**ESTUDO DE CASO – DESENVOLVIMENTO**

**DE UM SISTEMA DE GED COM OCR EM PYTHON**

**PARA AUXILIAR EM GERENCIAMENTO**

**DE DOCUMENTOS ELETRÔNICOS**

**EM ÁREAS RURAIS.**

## 1. Introdução

## 1.1 Desafio

Empresas e clientes em áreas rurais enfrentam desafios significativos no gerenciamento de documentos eletrônicos, devido à falta de acesso a serviços bancários e tecnologia adequada. O processo de coleta, armazenamento e recuperação de documentos muitas vezes é demorado e propenso a erros, causando inconveniência e perda de tempo para as partes envolvidas.

## 1.2 Justificativa do Estudo

Este estudo busca desenvolver um sistema de Gerenciamento Eletrônico de Documentos (GED) com Reconhecimento Ótico de Caracteres (OCR) em Python para mitigar as dores enfrentadas pelas empresas e clientes em áreas rurais. A implementação deste sistema proporcionará uma solução rápida e precisa para o gerenciamento de documentos eletrônicos, reduzindo o tempo gasto e os erros cometidos durante o processo.

## 1.3 Objetivos do Projeto

O objetivo principal deste projeto é desenvolver um sistema que permita às empresas e clientes em áreas rurais coletar, armazenar e recuperar documentos eletrônicos de forma eficiente e precisa. Além disso, o projeto visa proporcionar uma experiência mais conveniente e acessível para as partes envolvidas no processo de gerenciamento de documentos.

## 2. Contextualização do Problema

## 2.1 Desafios no Gerenciamento de Documentos em Áreas Rurais

Empresas e clientes em áreas rurais enfrentam dificuldades significativas no processo de gerenciamento de documentos eletrônicos. A falta de acesso a serviços bancários e tecnologia adequada torna o processo demorado e propenso a erros, causando frustração e perda de tempo para todos os envolvidos.

## 2.2 Dores dos Clientes do Rural

Os clientes do rural muitas vezes precisam lidar com documentos importantes, como contratos e registros bancários, que requerem coleta e análise cuidadosas. O processo de escrever à mão e depois passar para o computador é demorado e propenso a erros, resultando em perda de tempo e possíveis retrabalhos.

## 3. Proposta de Solução

## 3.1 Descrição do Sistema Proposto

O sistema proposto utilizará tecnologia de Reconhecimento Ótico de Caracteres (OCR) em Python para permitir a rápida digitalização e análise de documentos eletrônicos. Os documentos serão coletados e processados de forma rápida e precisa, reduzindo o tempo gasto e os erros cometidos durante o processo.

## 3.2 Requisitos do Sistema

## 3.2.1 Requisitos Funcionais:

1. RF1: Coleta de Documentos: O sistema deve permitir que os usuários coletem documentos eletrônicos por meio de upload de arquivos ou digitalização.
2. RF2: Reconhecimento Ótico de Caracteres (OCR): O sistema deve ser capaz de extrair texto de documentos digitalizados utilizando tecnologia OCR em Python.
3. RF3: Armazenamento de Documentos: Os documentos coletados e os metadados associados devem ser armazenados de forma segura nos bancos de dados.
4. RF4: Recuperação de Documentos: Os usuários devem poder buscar e recuperar documentos com base em critérios específicos, como nome do documento, data ou palavras-chave.
5. RF5: Visualização de Documentos: Os usuários devem poder visualizar os documentos armazenados no sistema.
6. RF6: Preenchimento Automático de Formulários: O sistema deve ser capaz de ler documentos digitalizados por OCR e preencher automaticamente formulários com as informações extraídas, agilizando assim o processo de obtenção de documentos importantes.

## 3.2.2 Requisitos Não Funcionais:

1. RNF1: Tecnologias Utilizadas: O sistema será desenvolvido em Python, utilizando bibliotecas de OCR para reconhecimento ótico de caracteres e com sistemas de gerenciamento de banco de dados.
2. RNF2: Interface Intuitiva: A interface do usuário deve ser amigável e de fácil utilização, proporcionando uma experiência intuitiva para os usuários.
3. RNF3: Segurança: O sistema deve implementar autenticação de usuários e controle de acesso para garantir a segurança dos documentos armazenados.
4. RNF4: Performance: O sistema deve ter uma boa performance, mesmo quando lidando com grandes volumes de documentos, garantindo tempos de resposta rápidos para as operações de busca e recuperação.

## 4. Metodologia

**4.1 Levantamento de Requisitos**

O levantamento de requisitos foi realizado por meio de diversas abordagens para identificar as necessidades específicas do sistema. Estas abordagens incluíram:

1. Questionários: Foram elaborados questionários estruturados para coletar informações quantitativas e qualitativas sobre os processos atuais de gerenciamento de documentos eletrônicos, as principais dificuldades enfrentadas e as expectativas em relação ao sistema.
2. Observação Direta: A equipe realizou observações diretas do processo de gerenciamento de documentos em empresas e clientes em áreas rurais para identificar pontos de fricção, gargalos e oportunidades de melhoria.
3. Análise Documental: Foi realizada uma análise detalhada de documentos existentes, como formulários, contratos, registros e outros tipos de documentos utilizados pelas empresas e clientes em áreas rurais, para compreender as informações necessárias e os padrões de uso.

Essas abordagens de pesquisa proporcionaram uma compreensão abrangente das necessidades e requisitos do sistema, garantindo que o sistema proposto seja capaz de atender de forma eficaz às demandas dos usuários e às especificidades do ambiente rural em que será implantado. Com base nesse levantamento, será possível realizar os próximos passos do projeto com maior precisão e eficiência.

**4.2 Design do Sistema**

Com base nos requisitos levantados, será elaborado o design do sistema, incluindo arquitetura de software e interface de usuário.

**4.3 Entidades**

Com base nos requisitos levantados, as seguintes entidades serão definidas para o sistema:

1. **Usuário**: Representa os usuários do sistema, como funcionários da empresa responsáveis pelo gerenciamento de documentos ou clientes que interagem com o sistema para acessar documentos. Cada usuário terá informações como nome, endereço de e-mail, senha e papel no sistema (por exemplo, administrador, funcionário, cliente).
2. **Documento**: Refere-se aos documentos eletrônicos a serem gerenciados pelo sistema, como formulários, contratos, registros, entre outros. Cada documento terá informações como título, tipo de documento, conteúdo e metadados associados.
3. **Metadados do Documento**: Informações adicionais associadas a cada documento, como nome do documento, data de criação, autor, palavras-chave, entre outros. Os metadados são importantes para facilitar a busca e a recuperação de documentos no sistema.
4. **Cadastro de Clientes**: Representa os dados dos clientes que precisam ser preenchidos pelos usuários, como parte do processo de coleta de informações. Cada formulário terá um conjunto de campos a serem preenchidos, e o sistema será capaz de preencher automaticamente esses formulários com base nas informações extraídas dos documentos digitalizados.
5. **Registro de Atividades**: Registra todas as atividades realizadas no sistema, como upload de documentos, preenchimento de formulários, buscas realizadas, entre outros, para fins de rastreabilidade e auditoria. Cada registro de atividade incluirá as seguintes informações:

* Data e hora da atividade
* Usuário responsável

1. **Configurações do Sistema**: Entidade responsável por armazenar as configurações do sistema, como detalhes de conexão com bancos de dados, configurações de segurança e preferências do usuário. Isso permite uma fácil personalização e manutenção do sistema.

Essas entidades serão essenciais para o desenvolvimento do sistema de GED com OCR em Python, garantindo que todas as informações necessárias sejam devidamente estruturadas e gerenciadas para atender às demandas dos usuários e do ambiente rural em que será implantado.

**4.4 Implementação**

O sistema será desenvolvido em Python, utilizando as tecnologias especificadas. Testes unitários e de integração serão realizados durante o desenvolvimento.

**4.5 Testes e Validação**

Serão realizados testes de sistema para garantir a funcionalidade e a qualidade do sistema. Os usuários finais serão convidados a testar o sistema e fornecer feedback.

**4.6 Implantação**

Após a conclusão dos testes e validação, o sistema será implantado em ambiente de produção e disponibilizado para uso.

**4.7 Cronograma**

A seguir, apresenta-se o cronograma detalhado para o desenvolvimento do sistema:

| Etapa | Data |
| --- | --- |
| Levantamento de Requisitos | 06/04/2024 |
| Design do Sistema | 06/04/2024 |
| Entidades | 06/04/2024 |
| Implementação | 17/04/2024 |
| Testes e Validação | 18/04/2024 |
| Implantação | 19/04/2024 |

**4.8 Ferramentas**

* Analise de imagens
  + OpenCV, OCR, Tesseract
* Interface
  + PyQt
* Banco de dados e Autenticação de usuário
  + Psycopg2 - Para acessar o PostgreSQL
    - [Psycopg – PostgreSQL database adapter for Python —Psycopg 2.9.9 documentation](https://www.psycopg.org/docs/)
  + SQLAlchemy - para relacionar com as classes do sistema
    - [SQLAlchemy Documentation — SQLAlchemy 2.0 Documentation](https://docs.sqlalchemy.org/en/20/)
  + Passlib - Para criptografar senhas da autenticação
    - [Documentação do Passlib 1.7.4 — Documentação do Passlib v1.7.4](https://passlib.readthedocs.io/en/stable/)
* Documentação e Gerenciamento de tasks
  + Notion
* Controle de Versão
  + Git e Github

## 5. Resultados Esperados

O sistema desenvolvido proporcionará uma solução eficiente para o gerenciamento de documentos eletrônicos em áreas rurais, reduzindo o tempo gasto e os erros cometidos durante o processo.

## 6. Discussão e Análise

O desenvolvimento do sistema pode enfrentar desafios técnicos relacionados à integração com bancos de dados e implementação de OCR em Python. A precisão do OCR pode ser afetada por diferentes tipos de documentos e qualidade da digitalização. Para resolver esses desafios, será fundamental realizar testes extensivos durante o processo de desenvolvimento, além de implementar estratégias de otimização para melhorar a precisão do OCR. O uso de algoritmos avançados de reconhecimento de caracteres e a calibração adequada dos parâmetros do sistema também podem contribuir significativamente para a melhoria da precisão e eficácia do sistema. Além disso, a equipe de desenvolvimento estará preparada para lidar proativamente com quaisquer problemas técnicos que possam surgir, buscando soluções criativas e eficientes para garantir o sucesso do projeto.

## 7. Conclusão

O sistema proposto atenderá às necessidades das empresas e clientes em áreas rurais, proporcionando uma solução eficiente e acessível para o gerenciamento de documentos eletrônicos.

Este estudo de caso destaca a dor enfrentada pelas empresas e clientes em áreas rurais relacionada ao processo demorado e propenso a erros de escrever à mão e depois passar para o computador. A solução proposta visa proporcionar uma maneira rápida e precisa de coletar, analisar e armazenar documentos eletrônicos, reduzindo assim o tempo gasto e os erros cometidos durante o processo. Com essa solução, as partes envolvidas poderão economizar tempo e, se necessário, corrigir erros no mesmo dia, sem a necessidade de retrabalhos ou viagens adicionais.