



**NE Soft**  
Factory

# ***NE-Soft Factory***

Versão 1.0

## **Documento de Arquitetura**

Responsável: *Alfeu Buriti*

NE-Soft Factory

### Histórico de Alterações

<b>Versão</b>	<b>Data</b>	<b>Responsável</b>	<b>Descrição</b>
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



## **SUMÁRIO**

1. Introdução
2. Visão Geral
3. Escopo
4. Escopo Negativo
5. Descrição do Problema
6. Responsáveis
7. Requisitos
  - 7.1. Requisitos funcionais
  - 7.2. Requisitos não-funcionais e de qualidade
8. Descrição dos componentes e interações

## 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, Extrator (Delimitador de Publicações), Classificador de tipos de publicações, além do Extrator detalhado de conteúdo. Estes módulos serão executados por meio de tarefas agendadas por um Scheduler EJB.

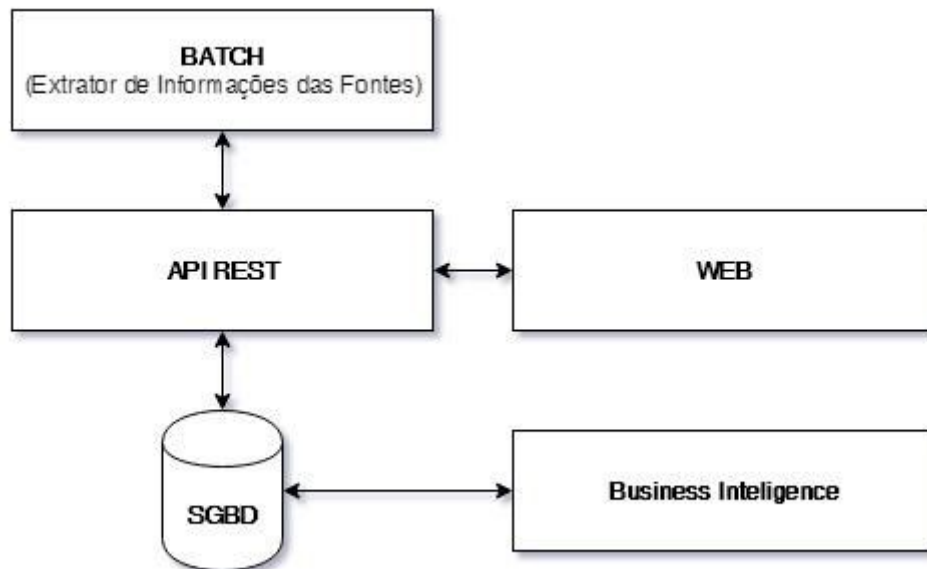
### **Módulo WEB**

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 classificar as publicações dos jurisdicionados do TCE-PI a fim de analisar e identificar indícios de fraudes em licitações. A solução irá gerar os seguintes entregáveis:

- Coletar Diários Oficiais das seguintes fontes:
  - o Diário Oficial do Estado do Piauí;
  - o Diário Oficial de Teresina;
  - o Diário Oficial de Parnaíba;
  - o Diário Oficial dos Municípios do Piauí.
- Converter Diários Oficiais para texto com uso da tecnologia *Optical Character Recognition* (OCR);
- Extrair de Publicações dos Diários Oficiais (Delimitador);
- Classificar tipos de Publicações (Licitação e Nomeações);
- Extrair conteúdo Detalhado;

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;
- Classificação de outros tipos de publicações diferentes das definidas no escopo não serão realizadas;
- 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 segunda solução proposta pela nossa equipe, definida a partir da regra de negócio estabelecida pelo tribunal, é a análise dos dados obtidos anteriormente. Esta funcionalidade deverá ser capaz de exibir a probabilidade de haver um possível fraude em licitações e nomeações.

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 Amorim.

### **Stakeholders externos**

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

## **7. REQUISITOS**

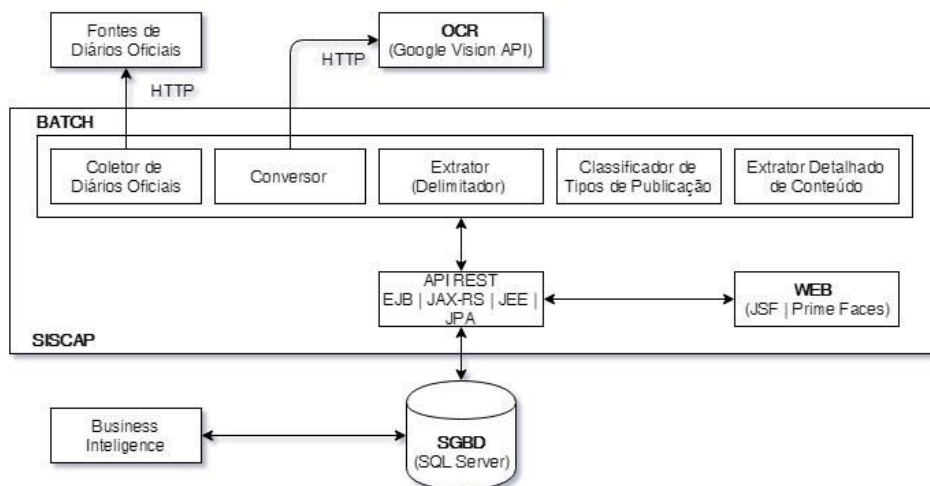
### **Requisitos Funcionais**

- Coletar os diários oficiais;
- Converter os diários oficiais de imagens para informações textuais;
- Extrair as publicações contidas nos diários oficiais;
- Classificar as publicações extraídas dos diários oficiais;
- Extrair informações relevantes para cada publicação de acordo com a sua classificação;
- Disponibilizar as análises a serem definidas pelo cliente;
- Disponibilizar informações das publicações para os auditores.

### **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. Como exemplo de fontes podemos citar Diário Oficial do Estado do Piauí e Diário Oficial dos Municípios do Estado do Piauí. A comunicação com as fontes de diários oficiais será através do protocolo HTTP.

#### Conversor

Como os diários oficiais coletados, em sua maioria, apresentam-se em arquivos no formato pdf não pesquisável, esse módulo será responsável por converter o conteúdo desses arquivos para texto com o uso da tecnologia OCR (Optical Character Recognition). A comunicação com a solução que será utilizada para realizar



**Extrator(Delimitador)**

Os diários oficiais são compostos por diversas publicações que serão delimitadas e extraídas por esse módulo, após a aplicação das técnicas realizadas pelo módulo Conversor.

**Classificador de Tipos de Publicação**

Responsável por classificar as publicações delimitadas no módulo Extrator em tipos pré-definidos como licitações e nomeações, por exemplo.

**Extrator Detalhado de Conteúdo**

Responsável por extrair informações relevantes para cada publicação de acordo com sua classificação.

**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.