

NE-Soft Factory

Versão 2.0

Documento de Arquitetura

Responsável: *Alfeu Buriti*

NE-Soft Factory



Histórico de Alterações

Versão	Data	Responsável	Descrição
1.0	10/08/2018	Alfeu Buriti Júnior Antônio Gomes Amorim Antonio Moreira da S. Filho Erick Guilherme Cavalcanti Fernando H. E. C. Guimarães Helton Santa Cruz Maria do Carmo Costa Vinícius B. Ferreira	Criação
1.1	12/08/2018	Antonio Moreira da S. Filho Helton Santa Cruz	Revisão do documento
2.0	29/08/2018	Antônio Gomes Amorim	Revisão do documento

SUMÁRIO

1. INTRODUÇÃO	4
2. VISÃO GERAL	4
3. ESCOPO	5
4. ESCOPO NEGATIVO	6
5. DESCRIÇÃO DO PROBLEMA	7
6. RESPONSÁVEIS E INTERESSADOS	7
7. REQUISITOS	8
8. DESCRIÇÃO DOS COMPONENTES E INTERAÇÕES	8

]

1. INTRODUÇÃO

Este Documento de Arquitetura de Software (DAS) tem o objetivo de oferecer uma visão geral do sistema SISCAP, para descrever diferentes aspectos e componentes da arquitetura do mesmo.

Pretende-se demonstrar de forma simples, os elementos que compõem o projeto de maneira que sejam visualizadas as ligações e interações entre as mesmas.

É composto por seções que compreendem introdução, visão geral, escopo, escopo negativo, descrição do problemas, responsáveis e interessados, requisitos e descrição dos componentes e interações.

2. VISÃO GERAL

A arquitetura do sistema irá ser dividida nos seguintes componentes:

Módulos Automatizados

Coletor de diário oficial, Conversor em OCR do diário oficial, Indexador de publicações, além do Buscador de palavras por meio de uso de linguagem natural. Estes módulos serão executados por meio de tarefas agendadas por um Scheduler EJB.

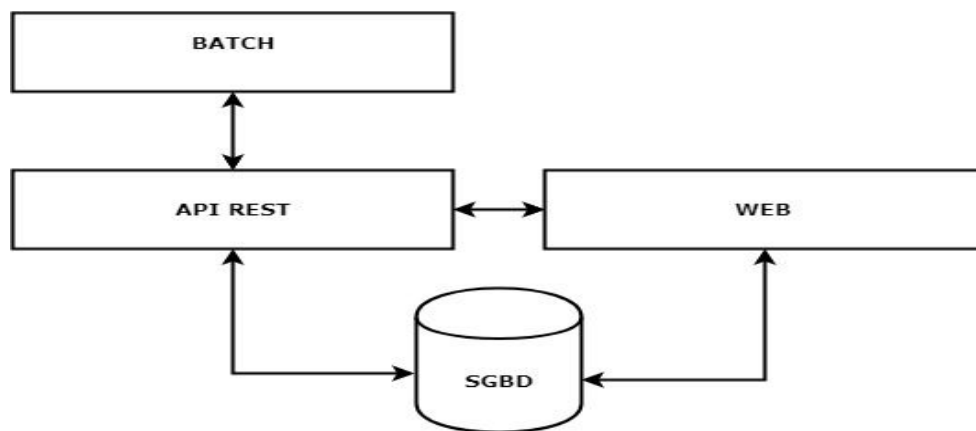
Módulo WEB

Interface de configuração para o usuário que se comunicará com a API REST para utilização do Banco de Dados.

Módulo Backend (API REST)

Responsável pela troca de dados entre os demais módulos do sistema e o SGBD, utilizando JPA (*Java Persistence API*).

3. ESCOPO



O projeto terá como escopo, a criação de um software que irá coletar, tratar, e indexar as publicações oficiais dos jurisdicionados do TCE-PI, armazenando os textos para realização de consultas otimizadas através de uma interface WEB. A solução irá prover os seguintes entregáveis:

- Coletar diários oficiais das seguintes fontes:
 - Diário Oficial do Estado do Piauí;
 - Diário Oficial de Teresina;
 - Diário Oficial de Parnaíba;
 - Diário Oficial dos Municípios do Piauí.

- Cadastrar fontes de diários oficiais
- Realizar upload de diários oficiais
- Converter os diários oficiais de imagens para informações textuais;
- Indexar textos extraídos dos diários oficiais;
- Fazer buscas nos diários oficiais com uso de linguagem natural;
- Consultar informações das publicações por parte dos auditores.

Além disso, os dados gerados serão disponibilizados no banco de dados para utilização das informações pelos usuários do TCE-PI.

4. ESCOPO NEGATIVO

- Não serão contempladas fontes distintas das definidas no escopo positivo;
- Treinamento dos usuários do TCE-PI não será realizado pela fábrica;
- Manutenção do sistema após o fim do projeto não será realizada pela fábrica;
- Não haverá suporte pela fábrica após a homologação final do cliente.

5. DESCRIÇÃO DO PROBLEMA

O principal objetivo deste projeto visa auxiliar os auditores com dificuldades na obtenção dos dados de licitações e nomeações a partir dos PDFs presentes nos websites dos diários oficiais do Estado do PI. O método utilizado atualmente para extração desses dados é executado de forma manual, o que gera um enorme esforço.

Uma solução proposta pela nossa equipe, definida a partir da regra de negócio estabelecida pelo tribunal, é a indexação dos textos obtidos anteriormente e a partir daí fazer buscas nos diários oficiais através do uso de linguagem natural.

Essas soluções apresentadas proporcionam uma maior agilidade e qualidade na auditoria de licitações e nomeações, o que minimizaria o desperdício dos recursos públicos. Além disso, o sistema se propõe a auxiliar a diminuição de fraudes em licitações o que geraria como efeito macro o auxílio no combate à corrupção.

6. RESPONSÁVEIS E INTERESSADOS

Equipe

- Alfeu Buriti;
- Antônio Filho;
- Helton Santa Cruz;
- Erick Guilherme;
- Maria do Carmo;
- Fernando Henrique Guimarães;
- Antônio Gomes Amorim.

Stakeholders externos

- Ricardo Almeida (representante do cliente);
- José Inaldo (product owner).

7. REQUISITOS

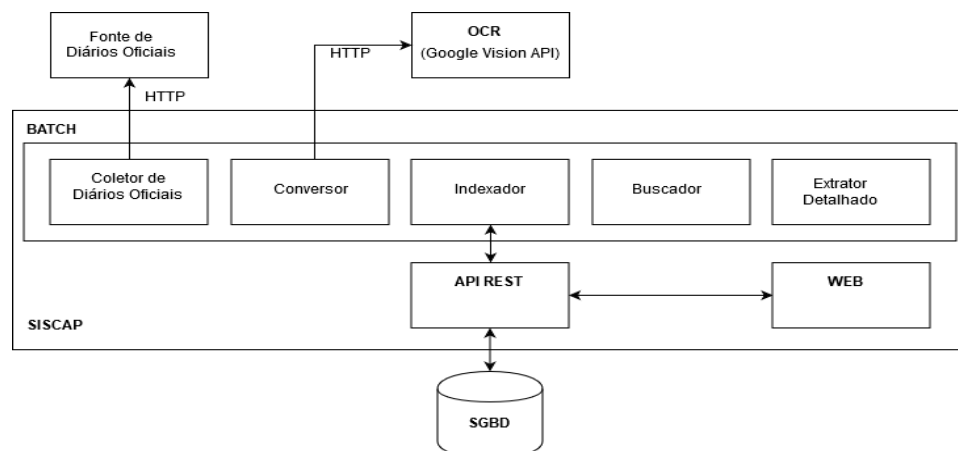
Requisitos Funcionais

- Cadastrar diários oficiais;
- Realizar *upload* manual de diários oficiais;
- Coletar os diários oficiais;
- Converter os diários oficiais de imagens para informações textuais;
- Indexar textos extraídos dos diários oficiais;
- Fazer buscas nos diários oficiais com uso de linguagem natural.

Requisitos não-Funcionais e de Qualidade

- Preferencialmente utilizar banco de dados MS-SQL Server;
- Preferencialmente utilizar tecnologias Java;
- O percentual obtido na conversão OCR deverá ser de pelo menos 90% de acerto.

8. DESCRIÇÃO DOS COMPONENTES E INTERAÇÕES



A aplicação será composta da seguinte forma:

Módulos Automatizados

Módulos que serão executados automaticamente por meio de tarefas agendadas e comunicar-se-ão com a API REST (Backend) para manipulação dos dados. São eles:

Coletor de diários oficiais

Responsável pela coleta de diários oficiais nas fontes cadastradas. A comunicação com as fontes de diários oficiais será através do protocolo HTTP.

Conversor

A maioria dos diários oficiais coletados apresentam-se em arquivos no formato PDF não pesquisável. Este módulo irá converter o conteúdo desses arquivos para texto com o uso da tecnologia *Optical Character Recognition* (OCR). A comunicação com a solução será através do protocolo HTTP.

Indexador

Os diários oficiais serão indexados por esse módulo, após a aplicação das técnicas realizadas pelo módulo Conversor. A tecnologia a ser utilizada será detalhada posteriormente.

Buscador

Fará busca de informações a partir das palavras digitadas pelos usuários. A tecnologia a ser utilizada será detalhada posteriormente.



Extrator Detalhado de Conteúdo

Responsável por extrair informações relevantes das publicações. A tecnologia utilizada será a API Vision do Google, ferramenta que detecta e extrai texto a partir de imagens.

Módulo WEB

Responsável pela interface de configuração para o usuário e também para realização upload manual de diários oficiais. A comunicação com o módulo API REST será através do protocolo HTTP.

Módulo API REST (Backend)

Responsável pela troca de informações com o SGBD utilizando JPA.