítulo foram apresentadas diversas consultas sobre o banco de dados da abordagem ACERPI e as extensões feitas no Capítulo 4. Foram descritas consultas referente às entidades, as assinaturas das portarias e por fim das progressões e aposentadorias, construindo um histórico da carreira dos servidores.

Todas essas consultas apresentadas visam atingir o objetivo da abordagem ACERPI: permitir uma maior facilidade na busca e filtragem das portarias. Com estas consultas foi possível extrair valor dos dados e agregar conhecimento ao usuário, como por exemplo o histórico de servidores na instituição.

6 CONCLUSÃO

Neste trabalho, foram apresentadas extensões a abordagem ACERPI (SCHMITZ et al., 2021) buscando aumentar a quantidade de metadados e informações para possibilitar consultas mais avançadas. Foram utilizadas expressões regulares para identificar as assinaturas e também o tipo das portarias, capturando trechos chaves. Por fim, foi apresentado uma série de consultas com o objetivo de extrair valor dos dados e facilitar aos usuários buscas nas portarias. Consultas sobre as entidades, sobre as assinaturas e também sobre progressões de carreira e aposentadoria, resultando na construção do histórico da carreira de cada servidor.

Este trabalho contribui adicionando a capacidade de identificar o tipo da portaria e capturar trechos de interesse referentes a progressões e aposentadorias, além de capturar as assinaturas das portarias. As principais contribuições deste trabalho foram expandir a abordagem ACERPI e demonstrar a possibilidade de fazer consultas sobre o seu banco de dados, contribuindo para atingir seu objetivo de facilitar buscas nas portarias. O trabalho também contribuiu criando uma fundação que permite maior facilidade de outras extensões ao estruturar a nova coleção de portarias e permitir que outras expressões regulares sejam facilmente criadas para identificar outros tipos de padrões nas portarias no futuro.

Por fim, alguns trabalhos futuros podem ser considerados para melhorar e expandir este trabalho e a abordagem ACERPI como um todo:

- A adição de mais expressões regulares identificando outros tipos de portarias que também seguem padrões. Possibilitando assim também outras consultas.
- O desenvolvimento de uma interface gráfica e um servidor *backend* estruturando a abordagem como uma aplicação, facilitando as consultas e adicionando uma visualização.

REFERÊNCIAS

- ARORA, R.; AGGARWAL, R. R. Modeling and querying data in mongodb. **International Journal of Scientific and Engineering Research**, v. 4, n. 7, p. 141–144, 2013.
- BRASIL. Lei nº 11.091, de 12 de janeiro de 2005. **Diário Oficial da República Federativa do Brasil**, Brasília, DF, 2005. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2005/lei/111091.htm>.
- BRASIL. Lei nº 12.772, de 28 de dezembro de 2012. **Diário Oficial da República Federativa do Brasil**, Brasília, DF, 2012. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/112772.htm>.
- EICH, L. Acerpi-link: Uma alternativa de resolução de entidades em portarias institucionais utilizando linkagem de registros. 2021.
- MARTINS, J. S. Acerpi-block : aplicação de técnicas de blocagem à abordagem acerpi. 2021.
- MONGODB. The database for modern applications. 2020. Disponível em: <<https://www.mongodb.com/>>.
- MONGODB. What is a Document Database? 2020. Disponível em: <<https://www.mongodb.com/document-databases>>.
- MONGODB. What is NoSQL. 2020. Disponível em: <<https://www.mongodb.com/nosql-explained>>.
- MONGODB. Aggregation Pipeline Stages. 2022. Disponível em: <<https://www.mongodb.com/docs/manual/reference/operator/aggregation-pipeline/>>.
- MONGODB. Databases and Collections. 2022. Disponível em: <<https://www.mongodb.com/docs/manual/core/>>.
- MONGODB. Documents. 2022. Disponível em: <<https://www.mongodb.com/docs/manual/core/document/#documents>>.
- PYTHON. re — Regular expression operations. 2022. Disponível em: <<https://docs.python.org/3/library/re.html>>.
- SCHMITZ, C. et al. Acerpi: An approach for ordinances collection, information extraction and entity resolution. In: **Anais do XXXV Simpósio Brasileiro de Bancos de Dados, SBBD 2021, online, October 4 - October 8, 2021**. [S.l.]: SBC, 2021.
- TANG, X. et al. Regular expression-based reference metadata extraction from the web. In: **2010 IEEE 2nd Symposium on Web Society**. [S.l.: s.n.], 2010. p. 346–350.
- UFRGS. Consulta a Portarias geradas pela Reitoria da UFRGS. 2016. Disponível em: <<https://www1.ufrgs.br/sistemas/sde/gerencia-documentos/index.php/publico/consultar/>>.
- W3C. Extensible Markup Language (XML) 1.0 (Fifth Edition). 2008. Disponível em: <<https://www.w3.org/TR/REC-xml/>>.