

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E
TECNOLOGIA DE SÃO PAULO
CAMPUS GUARULHOS

CLAUDIO SANTOS CABRAL

API PARA CONVERSÃO RETROSPECTIVA DE FICHAS CATALOGRÁFICAS

GUARULHOS

2022

CLAUDIO SANTOS CABRAL

API PARA CONVERSÃO RETROSPECTIVA DE FICHAS CATALOGRÁFICAS

Trabalho de Conclusão de Curso apresentado como exigência parcial para obtenção do diploma do Curso Tecnologia em Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Câmpus Guarulhos.

Orientador: Prof. Dr. Rodrigo Campos Bortoletto.

GUARULHOS

2022

CLAUDIO SANTOS CABRAL

API PARA CONVERSÃO RETROSPECTIVA DE FICHAS CATALOGRÁFICAS

Trabalho de Conclusão de Curso apresentado como exigência parcial para obtenção do diploma do Curso Tecnologia em Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Câmpus Guarulhos.

Orientador: Prof. Dr. Rodrigo Campos Bortoletto.

BANCA EXAMINADORA

Prof. Me. Cléber Silva de Oliveira
IFSP Campus Guarulhos

Prof. Me. Robson Ferreira Lopes
IFSP Campus Guarulhos

AGRADECIMENTOS

Agradeço ao meu orientador Rodrigo Campos Bortoletto pelo suporte técnico, apoio analítico prestado ao longo da fase de desenvolvimento do projeto, sendo sempre prestativo e propondo contribuições significantes para a solução dos problemas encontrados ao longo do caminho.

Agradeço ao Prof. Thiago Schumacher Barcelos, nas orientações prestadas referente ao desenvolvimento da fundamentação da parte escrita do trabalho.

Agradeço a Prof.^a Alexandra Aparecida de Souza, nas orientações referente o desenvolvimento de requisitos funcionais e elaboração da arquitetura do *software* através de diagramas.

Agradeço ao bibliotecário Rafael de Souza Feitosa, pelo apoio na fase de levantamento dos requisitos, com questionários respondidos e participação nas entrevistas, contribuindo com seu conhecimento de biblioteconomia para esse projeto.

Agradeço aos sócios Luiz Menici e Leonardo Mattes da empresa Docs & Bytes Informática por disponibilizarem a amostragem de fichas catalográficas para realização do nosso protótipo.

*"Inovação é driblar a necessidade
com criatividade".*

Gleber Santos

RESUMO

Com a evolução da ciência e tecnologia, os seres humanos passaram a adotar novas alternativas de resolver seus negócios cotidianos e profissionais, por exemplo, realizar compras online, pagar uma fatura pelo aplicativo, consultar saldo da conta bancária pela internet, atividades estas que nos séculos passados eram realizadas de modo manual, pessoas tinham que ir em lojas físicas e bancos para efetuar tais tarefas. A busca desenfreada pelas instituições privadas e públicas em inovar seus produtos, automatizar seus serviços, trazendo otimização e flexibilidade no seu fluxo de trabalho é consequência da concorrência e exigência do consumidor. Com isso, as instituições bibliotecárias também sentiram a necessidade de melhoria contínua no processo de catalogação usufruindo das tecnologias da informação atuais. Instituições bibliotecárias não disponibilizam a totalidade de conteúdo e riqueza de informação dos seus acervos nos catálogos online de acesso público, por limitações na compreensão e uso de recursos tecnológicos no compartilhamento cooperativo. Esta pesquisa aborda sobre o processo de conversão retrospectiva de fichas catalográficas para a migração dos catálogos online em detrimento dos catálogos bibliográficos impressos em fichas catalográficas. O trabalho apresenta um protótipo funcional que realiza a conversão de fichas catalográficas a partir do reconhecimento óptico dos caracteres das imagens das fichas e extração dos metadados selecionados para realizar a catalogação dos livros.

Palavras-chave: Conversão retrospectiva. Fichas. API. Tecnologia. Catalogação.

ABSTRACT

With the evolution of science and technology, human beings began to adopt new alternatives to solve their daily and professional business, for example, make online purchases, pay an invoice through the application, check bank account balance on the internet, activities that in the centuries in the past were performed manually, people had to go to physical stores and banks to perform such tasks. The unbridled search by private and public institutions to innovate their products, automate their services, bringing optimization and flexibility to their workflow is a consequence of competition and consumer demand. As a result, library institutions also felt the need for continuous improvement in the cataloging process, taking advantage of current information technologies, library institutions do not provide the entire content and wealth of information from their collections in online catalogs with public access, due to limitations in understanding and use of technological resources in cooperative sharing, this research addresses the process of retrospective conversion of catalog cards for the migration of online catalogs to the detriment of bibliographic catalogs printed on typed cards. The proposal is to present a functional prototype of an API that performs the conversion of catalog cards from the optical recognition of the characters of the images of the cards and extraction of the selected metadata to carry out the cataloging of the books.

Keywords: Retrospective Conversion. Files. API. Technology. Cataloguing.

LISTA DE ILUSTRAÇÕES

| | |
|--|----|
| Figura 1- Modelo ficha..... | 25 |
| Figura 2 - Pré-processamento ficha | 25 |
| Figura 3 - Carregamento de dados | 26 |
| Figura 4 - Tela home | 26 |
| Figura 5 - Tela de cadastro | 27 |
| Figura 6 - Tela de <i>login</i> | 27 |
| Figura 7 - Tela do administrador | 28 |
| Figura 8 - API Rest..... | 28 |
| Figura 9 - Diagrama de caso de uso | 29 |
| Figura 10 - Diagrama de classe de domínio..... | 42 |
| Figura 11- <i>Login</i> de usuário..... | 43 |
| Figura 12 - Cadastro de usuário..... | 43 |
| Figura 13 - Consultar ficha na API | 44 |

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1 - Requisitos do usuário..... | 37 |
| Tabela 2 - Planejamento das <i>sprints</i> | 38 |
| Tabela 3 - Gestão de riscos | 40 |

SUMÁRIO

| | |
|--|-----------|
| 1 INTRODUÇÃO..... | 12 |
| 2 OBJETIVOS..... | 14 |
| 2.1 Objetivo Geral..... | 14 |
| 2.2 Objetivo Específicos..... | 14 |
| 3 FUNDAMENTAÇÃO TEÓRICA..... | 15 |
| 3.1 Fichas Catalográficas e o Código de Catalogação Anglo-Americano - AACR2..... | 15 |
| 3.2 Processo de Conversão Retrospectiva e a Catalogação Cooperativa..... | 16 |
| 3.3 Catálogos <i>Online</i>..... | 17 |
| 4 METODOLOGIA DE PESQUISA..... | 18 |
| 5 METODOLOGIA DE DESENVOLVIMENTO..... | 19 |
| 5.1 Metodologia Ágil e as Técnicas de Gestão de Desenvolvimento de Software..... | 19 |
| 5.2 Tecnologias..... | 20 |
| 5.2.1 Reconhecimento óptico de caracteres..... | 20 |
| 5.2.2 Figma..... | 21 |
| 5.2.3 Ambiente de desenvolvimento integrado..... | 22 |
| 5.2.4 <i>Frameworks</i> | 22 |
| 5.2.5 Biblioteca de expressão regular..... | 23 |
| 5.2.6 Banco de dados relacional..... | 23 |
| 5.3 Prototipação..... | 24 |
| 5.3.1 Coleta de fichas..... | 24 |
| 5.3.2 Agrupamento de fichas e tratamento dos seus dados..... | 24 |
| 5.3.3 Criação de bash shell script..... | 25 |
| 5.3.4 Apresentação do sistema - tela <i>home</i> | 26 |
| 5.3.5 Apresentação do sistema - tela cadastro..... | 26 |
| 5.3.6 Apresentação do sistema - tela <i>login</i> | 27 |
| 5.3.7 Apresentação do sistema - tela do administrador..... | 28 |
| 5.3.8 Apresentação do sistema - API Rest..... | 28 |
| 6 FUNCIONALIDADES..... | 29 |
| 7 CONSIDERAÇÕES FINAIS..... | 30 |
| REFERÊNCIAS..... | 31 |
| APÊNDICE A - Questionário – projeto de conversão retrospectiva de fichas catalográficas para bibliotecas..... | 33 |
| APÊNDICE B - Histórias dos usuários..... | 37 |

1 INTRODUÇÃO

Em meados do terceiro milênio a.C houve a necessidade de organizar registros bibliográficos, naquela época as tecnologias utilizadas pelos babilônicos e egípcios eram tábuas de argila, papiros e pergaminhos, nesses tabletes eram identificados o número do tablete em uma série, o título e o escriba. Já no século IV, na idade média em Monte Cassino no sudeste da Roma, monges e freiras copistas eram empenhados em produzir livros manuscritos na sala dos mosteiros também conhecida como *scriptorium*, primordialmente no renascimento do século XII, a produção dessas coleções de livros foi descentralizada por meio da origem de universidades e instituições eclesiásticas. Além disso, uma das primeiras iniciativas na sistematização para a construção de catálogos emergiu também durante a civilização grega por meio da elaboração dos *pinakes*, instrumentos de estruturação de registros bibliográficos, feitos por Calímaco nas bibliotecas de Alexandria e Pérgamo, esse marco histórico representa uma das maiores contribuições dignas para o esquema metódico de catálogos da época. “Um processo de representação documentária que desde a antiguidade atua como instrumento de acesso à informação e ao documento” (PEREIRA; SANTOS, 1998, p. 123).

No Brasil, a iniciativa de identificar coleção de livros por meio de fichas datilografadas, surge segundo relatos que aconteceram durante uma conversa entre o poeta modernista Mario de Andrade e seu primo Pio Lourenço Corrêa, fazendeiro e intelectual de Araraquara, que lhe deu conselhos sobre sistematização de estudos em fichas. Na década de 1920 são datadas as primeiras fichas no fichário analítico de Mario de Andrade criado há mais de duas décadas, o intuito de Mario, era prover para as gerações futuras o acesso público de seus acervos construídos no percorrer de um trabalho minucioso da organização de seus materiais.

Conforme o avanço dos meios de telecomunicações, instituições públicas e privadas discutem no que tange aplicações que envolvem inovação e transformação no fluxo de seus produtos e serviços em seus acervos, dessa forma, surgem soluções tecnológicas para auxiliarem na busca iminente à informação e manipulação desses dados, viabilizando a entrega computadorizada aos interesses do usuário final. Em 1980, catalogadores brasileiros sentiram a necessidade de aderir o processo de conversão retrospectiva RECON (*Retrospective Conversion*) de fichas catalográficas, justamente pela problemática em registrar os metadados em formato eletrônico, nesse

trônico, nesse sentido, por meio dessa implantação, bibliotecas puderam realizar procedimentos e rotinas de trabalho automatizadas. Oliveira *et al.* (1998, p.42) destacam que:

A escolha de um método eficaz que seja adequado às características e necessidades das bibliotecas e que assegure a conversão de catálogos manuais para informatizados, com prazos de execução razoáveis, custo baixo e garantindo a qualidade das informações, são critérios que devem prevalecer na definição de um modelo de conversão retrospectiva.

Haja vista que, ocorreram interesses concernentes para o aproveitamento da catalogação cooperativa, que visa a conexão centralizada de redes cooperativas para o compartilhamento de registros bibliográficos entre bibliotecas, para posterior análises das necessidades e exigências desejáveis das instituições mantenedoras em relação a duplicação de tarefas e redução de custo na catalogação. No entanto, bibliotecas não disponibilizam a totalidade de materiais de seus acervos nas suas fontes de pesquisa conhecidas como catálogos online OPAC (*Online Public Access Catalogue*), devido à ocorrência dos grandes volumes de registros não processados, fazendo proveito sobretudo de catálogos manuais confeccionados em fichas através de máquinas datilográficas.

Desse modo, este trabalho tem por objetivo desenvolver um *software* que promova o reaproveitamento integral dos dados das fichas catalográficas por meio do reconhecimento óptico de caracteres durante a fase de migração dos catálogos manuais para o catálogo online, os quais passarão por um algoritmo alimentado de expressões regulares (*Regular Expression*) para mapeamento e captura dos metadados que selecionados serão lançados na base local do sistema Alexandria que disponibiliza catálogos online para seus usuários.

2 OBJETIVOS

2.1 Objetivo Geral

Desenvolver e validar um programa que proporcione o processo de conversão retrospectiva de fichas catalográficas.

2.2 Objetivos Específicos

- Realizar pesquisas de trabalhos anteriores que abordam quais metodologias e ferramentas de desenvolvimento de software aplicadas para a modularização e codificação de programas capazes de efetuar o procedimento de conversão retrospectiva em fichas datilografadas.
- Efetuar ensaios periódicos dos lotes de fichas submetidos às fases de execução do *software*.
- Entregar e validar um *software* funcional que execute a extração de dados selecionados parametricamente para inserção desses dados numa base local, para serem consultados nas fontes de pesquisa do sistema integrado de gerenciamento de bibliotecas Alexandria.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 Fichas Catalográficas e o Código de Catalogação Anglo-Americano - AACR2

No século XIX, a descentralização de catálogos manuais compostos por fichas catalográficas no formato 12cm x 7.50cm, impulsionou a necessidade de implementar normas para elaboração de fichas na representação descritiva de cada produto em específico. Segundo Mey (1995), com a ascensão de bibliotecas públicas no período da Revolução Francesa, o governo francês adotou regras e normas para que os materiais pudessem ser confiscados, nessa época não existia um padrão universal oficializado, então para organizar registros bibliográficos diversas entidades adotaram normas próprias. Com a realização da conferência pela UNESCO em Paris no século XX, ocorreram discussões de princípios e fundamentos da padronização na catalogação descritiva, o resultado foi o lançamento em pauta da norma AACR (Anglo-American Cataloging Rules) em 1967. Mey (1995, p.23) afirma que:

[...] vem mudar a perspectiva dos códigos: a impressão e venda de fichas catalográficas pela Library of Congress dos Estados Unidos (LC).

Ao invés de cada biblioteca fazer sua própria catalogação de seus livros, a LC passou a vender cabeçalhos, também por ela indicados.

Em 1976, foi apresentado o ISBD (*International Standard Bibliographic Description*) na Reunião Internacional de Especialistas em Catalogação (RIEC), cujo intuito foi contribuir para a formalização sistemática da escrita no processo da descrição bibliográfica e das definições de pontuações nas fichas catalográficas, propondo a facilidade no intercâmbio de dados e registros de diversos materiais distintos. Segundo Campello (1997, p.61) essas descrições são compostas de oito campos:

- 1) título e responsabilidade;
- 2) edição;
- 3) detalhes específicos de material (não vale para livros);
- 4) dados de publicação (local, editora, data);
- 5) descrição física;
- 6) série;
- 7) notas;
- 8) ISBN.

Desse modo, extinguiu-se a era dos padrões nacionalistas, na relevância do ISBD para o aperfeiçoamento da norma AACR2 traduzido para 25 países.

3.2 Processo de Conversão Retrospectiva e a Catalogação Cooperativa

Conversão retrospectiva, do inglês *retrospective conversion*, é a atividade de converter fichas manuscritas em registros por linhas de computador, é recurso tecnológico de compartilhamento cooperativo que permite a totalidade das bases de dados para o acesso e intercâmbio de informação (GREGOR, 1984; REED-SCOTT, 1984), seu propósito é diminuir a propagação dos catálogos manuais e difundir a evolução dos catálogos online OPAC.

Eles são variadamente designados por catálogos de computador (computer catalogs), catálogos online (online catalogs), catálogos de fichas automatizados (automated card catalogs), catálogos de acesso de cliente (patron Access catalogs), ou catálogo em linha de acesso público (online público Access catalogs), sendo este último o mais adotado. São conhecidos também por siglas como OLC, PAC e OPAC (HILDRETH, 1985, p.233-285).

No século XX, bibliotecas que fizeram parte da CSIC (*Unidad de Coordinación de Bibliotecas*) na Espanha e da universidade de Lancaster no Reino Unido, adotaram técnicas e métodos para a implementação do processo de conversão retrospectiva de seus acervos. Conforme Dwyer (1987) o processo de conversão retrospectiva propicia na fonte de pesquisa o acesso bibliográfico, que completa a disponibilidade dos materiais em seus acervos nos catálogos online.

No Brasil desde 1989, a Universidade Estadual de Campinas (UNICAMP) possui o sistema de bibliotecas da UNICAMP (SBU) integrado nas redes cooperativas BIBLIODATA/CALCO administrada pela Fundação Getúlio Vargas (FGV) sendo reconhecido como um case de sucesso de compartilhamento de recursos. Nessa época houve muitas reclamações dos usuários justamente pelo sistema não garantir a totalidade de informação em seus catálogos online, para sanar esse problema a UNICAMP juntamente com o SBU iniciaram um projeto de transformação dos registros bibliográficos do acervo que passaram pelo procedimento de conversão retrospectiva de fichas para completar a base de dados da UNICAMP (ACERVUS) que provê as fontes de pesquisa em seus catálogos online, e que necessitavam fornecer a completude de informações dispostas no acervo físico.

Para isso, foram analisados *softwares* para conversão retrospectiva, o software escolhido foi o CatCD para *Windows* da OCLC (*Online Computer Library Center*),

com objetivo de capturar, exportar e tratar os dados bibliográficos em formato MARC (*Machine Readable Cataloging*),¹ desenvolvido nos Estados Unidos na década de 60, para compartilhar registros de livros e teses pelas redes cooperativas, o formato MARC é a adoção digital dotada para subsidiar a interoperabilidade com outros sistemas e mecanismos de busca entre as bibliotecas. Vale ressaltar que o CatCD disponibilizado em um CD-ROM compreende a conversão secundária, isto é, existe a conversão primária, e por conseguinte a conversão secundária, que trata do formato MARC.

Este trabalho desenvolveu um protótipo funcional que realiza somente a conversão primária, que propõe a digitalização e captura dos registros das fichas datilográficas manuscritas ou impressas que são armazenadas em uma base de dados para serem consultadas em um sistema web de biblioteca.

3.3 Catálogos Online

A disposição de catálogos online surge com o intuito de suprir a defasagem de informação que acontece nos catálogos manuais, que possuem ineficiência no resgate de informação devido às suas desatualizações. Com o advento da automação de acervos, a rastreabilidade dos registros são mais assertivas e promovem a completude dos dados durante a fase de recuperação da informação, em decorrência do compartilhamento e cooperação entre as bibliotecas, que favorecem a disponibilidade via Internet dos catálogos virtuais para um melhor aproveitamento no atendimento do usuário.

No Brasil, os sistemas de gerenciamento de bibliotecas da Universidade de São Paulo (USP) e Universidade Estadual de Campinas (UNICAMP) concedem o acesso aos catálogos virtuais utilizando a consulta pelo nome do autor, título e assunto, dessa forma, o usuário tem conhecimento do que a biblioteca dispõe em seu acervo. Sendo assim, o usuário pode realizar um empréstimo, reservar o livro, ou seja, isso traz benefícios para os usuários no processo de aquisição, tempo e ponto de acesso.

1 "Oliveira et al. - Compact Disc Cataloging - CatCD análise de um in.txt".

4 METODOLOGIA DE PESQUISA

Para o desenvolvimento deste trabalho foi adotada a utilização da pesquisa exploratória que visa fazer o levantamento bibliográfico de dados relacionados ao escopo do objeto desse estudo, de modo que, o pesquisador se empenhe em coletar referências que abordem sobre o processo de conversão retrospectiva de fichas catalográficas. A maneira como essas informações foram coletadas foi por meio de livros, artigos, revistas científicas, sites, vídeos e trabalhos acadêmicos anteriores, o conjunto dos dados coletados serviram de base para compreensão do tema proposto.

Além disso, foi desenvolvida uma pesquisa qualitativa, que teve como base o estudo de análises e concepções do cliente, para o colhimento das respostas em relação ao comportamento humano no que tange a usabilidade dos mecanismos de catalogação empírica e a adaptação das novas tecnologias empregadas no processo de transformação e automação dos serviços bibliotecários foi enviado um questionário fechado desenvolvido no *Google Forms* para coleta de informações, logo foi direcionado para o bibliotecário do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo do Campus de Guarulhos (IFSP-GRU) que foi o cliente do projeto responder o formulário. Também foi enviado um termo de consentimento livre esclarecido, para registrar a concordância do bibliotecário em contribuir com seus conhecimentos de biblioteconomia para este projeto. Por fim, após respondido tal questionário, foi realizado uma entrevista de alinhamento para discussão das respectivas respostas, isso trouxe sustentação na definição dos requisitos usuários e requisitos do sistema para o desenvolvimento do *software*.

5 METODOLOGIA DE DESENVOLVIMENTO

5.1 Metodologia Ágil e as Técnicas de Gestão de Desenvolvimento de Software

Na década de 90, o termo ágil surge em uma conferência entre 17 desenvolvedores que conversavam sobre novas práticas e valores na gestão de projetos para o desenvolvimento de software em pequenas, médias e grandes organizações (Beck *et al.*, 2001). Em decorrência dos movimentos agilistas, originou-se doze manifestos, que se opuseram a *mindset* (mentalidade) de metodologias tradicionais fundamentadas em processos de documentação rígidas e mais controladas. Todavia, a metodologia ágil não descarta os processos, ferramentas, documentação, contratos e planejamento, sua diferença está na priorização das interações entre os indivíduos envolvidos no projeto e entregas periódicas de *software* funcional, cuja preocupação é reduzir o tempo na documentação extensiva e gastar mais empenho na implementação, viabilizando assim o modelo iterativo e incremental.

Nesse sentido, neste projeto foi utilizado o *framework* Scrum para gerenciar o ciclo de desenvolvimento do software, que teve sua origem na gestão de projetos no sistema Toyota para a produção de automóveis e na estratégia militar para tomada de decisões rápidas OODA (*Observe, Orient, Decide and Act*) utilizado pelas Forças Aéreas dos Estados Unidos, sendo também adotado para gestão de construção de *softwares*. Adotou-se esse *framework* justamente pela aplicação de métodos iterativos e incrementais que visa o refinamento sucessivo do programa até que apresente um nível de aceitação no decorrer do trabalho dos critérios de aceite das funcionalidades desejadas pelo cliente durante os cenários de testes.

Scrum tem sido usado para desenvolver software, hardware, software embarcado, redes de funções interativas, veículos autônomos, escolas, governo, marketing, gerenciar a operação da organização e quase tudo que usamos em nosso dia-dia nas nossas vidas, como indivíduos e sociedades. [...] Scrum demonstra efetividade especialmente na transferência de conhecimento iterativo e incremental. Scrum é agora amplamente usado para produtos, serviços e no gerenciamento da própria empresa (SCHWABER; SUTHERLAND, 2017, p.4).

No pré-planejamento será feito o levantamento de requisitos funcionais e não funcionais por meio das histórias dos usuários, requisitos estes que serão documentados no *backlog* do produto e *backlog* da *sprint*, artefatos utilizados para

garantir a transparência e comunicação entre o time *Scrum*, levando em consideração suas prioridades, prazos de entrega das tarefas, definição da equipe e das tecnologias no projeto.

Neste projeto o time *Scrum* é constituído pelo *Product Owner*, responsável em interagir com o bibliotecário e ao mesmo tempo compreender as suas necessidades e desejos em adquirir um software para o seu modelo de negócio, o *Product Owner* e *Development Team*, neste caso foi o aluno, outro membro importante é o *Scrum Master*, que foi o orientador desta pesquisa responsável em gerenciar e acompanhar o ciclo do desenvolvimento propondo soluções e auxiliando na mitigação dos possíveis riscos do sistema. Em suma, realizaremos a *Sprint Planning* para definir as tarefas a serem entregues para o escopo de cada *sprint* seguindo seus prazos de acordo o cronograma de gestão de projetos desenvolvido. Desse modo, foram feitas reuniões semanais de 15 minutos para averiguar pontos de sucesso, fracasso e ocorrência de impedimentos no decorrer do desenvolvimento, a reunião é realizada juntamente com o *Team* e *Scrum Master*.

5.2 Tecnologias

Neste tópico, serão apresentadas as ferramentas de *software* necessárias para o desenvolvimento do sistema.

5.2.1 Reconhecimento óptico de caracteres

Originado no início da década de 60, marco também dos primeiros computadores digitais, surge a criação de máquinas OCR (*Optical Character Recognition*), sub área da visão computacional que estuda a identificação automática de textos impressos ou escritos à mão nos documentos digitalizados, viabilizando a extração de todo o conteúdo textual em formato ASCII (*American Standard Code for Information Interchange*) para análise computacional. Nessa época as máquinas eram muito limitadas, reconheciam algumas letras e se restringiam a símbolos, outro problema era a complexidade de garantir alta qualidade, precisão na detecção dos caracteres em documentos mais antigos.

Com melhorias contínuas aplicadas no reconhecimento óptico de modelos não estruturados, tem-se difundido essa prática suprimindo a necessidade de diversificados nichos de mercado, como por exemplo, nos correios, em instituições bancárias e

sobretudo, ganhou popularidade na automação de acervos, visando a manutenção preventiva em formato digital dos dados de várias obras raras, mitigando o risco de perder conjuntos de dados de cada obra que posteriormente poderiam sofrer incidentes inesperados, tais como: incêndios, furtos, desgaste do documento ao longo dos anos. Contudo, existem condições adversas que impactam na resolução do OCR, isto é, determinados caracteres causam imprecisão no seu reconhecimento, como por exemplo, distinguir a letra “o” do numeral “0” (BOROVIKOV, 2014).

A plataforma *Amazon Web Services* no seu portfólio orientado a prestação de serviços, disponibiliza quatro API's da *Amazon Textract* para detectar e analisar diversificados tipos de documentos digitalizados. De acordo com a Amazon Web Services (2002), *Amazon Textract* é um serviço de OCR que extrai automaticamente documentos manuscritos e impressos estruturados e não estruturados, possui uma versão gratuita limitada e dispõe no guia do desenvolvedor o código-fonte da API *Detect Document Text* desenvolvida na linguagem de programação Python que foi utilizado neste projeto para reconhecimento dos caracteres em texto corrido, o serviço também possui outras API's que são pagas para analisar e capturar dados de outros documentos que possuem uma estrutura diferente de um texto corrido, como por exemplo, formulários, tabelas, documentos de identidade como carteira de motorista, R.G, passaporte, entretanto, permite a integração com o *Amazon S3* que foi utilizado como repositório em nuvem para armazenar a imagem das fichas catalográficas.

5.2.2 Figma

É uma plataforma amigável que tem a finalidade de trazer para as equipes de desenvolvimento a criação visual de um produto, seja uma aplicação mobile ou web e serve de ilustração das interfaces *front-end* para o cliente. Com o Figma, pode-se criar um protótipo de design navegável usando o recurso de prototipação disponibilizado pela ferramenta.

A plataforma possui uma versão gratuita e uma versão paga, utilizou-se a versão livre que foi o suficiente para o desenvolvimento visual do sistema, além do que o Figma funciona em qualquer sistema operacional pelo navegador web ou computador, realizando apenas o *download* do pacote de instalação no próprio site da plataforma.

5.2.3 Ambiente de desenvolvimento integrado

O PyCharm é um ambiente de desenvolvimento integrado (IDE) para a linguagem de programação Python, oferece a versão *Community* (gratuita) e *Professional* (paga), como o PyCharm disponibiliza um contrato de licença para estudantes matriculados em universidades. Optou-se em utilizar a IDE *Professional* usufruindo de recursos que a versão *Community* não possui, por exemplo, com a versão *Professional* é possível, ter suporte a banco de dados, acesso a ferramentas de big data, suporte ao HTML5, CSS e JS, efetuar configurações remotas, implantação e fazer um desenvolvimento web usando o *Framework* Django apenas criando um projeto novo na IDE, sem precisar baixar o *Framework* em seu site oficial e criar um ambiente virtual (env) para baixar suas dependências, o que é necessário realizar em caso de uso da versão *Community*, já que não tem suporte para desenvolvimento de estrutura web. Ademais, o PyCharm tem suporte para o python2 e python3, diferente de algumas linguagens de programação que tem compatibilidade com suas versões anteriores e atuais, o python não é retrocompatível, neste projeto utilizou-se o python3, pois no seu lançamento em 2008 foram feitas retificações de design que a linguagem não trazia em versões anteriores.

5.2.4 Frameworks

Durante a fase de desenvolvimento de um projeto, programas desenvolvidos em linguagem de programação pura, exigem uma sobrecarga exaustiva de estudo e trabalho para o programador, resultando na prorrogação dos prazos de entrega e improdutividade na codificação. Com o objetivo de sanar esses problemas, foram criados *middlewares*, programas de configuração, que fornecem uma infraestrutura pré-programada preparada no ambiente de desenvolvimento, essa estrutura pré-configurada orientada a objetos é chamada de *framework*, voltada para ganhos de produtividade, reutilização de código e injeção de dependência.

Existem vários frameworks desenvolvidos em python, no entanto, neste projeto utilizou-se o Django, que estimula o *clean code*, separação dos módulos da estrutura e facilidades de manutenibilidade em aplicações web e “torna tarefas comuns do desenvolvimento web rápidas e fáceis” (DJANGO SOFTWARE FOUNDATION, 2017, p. 5), para tanto, compõe tratamento de exceções, requisições HTTP, mapeamento de objeto-relacional também. Em síntese, o Django suporta os modelos de arquitetura

de software MVC e MVT, neste projeto utilizou-se o padrão MVT (*Model-View-Templates*), porque a parte do Controlador é gerenciada pelo próprio Django, diferente do MVC em que o desenvolvedor escreve o código do módulo *Controller*.

Além disso, utilizou-se o *Bootstrap*, *framework front end* que dispõe a maioria dos seus recursos gratuitamente para o desenvolvimento *web* de um site responsivo e interface amigável que possui suporte para todos os navegadores de internet populares, é uma estrutura popular composta de componentes do HTML5, CSS e Javascript, que contribuem para a simplificação de páginas *web's* dinâmicas ou estáticas responsivas.

5.2.5 Biblioteca de expressão regular

Expressões regulares, do inglês *regular expression* ganharam renome no trabalho do matemático norte-americano Stephen Cole Kleene conhecida como álgebra de conjuntos regulares, que conseqüentemente resultou na criação da biblioteca *regex* por Henry Spencer. Em meados de 1968 foi desenvolvido um editor de texto constituído de expressões regulares, comandos como *grep* para sistemas baseados em Unix que compreendem o uso de ER (*Regular Expression*) (JARGAS, 2012). Sendo assim, expressões regulares são padrões algébricos determinísticos desenvolvidos em metacaracteres ou funções especiais para identificar padrões de um texto, utilizando uma linguagem de programação que possui suporte de bibliotecas *regex*, nesse projeto a linguagem de programação utilizada para o desenvolvimento das expressões foi o *Python*. Na Hierarquia de Chomsky as linguagens regulares são mais fáceis de entender, facultando algoritmos de menor complexidade (ROZENBERG; SALOMAA, 1997). Em suma, foi utilizado neste projeto o módulo do python que promove uma micro linguagem embutida composta de funções capazes de identificar, substituir um determinado objeto, de acordo a expressão regular aplicada pelo desenvolvedor.

5.2.6 Banco de dados relacional

O Banco de dados relacional MySQL foi desenvolvido em 1994 pela empresa sueca chamada MYSQL AB, em 2008 a companhia norte-americana Sun Microsystems compra o MYSQL AB, que posteriormente foi vendido para a Oracle em 2010. Os bancos de dados são sistemas computadorizados que armazenam e

atualizam um volume de dados para os usuários consultarem uma determinada informação no seu modelo de negócio.

Neste projeto utilizou-se o servidor de banco de dados MySQL e a IDE DBeaver para realizar o desenvolvimento das tabelas, as tabelas foram responsáveis em guardar os registros dos campos das fichas na base de dados, além do que optou-se em utilizá-lo por ser um *software* gratuito para uso não comercial e de código-aberto, uma de suas vantagens é que a versão gratuita do MySQL não se limita às suas funcionalidades.

5.3 Prototipação

Este tópico aborda os procedimentos adotados para o desenvolvimento da API de conversão retrospectiva de fichas catalográficas e sua demonstração.

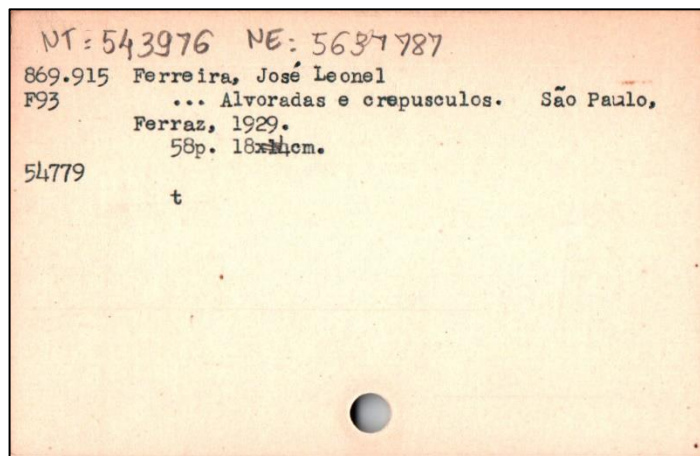
5.3.1 Coleta de fichas

Em primeiro lugar, ocorreu o auxílio da empresa Docs & Bytes Informatica que presta serviço para bibliotecas de São Paulo, para a realização da coleta de fichas catalográficas para amostragem e ensaios de teste do projeto. Nesse sentido, a empresa Docs & Bytes Informatica solicitou para um de seus clientes a aquisição de amostras de fichas para contribuição desta pesquisa, com isso as fichas de papel foram digitalizadas e armazenadas no formato .jpg em um diretório local. Posteriormente, recebemos o diretório compactado por e-mail e exportamos essas fichas no repositório s3 da Amazon.

5.3.2 Agrupamento de fichas e tratamento dos seus dados

Para implementação das expressões regulares no código, sobretudo, foi necessário fazer uma análise dos textos que contém nas fichas tendo como base a norma AACR2 (Código de Catalogação Anglo Americano), então adquiriu-se uma amostra de 119 fichas para identificar determinados padrões e distinguir seus subgrupos. Para tanto, foi implementado estratégias de extração no regex diferentes para cada configuração de subgrupo de ficha, para um melhor resultado esperado do nosso programa e após migrar os dados tratados no *database*, esses dados foram colocados em um modelo de acurácia para verificar e validar sua assertividade. A figura abaixo, mostra o formato da ficha.

Figura 1- Modelo ficha



Fonte: Elaborado pelo autor.

5.3.3 Criação de bash shell script

Para execução do *software* que faz a seleção e mapeamento das fichas utilizando Regex, foi desenvolvido um bash shell script chamado main.sh que executa o *software* automaticamente e após persiste os registros coletados no Banco de Dados para consumo da API. No terminal aparece o carregamento da ficha candidata e o seu pré-processamento, conforme ilustra a figura abaixo.

Figura 2 - Pré-processamento ficha

```

claudio@claudio-Inspiron-5558:/home/desenvolvimento$ ./main.sh
NT = 543976 NE: 5637787
869.915 Ferreira, Jose Leonel
F93
Alvoradas e crepusculos.
Sao Paulo,
Ferraz, 1929.
58p. 18x14cm.
54779
t
Contador de Lines na Lista é: 9
1 registro inserido
ficha processada: Catalogado/20210818-000.jpg

```

Fonte: Elaborado pelo autor.

Portanto, para este trabalho foi validada e averiguada a carga da ficha na base de dados chamada fichas na tabela usuarios_fichas.

Figura 3 - Carregamento de dados

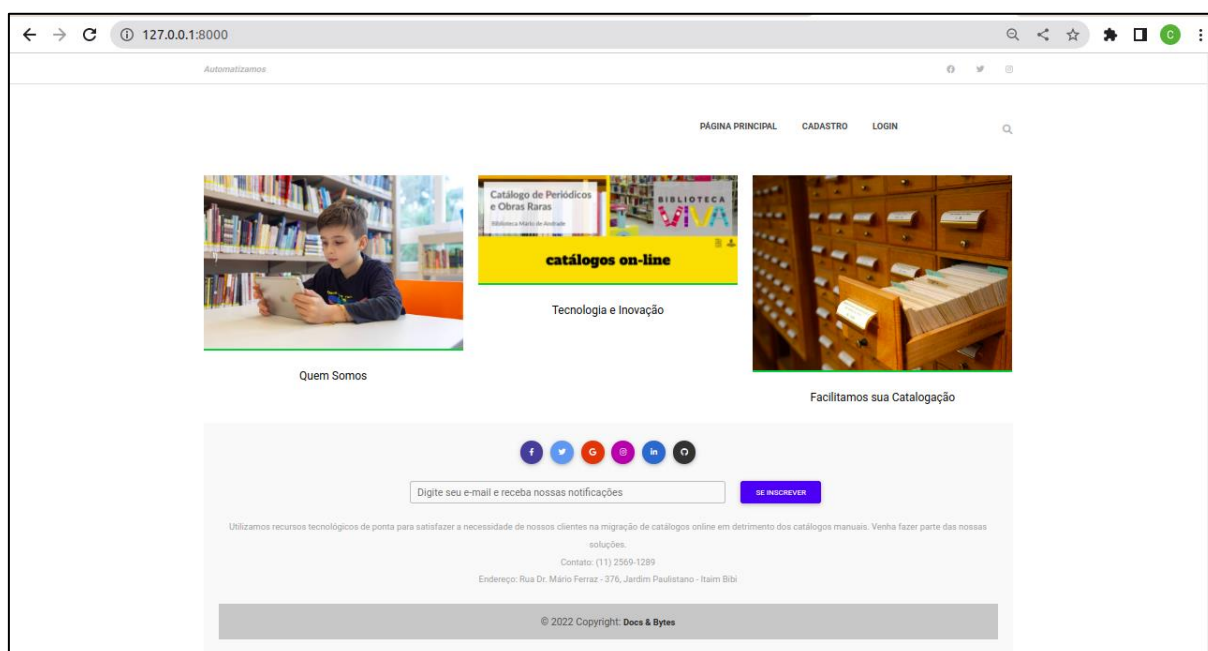
| | id | url | nat | nae | classificacao | autor | notacao | titulo |
|----|----|-------------------------|-----|-----|---------------|-------------------------------|---------|----------|
| 4 | 8 | Modelo/20210818-007.jpg | 0 | 0 | 869.909 | Lins, Alvaro de Barros | L3 | Jornal d |
| 5 | 9 | Modelo/20210818-008.jpg | 0 | 0 | 869.909 | Linhares, Mario | L4 | Historia |
| 6 | 10 | Modelo/20210818-009.jpg | 0 | 0 | 869.909 | Lins, Ivan Monteiro de Barros | L5 | Martins |
| 7 | 11 | Modelo/20210818-010.jpg | 0 | 0 | 869.909 | Lins, Alvaro de Barros | L7 | Jornal d |
| 8 | 12 | Modelo/20210818-011.jpg | 0 | 0 | 869.909 | Lins, Alvaro de Barros | L6 | Jornal d |
| 9 | 14 | Modelo/20210818-013.jpg | 0 | 0 | 869.909 | Lima, Abdias | L9 | Falam o |
| 10 | 15 | Modelo/20210818-014.jpg | 0 | 0 | 869.909 | Lima, Alceu Amoroso | L14 | dos. |
| 11 | 16 | Modelo/20210818-015.jpg | 0 | 0 | 869.909 | Lima, Heitor Ferreira | L15 | Castro A |
| 12 | 17 | Modelo/20210818-016.jpg | 0 | 0 | 869.909 | Lima, Alceu Amoroso | L16 | Tres ens |
| 13 | 18 | Modelo/20210818-017.jpg | 0 | 0 | 869.909 | Lins, Alvaro | L20 | Jornal d |
| 14 | 19 | Modelo/20210818-018.jpg | 0 | 0 | 869.909 | Leme, Jose Dias | L23 | Gustavo |

Fonte: Elaborado pelo autor.

5.3.4 Apresentação do sistema - tela *home*

A tela *home* é a primeira página de interação inicial com o usuário, nessa mesma página o cliente pode inserir seu endereço de e-mail para receber notificações e novidades de nossos serviços.

Figura 4 - Tela home



Fonte: Elaborado pelo autor.

5.3.5 Apresentação do sistema - tela cadastro

A tela de cadastro consiste na autenticação do cliente e regime da política de segurança de acesso a API que, nesse sentido, somente é utilizada por clientes cadastrados.

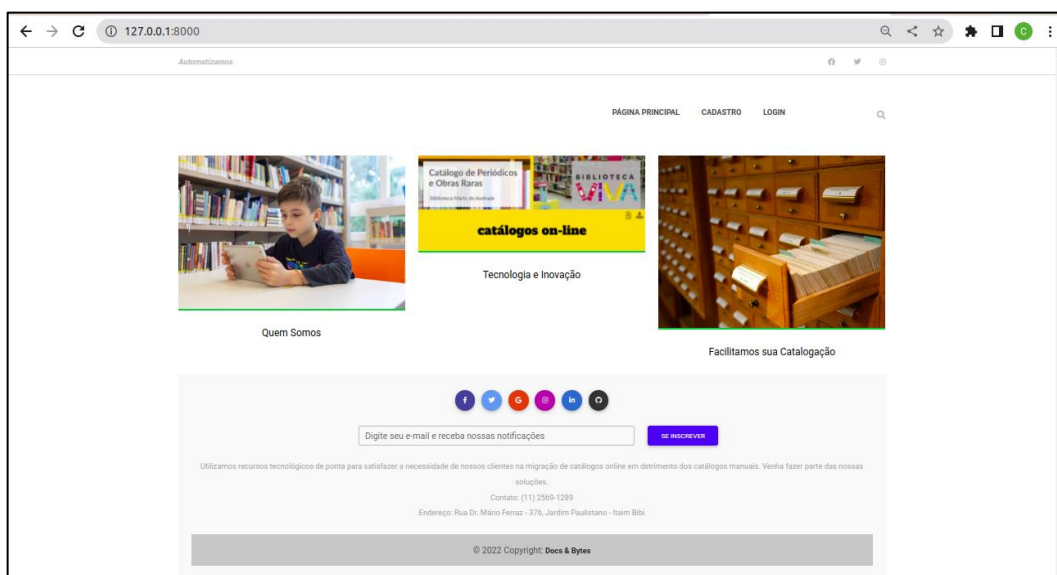
Figura 5 - Tela de cadastro

Fonte: Elaborado pelo autor.

5.3.6 Apresentação do sistema - tela *login*

A tela de *login* consiste no acesso do cliente, que está cadastrado na base de dados da biblioteca e consequentemente poderá usufruir dos serviços da biblioteca.

Figura 6 - Tela de *login*

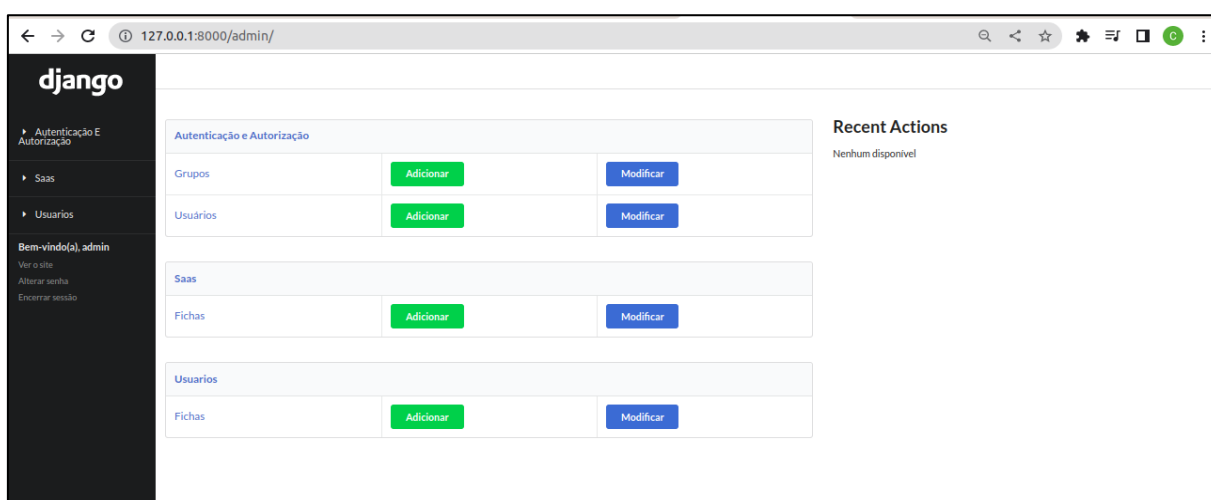


Fonte: Elaborado pelo autor

5.3.7 Apresentação do sistema - tela do administrador

Com o objetivo de controlar as fichas processadas/cadastradas e informações sigilosas de seus clientes, apenas super usuários tem permissão de acesso nessa sessão.

Figura 7 - Tela do administrador

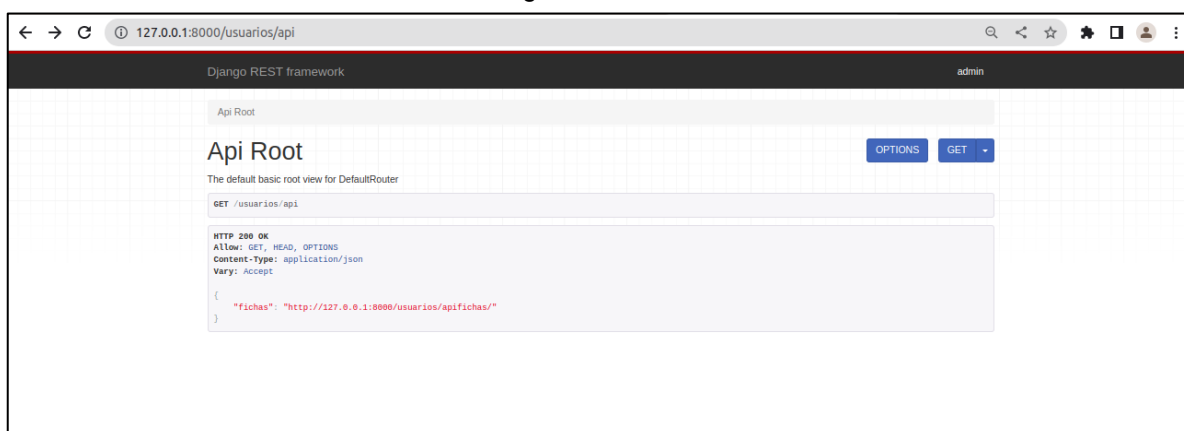


Fonte: Elaborado pelo autor.

5.3.8 Apresentação do sistema - API Rest

Após acesso ao sistema, o bibliotecário digita a url da API e consulta as fichas desejadas.

Figura 8 - API Rest

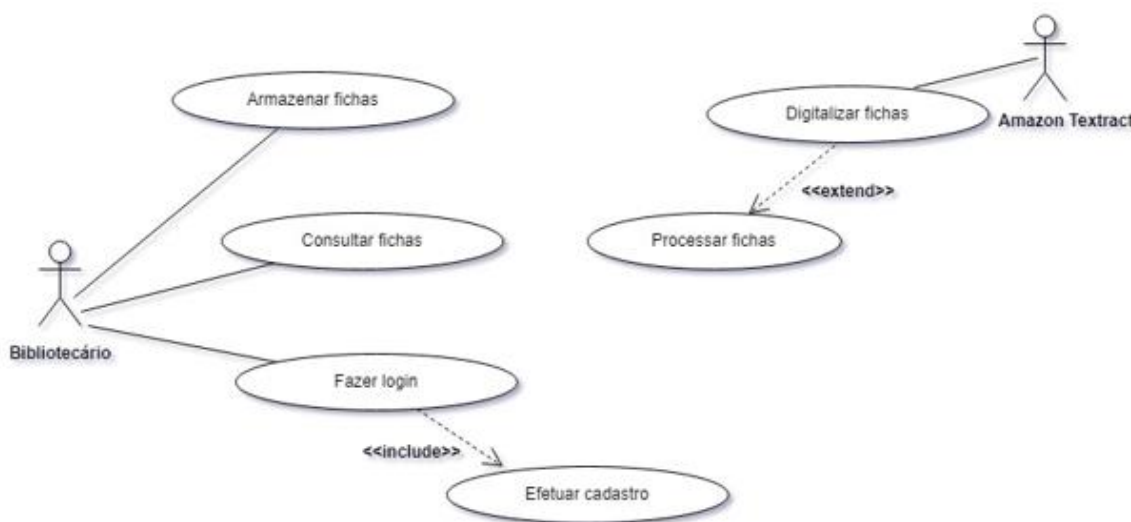


Fonte: Elaborado pelo autor.

6 FUNCIONALIDADES

Abaixo mostra as funcionalidades do sistema por meio do Diagrama de Caso de Uso e Histórias de Usuários criadas para modelagem UML.

Figura 9 - Diagrama de caso de uso



Fonte: Elaborado pelo autor.

O bibliotecário digitaliza as fichas por meio de um digitalizador de documentos, por conseguinte, é gerado um arquivo .jpg da ficha em sua máquina, arquivos estes que foram armazenados no repositório criado no serviço da *Amazon s3*. Para processar as fichas, isto é, converter em formato digital e extrair os dados desejados para iniciar o processo de catalogação, o bibliotecário acessa a página web do site da biblioteca, antes de logar na plataforma é necessário efetuar seu cadastro se for seu primeiro acesso, após logado no sistema, o mesmo deverá procurar por serviços de conversão retrospectiva e catalogação, acessar essa página e clicar no botão processar ficha, que acionará o evento de digitalização e conversão das fichas que estão no repositório da *Amazon s3*, após sua conclusão o sistema exibe uma mensagem de sucesso. Sendo assim, nesse momento será possível catalogar o livro que pertence a ficha processada, o bibliotecário irá procurar na barra de pesquisa pelo autor daquele livro para consultar os dados daquela ficha processada, quando os resultados da ficha forem exibidos na tela, o bibliotecário deverá clicar no botão de catalogar livro.

7 CONSIDERAÇÕES FINAIS

Esse trabalho teve sua iniciativa por meio de um projeto na empresa Docs & Bytes Informática que fomenta e presta serviços de *software* no setor bibliotecário, com intento de otimizar serviços bibliotecários carentes de aplicações automatizadas.

Portanto, foi desenvolvido um protótipo para ilustrar o comportamento de um programa que carrega fichas de um repositório e extrai informações relevantes automaticamente para a efetivação da conversão retrospectiva, fluxo primário da catalogação. Sendo assim, durante o processo de coleta e análise de amostra das fichas foram encontrados subgrupos de um grupo de fichas, dessa forma, aderiu-se estratégias diferentes para captura dos dados desejados de cada ficha tratando uma estrutura de expressões regulares diferente para cada subgrupo encontrado, estimou-se a porcentagem de acertos por campos capturados, e com isso pode-se concluir que a média global de acertos foi de 0,83 o equivalente à 83% dos resultados obtidos, esses acertos estão relacionados ao percentual de acurácia obtido para cada campo capturado dentro da amostragem.

Devido o levantamento da proporção estimada, leva-se em consideração que a proposta do projeto foi cumprida, mas que o *software* precisa de melhorias periódicas, contudo, isso não impacta na ideia inicial do projeto, que enfatiza a extração e tratamento dos registros bibliográficos para serem consumidos por uma API Rest. Em suma, esse trabalho ficará disponível para expansão do ramo de negócio de empresas de tecnologia que prestam serviços e produtos para bibliotecas e instituições diversas que desejam migrar seus catálogos manuais para catálogos online.

REFERÊNCIAS

- ALBUQUERQUE, Ana Cristina de. **Catálogo e descrição de documentos fotográficos em bibliotecas e arquivos**. p. 18.
- ASSUMPÇÃO, Fabricio Silva. **Conversão de registros em XML para MARC 21: um modelo baseado em folhas de estilo XSLT**. p. 137, 2013.
- BOTTA, Andre Luiz Costantino; LEYZA ELMERI BALDO DORINI; MINETTO, Rodrigo. Análise comparativa de métodos de segmentação e técnicas de aprendizado de máquina com aplicação no reconhecimento automático de placas de identificação de veículos. *In: SEMINÁRIO DE INICIAÇÃO CIENTÍFICA E TECNOLÓGICA DA UTFPR*, 20. Campo Mourão, 2015.
- CAMPELLO, B. S. **Introdução ao controle bibliográfico**. 2. ed. Brasília: Briquet de Lemos, 2006.
- CAMPOS, H. **Manual de catalogação do sistema de bibliotecas da universidade federal do Ceará**. p. 173, [s.d.].
- CRISTOVÃO, S. **As relações entre resource description and access (RDA) e o código de catalogação anglo-americano (aacr2): a busca de uma descrição bibliográfica completa**. p. 87, 2014.
- DOGNINI, Marlon José; RAABE, André Luís Alice. Eduling software educacional para linguagens regulares. p. 10, 2003.
- DOGNINI, Marlon José; RAABE, André Luís Alice. **Um software para auxílio a aprendizagem de linguagens regulares**. p. 11, 2003.
- FELTRIM, Valéria de Lisandra. PyGER: **Uma ferramenta geradora de expressões regulares a partir de um conjunto de expressões em linguagem natural**. p. 5,
- FUSCO, Elvis. **Modelos conceituais de dados como parte do processo da catalogação: perspectiva de uso dos FRBR no desenvolvimento de catálogos bibliográficos digitais**. p. 251, 2010.
- GREGOR, D. (comp.). **Retrospective conversion: report of a meeting sponsored by the Council on Library Resources**. Washington, D.C.: CLR, 1984. 140 p.
- HILDRETH, Charles R. **Online Public Access Catalogs**. Annual Review of Information Science and Technology (ARIST), v. 20, p.233-285, 1985.
- JARGAS, Aurelio M. **Expressões regulares: uma abordagem divertida**. 4. ed. São Paulo: Novatec, 2012.
- MODESTO, Fernando. **Panorama da catalogação no Brasil: da década de 30 aos primeiros anos do Século 21**. p. 22, 2007.
- MORAIS, José João Gonçalves. **Obtenção de expressões regulares pequenas a partir de autómatos finitos**. p. 128, 2004.
- MARÇAL, Leonardo; MARCELLE, Bianca. **Código de catalogação anglo-americano - AACR**. 2017. Disponível em: <http://rgdoi.net/10.13140/RG.2.2.13013.93925>. Acesso em: 6 maio 2022.

OLIVEIRA, N. M. *et al.* Compact disc cataloging - CatCD: análise de um instrumento para conversão retrospectiva no Sistema de bibliotecas da. **Perspect. cienc. inf.**, v. 3, n. 1, p. 6, [s.d.].

QUEIROZ, Nathalia Guedes de; ARAUJO, Samantha Andrade de. **Catálogos on-line:** um breve estudo dos catálogos on-line de acesso público (OPAC'S) 1. p. 17.

SOMMERVILLE, I. **Software engineering**. 8. ed. Harlow: Pearson Education, 2007.

SOTTOMAYOR, José Carlos. **Regras portuguesas de catalogação**. p. 5.

VOSGRAU, S. R. C.; SANTOS, A. A. C. Formato MARC 21 holdings para publicações seriadas. p. 15, [s.d.].

ZANAGA, M. P. **Conversão retrospectiva no sistema de bibliotecas da Unicamp:** integração de registros MARC a catálogo on-line. p. 11, [s.d.].

APÊNDICE A - Questionário – projeto de conversão retrospectiva de fichas catalográficas para bibliotecas

- **Cliente:** Rafael de Souza Ramalhães Feitosa - Bibliotecário Documentalista
- Coordenador do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Campus Guarulhos.

- **Descrição do Formulário:** Tem como objetivo colher informações referente o dia a dia de um funcionário de biblioteconomia durante o processo de catalogação. A pesquisa é **CONFIDENCIAL** e as informações coletadas NÃO serão disponibilizadas a nenhuma empresa terceira. Sua participação é essencial para nós! Agradecemos, desde já!

***Obrigatório.**

1) Marque em uma das opções abaixo, qual é a norma de catalogação utilizada pela biblioteca do Instituto Federal de Educação Ciência e Tecnologia de São Paulo? *

- ☐ RDA - Resource Description and Access
☐ AACR2 - Anglo-American Cataloguing Rules

2) A biblioteca, participa de algum intercâmbio / cooperação na catalogação? *

- ☐ Não
☐ Sim

3) A biblioteca oferece acesso aos catálogos online? *

- ☐ Não
☐ Sim

4) Quais dificuldades você encontra no processo de catalogação em sua biblioteca?
*

5) A biblioteca, faz proveito de algum software que efetue a conversão e extração automática dos registros bibliotecários das fichas? Se sim, qual é nome do software?
*

6) Marque nas opções abaixo, quais são os campos do Formato Marc 21 que são trabalhados na catalogação? *

- ☐ 1. ISBN (Número normalizado da obra)
- ☐ 2. Idioma
- ☐ 3. CDU (Código do livro para ser colocado no estante)
- ☐ 4. Autor
- ☐ 5. Título
- ☐ 6. Edição
- ☐ 7. Editorial, lugar de impressão da obra
- ☐ 8. Ano de publicação da obra
- ☐ 9. Quantidade de páginas, ilustrações
- ☐ 10. Título da série (quando faz parte de uma coleção)
- ☐ 11. Notas gerais (informações adicionais ex.: dissertação ou tese, entre outros)
- ☐ 12. Nota de conteúdo
- ☐ 13. Nota de resumo
- ☐ 14. Nota de formato físico adicional disponibilidade
- ☐ 15. Assuntos, temas ou descritores da obra Há em sua opinião, alguma informação relevante que não está nas opções acima?



ÉTICA EM PESQUISA DO INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE SÃO PAULO

Av. Salgado Filho, 3501 – Centro, Guarulhos – SP, CEP: 07115-000 – Telefone: (11) 2304-4250

<http://gru.ifsp.edu.br/>

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Título do Projeto de pesquisa: API para Conversão Retrospectiva de Fichas Catalográficas

Pesquisador Responsável: Claudio Santos Cabral

Nome do participante: Rafael de Souza Ramalhães Feitosa

Você está sendo convidado (a) para ser participante do Projeto de pesquisa intitulado “API para Conversão Retrospectiva de Fichas Catalográficas” de responsabilidade do (a) pesquisador (a) Claudio Santos Cabral. Leia cuidadosamente o que se segue e pergunte sobre qualquer dúvida que você tiver. Caso se sinta esclarecido (a) sobre as informações que estão neste Termo e aceite fazer parte do estudo, peço que assine ao final deste documento, em duas vias, sendo uma via sua e a outra do pesquisador responsável pela pesquisa. Saiba que você tem total direito de não querer participar.

1. A participação nesta pesquisa consistirá em ... *(detalhe aqui a metodologia da pesquisa de com linguagem acessível e compreensível ao público-alvo incluindo a frequência e; tempo de duração de cada encontro, número de vezes que deverá retornar, as condições do local/infraestrutura onde será realizada a coleta de dados e quem as fará, entre outras informações relevantes ao participante).*
2. Durante a execução da pesquisa poderão ocorrer riscos de ... *(descrever os possíveis riscos e as formas pelas quais serão minimizados cada um deles).*
3. Os benefícios com a participação nesta pesquisa serão... *(descrever o(s) benefício(s) diretos (para o próprio participante) e/ou indiretos (para a comunidade e/ou outras pessoas) que a pesquisa trará);*
4. Os participantes não terão nenhuma despesa ao participar da pesquisa e poderão retirar sua concordância na continuidade da pesquisa a qualquer momento.
5. Não há nenhum valor econômico a receber ou a pagar aos voluntários pela participação, no entanto, caso haja qualquer despesa decorrente desta participação haverá o seu ressarcimento pelos pesquisadores.

6. O nome do participante será mantido em sigilo, assegurando assim a sua privacidade, e se desejarem terão livre acesso a todas as informações e esclarecimentos adicionais sobre o estudo e suas consequências, enfim, tudo o que queiram saber antes, durante e depois da sua participação.

7. Os dados coletados serão utilizados única e exclusivamente para fins desta pesquisa, e os resultados poderão ser publicados.

Qualquer dúvida, pedimos a gentileza de entrar em contato com Claudio Santos Cabral, pesquisador (a) responsável pela pesquisa, e-mail: claudio.cabral@aluno.ifsp.edu.br.

RAPHAEL DE SOUZA R. FEITOSA Eu ,declaro ter sido informado e concordo em ser participante do Projeto de pesquisa acima descrito.

Guarulhos, 07 de Junho de 2022.


Assinatura do participante

APÊNDICE B - Histórias dos usuários

Tabela 1 - Requisitos do usuário

| Dados do Projeto | | | |
|---|---|------------|--------|
| Nome do Projeto: | API - Conversão Retrospectiva de Fichas Catalográficas | | |
| Número de Requisitos do usuário | Descrição - Requisitos do Usuário (Desejo) | Prioridade | Sprint |
| RF001 | História do usuário: Como um bibliotecário, gostaria de escanear lotes de fichas físicas, para armazená-las no repositório remoto. | P1 | 1 |
| | Crítérios de aceite: Finalizar o escaneamento de todas as fichas. | | |
| RF004 | História do usuário: Como um bibliotecário, gostaria de um programa externo que digitalize fichas, para selecionar seus campos desejados. | P2 | 2 |
| | Crítérios de aceite: Testar e verificar a exibição do texto da ficha em formato digital. | | |
| RF005 | História do usuário: Como um bibliotecário, gostaria de um programa interno que extraia os registros bibliográficos selecionados nas fichas, para registrá-los na base de dados da biblioteca. | P2 | 2 |
| | Crítérios de aceite: Verificar e validar os campos capturados nas fichas, que são: Nº do título (NAT), Nº de exemplar (NAE), Classificação, Autor, Notação do Autor, Título, Subtítulo, Edição, Local, Editora, Ano, Detalhe, Série, LBN, ISBN, Notas. | | |
| RF003 | História do usuário: Como um bibliotecário, gostaria que o sistema de catalogação tenha controle de acesso login e logoff, para que usuários não cadastrados não tenham acesso. | P1 | 1 |
| | Crítérios de aceite: Efetuar o acesso no sistema e sair do mesmo. | | |
| RF002 | História do usuário: Como um bibliotecário, gostaria de efetuar o cadastro de primeiro acesso, para acessar o sistema de catalogação. | P1 | 1 |
| | Crítérios de aceite: Testar a validação do cadastro de novos usuários no sistema. | | |
| RF006 | História do usuário: Como um bibliotecário, gostaria de consultar os registros de uma ficha pelo autor, para que a catalogação seja feita por autor. | P1 | 3 |
| | Crítérios de aceite: Testar a busca de registros bibliográficos de uma ficha, por meio do ponto de acesso principal (palavra-chave) que será o nome do autor. | | |
| RNF001 | O programa deve processar cada ficha menos de 1s. | P1 | 1 |
| RNF002 | O programa deve garantir a elegibilidade, integridade e confiabilidade dos dados durante as consultas dos registros bibliográficos. | P2 | 3 |
| * 1 = Imprescindível; 2 = Importante; 3 = Pouco Importante; 4 = Desejável | | | |

Tabela 2 - Planejamento das sprints

| Nome do Projeto | | Data de Início | Data Final | Progreso Geral | | Entrega do Projeto | |
|---|--|----------------|------------|----------------|--|----------------------|--|
| API para conversão retrospectiva de fichas catalográficas | | 01/mar | 20/dez | 100% | | Declaração do Escopo | |

| AT RISK | Nome da tarefa | Responsável | Story Points | Início | Fim | Duração (dias) | Status |
|-------------------------------------|--|----------------|--------------|------------|------------|----------------|----------|
| <input checked="" type="checkbox"/> | Sprint 1 - Desenvolvimento Fase 1 | | | | | | |
| <input type="checkbox"/> | Feature 1 - RF001 | Claudio Cabral | | 2022-03-21 | 2022-03-31 | 10 | Completo |
| 1.1 | Desenvolver a parte View da tela home do sistema | | | 2022-03-21 | 2022-03-22 | 1 | Completo |
| 1.2 | Desenvolver a parte View da tela de login do sistema | | | 2022-03-21 | 2022-03-23 | 2 | Completo |
| 1.3 | Desenvolver a parte View da tela de cadastro do sistema | | | 2022-03-24 | 2022-03-25 | 1 | Completo |
| 1.4 | Desenvolver a parte View da tela de serviço de conversão e catalogação de fichas do sistema | | | 2022-03-25 | 2022-03-31 | 6 | Completo |
| | Feature 2 - RF002 | Claudio Cabral | | 2022-03-28 | 2022-04-08 | 11 | Completo |
| 2.1 | Desenvolver a Controller da tela de cadastro do usuário | | | 2022-03-28 | 2022-03-31 | 3 | Completo |
| 2.2 | Desenvolver a Controller da tela de login, para acesso do usuário no sistema | | | 2022-04-01 | 2022-04-08 | 7 | Completo |
| | Feature 3 - RF003 | Claudio Cabral | | 2022-04-18 | 2022-06-30 | 73 | Completo |
| 3.1 | Desenvolver e validar restrições em caso de dados incorretos no cadastramento do usuário | | | 2022-04-18 | 2022-04-30 | 12 | Completo |
| 3.2 | Desenvolver a classe model [Usuario] para persistir seus dados na base de dados | | | 2022-05-02 | 2022-05-31 | 29 | Completo |
| | Sprint 2 - Desenvolvimento Fase | | | | | | |
| | Feature 4 - RF004 | Claudio Cabral | | 2022-07-18 | 2022-08-18 | 31 | Completo |
| 4.1 | Utilizar API Amazon Textract para digitalizar as fichas que estão no repositório S3 | | | 2022-07-18 | 2022-07-25 | 7 | Completo |
| 4.2 | Criar um motor (batch process) para buscar no S3 as fichas para serem digitalizadas em filas | | | 2022-07-26 | 2022-08-08 | 13 | Completo |
| 4.3 | Criar o FichaDao para persistir os dados capturados na base de dados | | | 2022-08-09 | 2022-08-18 | 9 | Completo |
| | Feature 5 - RF005 | Claudio Cabral | | 2022-08-19 | 2022-09-18 | 30 | Completo |
| 5.1 | Desenvolver a classe Ficha e seus métodos | | | 2022-08-19 | 2022-08-26 | 7 | Completo |

| | | | | | | | |
|--------------------------|--|----------------|--|------------|------------|----|----------|
| 5.2 | Desenvolver e testar métodos da classe Ficha | | | 2022-09-02 | 2022-09-17 | 15 | Completo |
| <input type="checkbox"/> | Sprint 3 - Desenvolvimento Fase | | | | | | |
| <input type="checkbox"/> | Feature 6 - RF006 | Claudio Cabral | | 2022-09-19 | 2022-11-20 | 66 | Completo |
| 6.1 | Desenvolver a API Rest na tela de serviço, para buscar os registros das fichas que estão no banco de dados | | | 2022-09-19 | 2022-09-26 | 7 | Completo |
| 6.2 | Integração da API Rest com o Front-end | | | 2022-09-27 | 2022-10-15 | 18 | Completo |
| 6.3 | Realizar testes de integração | | | 2022-10-17 | 2022-11-24 | 38 | Completo |

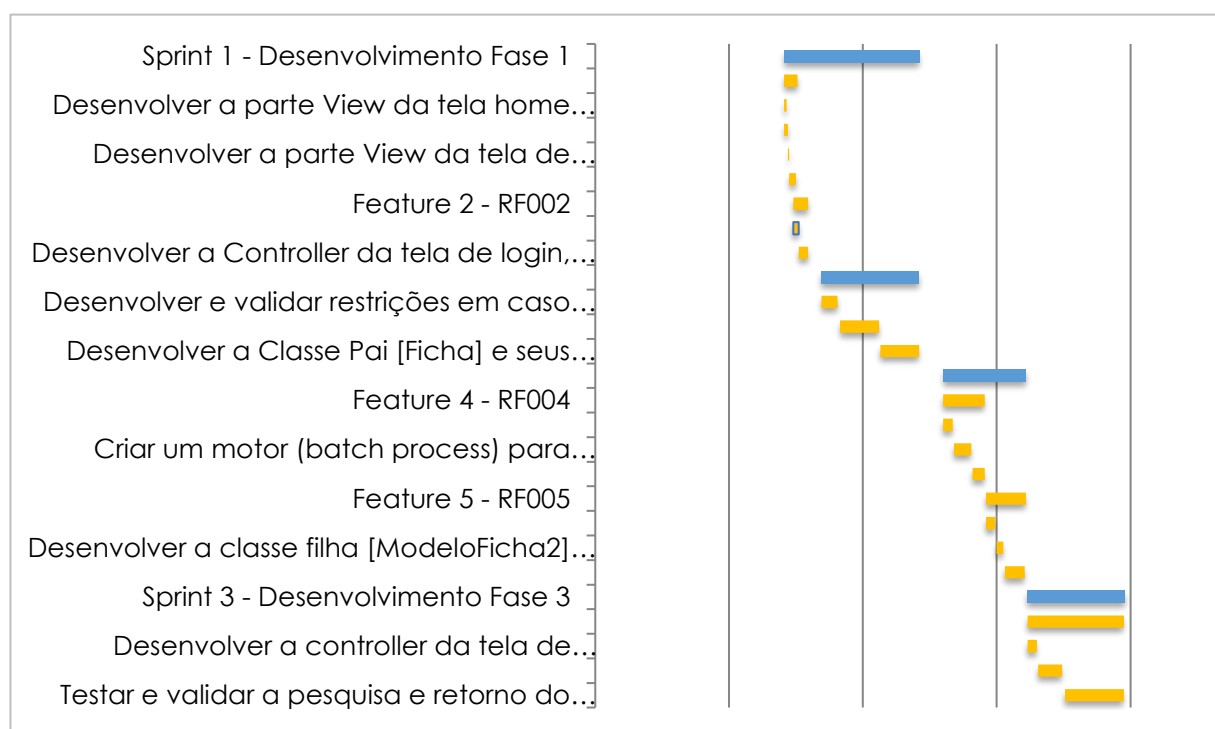


Tabela 3 - Gestão de riscos

NOME DO PROJETO:

API para Conversão Retrospectiva de Fichas Catalográficas

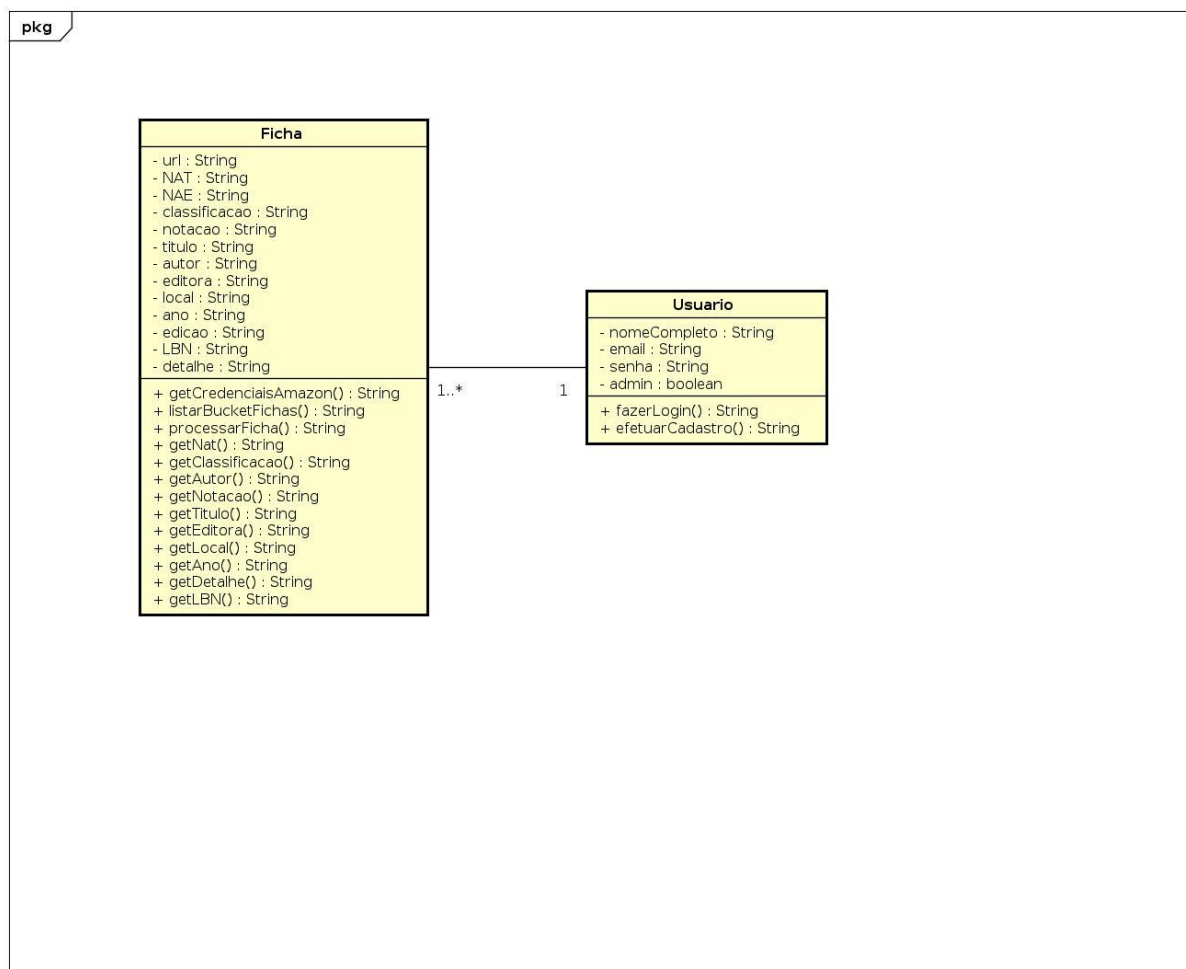
| DEFINA COMO SERÃO PONTUADOS OS DADOS DA MATRIZ | | | | | |
|--|-------------------|-------------|----------------|------------|------------------|
| PROBABILIDADE | DESPREZÍVEL (10%) | BAIXO (30%) | MODERADO (50%) | ALTO (70%) | MUITO ALTO (90%) |
| | 0,1 | 0,3 | 0,5 | 0,7 | 0,9 |
| IMPACTO | DESPREZÍVEL (10%) | BAIXO (30%) | MODERADO (50%) | ALTO (70%) | MUITO ALTO (90%) |
| | 0,05 | 0,1 | 0,2 | 0,4 | 0,8 |

| Riscos / Perigos | | | | | |
|--|----------------------|---------------|------------|--|--|
| DESCRIÇÃO DO RISCO | TIPO | PROBABILIDADE | IMPACTO | PREVENÇÃO | PLANO DE CONTINGÊNCIA |
| Prazos de entrega não cumpridos, falta de dimensionamento do escopo do projeto | Risco Organizacional | Moderado | Muito Alto | Desenvolvimento do Plano de Cronograma das Sprints e Diagramas UML | Replanejamento do cronograma das Sprints, priorizando as funcionalidades críticas do sistema |
| Falta de aprovação das partes envolvidas no projeto | Risco Organizacional | Moderado | Muito Alto | Comunicação efetiva e envio de e-mails, mensagens no WhatsApp | Enviar um e-mail informativo solicitando o consentimento das partes envolvidas sobre alguma <i>change</i> ou toma de decisão durante o projeto |
| Ausência de monitoramento e comunicação com o orientador do projeto | Risco Organizacional | Moderado | Alto | Reuniões semanais de 15 minutos | Remarcar o agendamento das reuniões em casos de imprevistos, alinhar com o orientador horário e sua disponibilidade na semana |
| Gestão de levantamento de requisitos funcionais e não funcionais mal definidos | Risco de Negócio | Alto | Muito Alto | Realizar entrevistas para aprovação das histórias de usuário com o cliente e meetings com orientador para definir a prioridade das histórias | Reconstruir os requisitos que não satisfazem o desejo do cliente |
| Inexperiência do programador em relação a linguagem a ser desenvolvida | Risco Técnico | Alto | Muito Alto | Uso de plataformas que oferecem cursos sobre programação como Alura e Udemy | Solicitar auxílio às pessoas que trabalham com a linguagem de programação utilizada neste projeto, buscar alternativas de conteúdo nos fóruns de programação e documentação da própria linguagem |

| | | | | | |
|--|---------------|------------|------------|---|---|
| Programação irrealista | Risco Técnico | Alto | Muito Alto | Isolar e classificar os conceitos da linguagem que serão essenciais para desenvolvimento do projeto | Redimensionamento do escopo (arquitetura) da estrutura do sistema e realização do procedimento de refatoração se necessário |
| Ocorrência de bugs durante os testes unitários | Risco Técnico | Muito Alto | Muito Alto | Aderir as boas práticas de desenvolvimento de software | Utilizar a ferramenta de Depuração instalada na IDE do PyCharm, para identificação dos erros de lógica e sintaxe. |
| Dependência de serviços e produtos externos gratuitos em um determinado intervalo de tempo (<i>Trial Free</i>) | Risco Externo | Moderado | Alto | Averiguar se os produtos e serviços utilizados possuem versão gratuita por longo prazo | Pesquisar mais de uma ferramenta, e analisar seus custos. |
| Corrupção do S.O, defeito de Hardware -perda do código-fonte | Risco Externo | Baixo | Muito Alto | (Backup) - Criação de um repositório remoto GitHub e versionamento de código | Restauração do Sistema Operacional e sistemas de arquivos, efetuar inspeções corretivas no computador |

Diagrama de Classe de Domínio

Figura 10 - Diagrama de classe de domínio



Diagramas de Sequência

Figura 11- Login de usuário

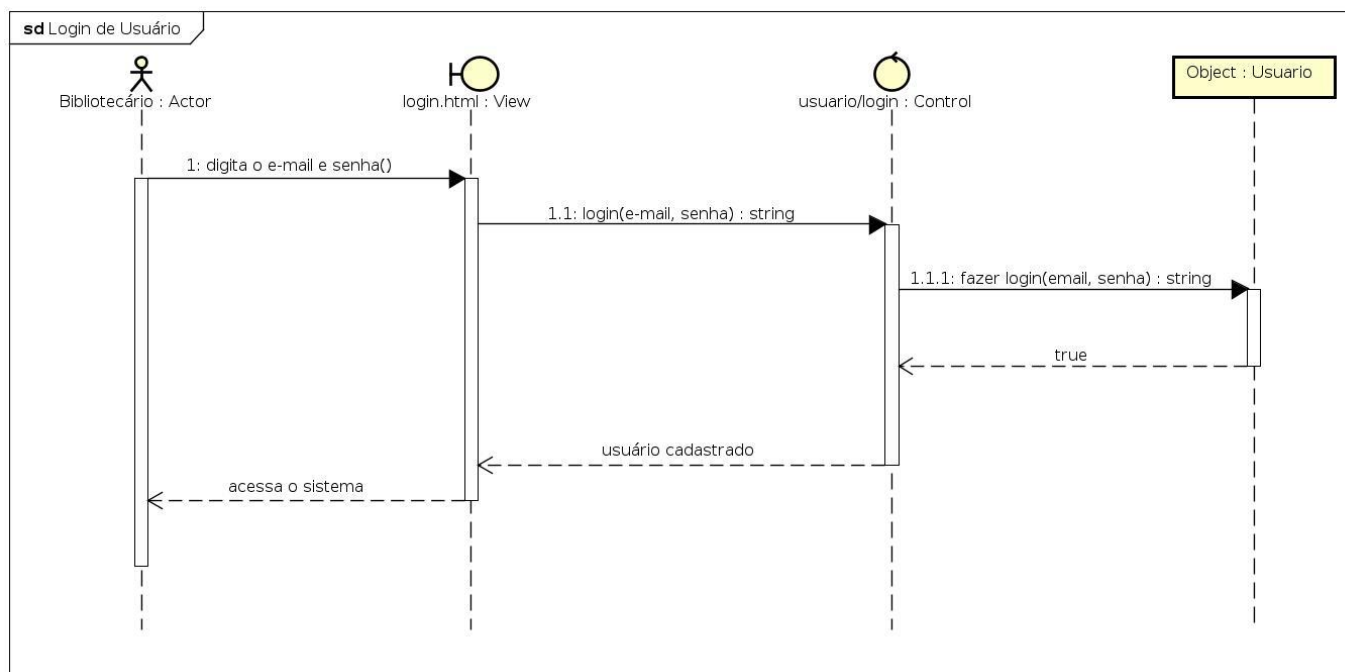


Figura 12 - Cadastro de usuário

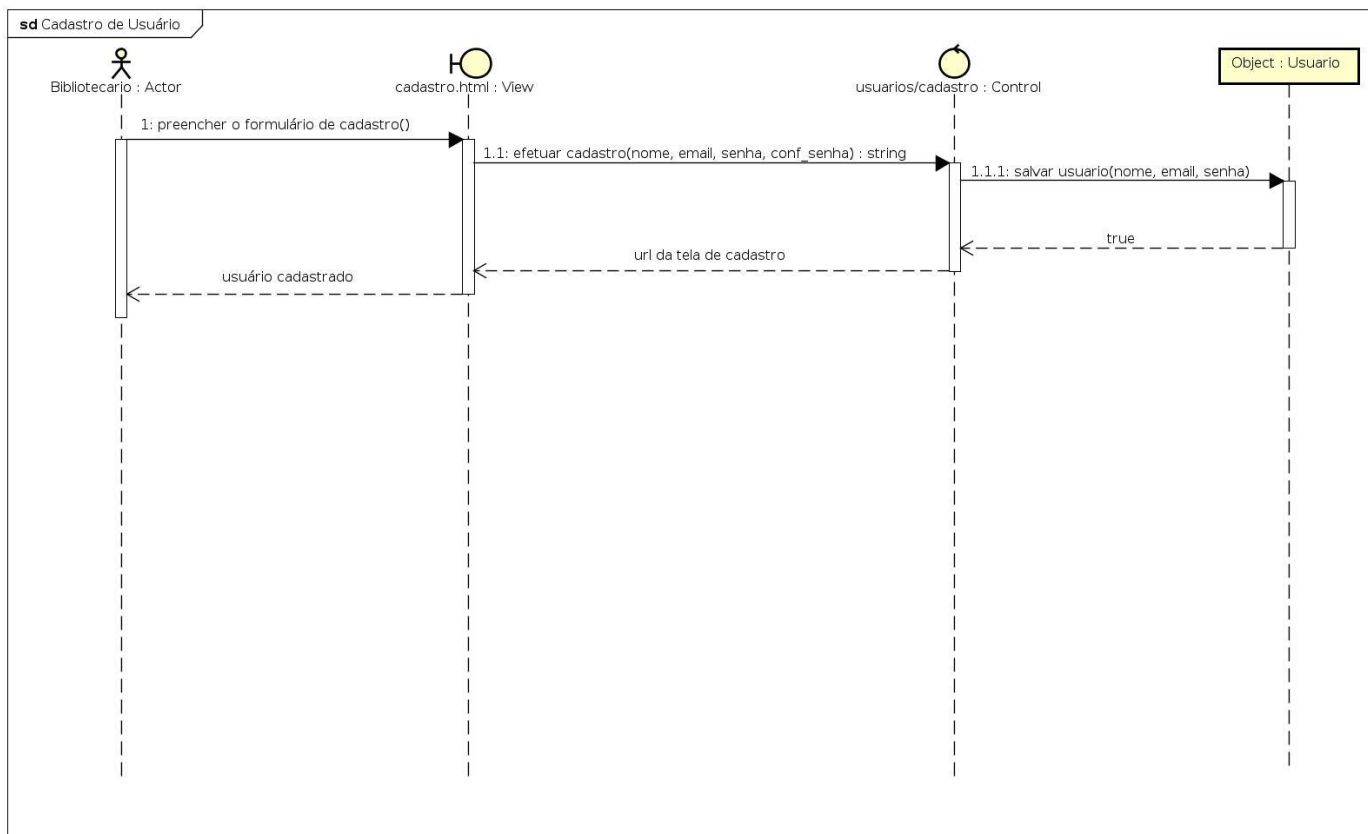


Figura 13 - Consultar ficha na API

sd Consultar Ficha na API

