

Programa para consulta em arquivos sobre a produção científica do DCC



Aluno: Pedro Vinícius da Silva Ribeiro

Matrícula: 2019033903

1. Visão Geral

Este projeto é uma solução automatizada para a consulta e extração de dados de arquivos XML relacionados à produção científica do Departamento de Ciência da Computação (DCC). O programa foi projetado para processar arquivos de currículo Lattes em formato XML e gerar tanto relatórios detalhados quanto resumos numéricos sobre as produções acadêmicas, como artigos publicados e trabalhos apresentados em eventos.

O objetivo principal é fornecer uma visão clara e organizada da produção científica do DCC, bem como facilitar o acompanhamento da produção individual dos pesquisadores.

2. Principais Funcionalidades

- Leitura e extração de dados de currículos Lattes em arquivos XML.
- Geração de relatórios detalhados com os dados do pesquisador.
- Geração de resumos numéricos das produções acadêmicas.
- Atualização de planilhas com os dados numéricos das produções acadêmicas, evitando duplicidade de pesquisadores.
- Processamento de múltiplos arquivos XML de uma pasta automaticamente.

3. Arquitetura do Sistema

O sistema foi desenvolvido seguindo o Modelo Incremental. Isso permitiu que o desenvolvimento fosse feito em pequenas etapas, com melhorias e ajustes contínuos. O sistema foi projetado em Python, utilizando as bibliotecas adequadas para processamento de arquivos XML e geração de planilhas.

Fonte: SOMMERVILLE, I. Engenharia de Software. 9. ed. São Paulo: Pearson Prentice Hall, 2011. 529 p.

4. Principais Componentes

- **Processador XML:** Extrai informações detalhadas dos currículos Lattes em formato XML, incluindo dados gerais, artigos publicados, trabalhos em eventos e outras produções acadêmicas.
- **Salvador de Dados:** Responsável por salvar os dados extraídos nos formatos JSON ou Excel. Também é responsável por atualizar a planilha geral com o resumo numérico das produções, verificando se o pesquisador já está na planilha e, se necessário, atualizando os dados.

- **Integração com Pastas:** Processa automaticamente todos os arquivos XML localizados em uma pasta, gerando as saídas conforme as opções selecionadas.

5. Codificação

Durante a fase de Codificação, aplicamos a filosofia de código limpo, que foca na legibilidade, manutenção e eficiência. Isso inclui:

- Uso de nomes significativos para variáveis, funções e classes.
- Divisão do código em pequenas funções coesas, cada uma com uma tarefa clara.
- Minimização da complexidade, evitando estruturas condicionais e loops aninhados.
- Uso de comentários claros para explicar partes importantes do código.
- Aplicação de boas práticas de programação e organização modular.
- Revisões regulares para garantir a qualidade e aderência aos princípios de código limpo.

6. Testes e Integração

Após a codificação, o programa foi testado para verificar sua funcionalidade e corrigir bugs. Todos os recursos foram validados, incluindo a consulta de arquivos XML, a extração de dados e a atualização das planilhas de produção. O software está funcionando de acordo com o esperado.

7. Operação e Manutenção

Após os testes, o software foi implantado e está em uso. Manutenções foram realizadas para atender a novas necessidades e atualizações tecnológicas, com ajustes no código para garantir seu funcionamento contínuo.

8. Requisitos do Projeto

8.1. Funcionais

- Leitura e extração de dados de currículos Lattes em formato XML.
- Escrita e atualização de planilhas com os dados extraídos.
- Geração de relatórios detalhados sobre as produções acadêmicas.
- Atualização de planilhas com os dados numéricos das produções, organizadas por pesquisador, evitando duplicidade.

8.2. Não Funcionais

- **Desempenho:** O sistema deve ser capaz de processar grandes volumes de dados de maneira eficiente.
- **Manutenção:** O código deve ser de fácil compreensão e manutenção, com funções bem definidas e organizadas.
- **Escalabilidade:** O sistema deve ser capaz de processar múltiplos arquivos XML sem comprometer o desempenho.

9. Tecnologias Utilizadas

Linguagem:

- **Python 3.x**

Bibliotecas:

- **lxml**: Para leitura e processamento de arquivos XML.
- **pandas**: Para manipulação de dados e geração de planilhas Excel.
- **openpyxl**: Para manipulação de arquivos Excel, incluindo a criação de múltiplas abas.
- **os**: Para manipulação de arquivos e diretórios.
- **json**: Para manipulação de dados em formato JSON.

Interface Gráfica:

- **Tkinter**: Biblioteca padrão do Python para desenvolvimento de interfaces gráficas (GUI). Usada para criar a interface gráfica amigável, com opções de seleção de arquivos, pastas, escolha de formato de saída e tipo de geração.
- **ttk (Themed Tkinter)**: Conjunto de widgets modernos fornecidos pelo Tkinter, usados para dar um visual mais atualizado à interface com temas, cores e componentes estéticos.

Estilo:

- **Tema Clam**: Utilizado para modernizar a interface gráfica, dando um visual mais limpo e profissional aos elementos gráficos.

10. Interface Gráfica

Uma interface gráfica foi adicionada ao projeto para tornar a interação com o sistema mais amigável e acessível, mesmo para usuários sem conhecimentos técnicos avançados. Desenvolvida em **Tkinter** (a biblioteca nativa de interface gráfica para Python), a interface oferece uma experiência de uso moderna e simplificada.

Principais Funcionalidades da Interface Gráfica:

1. **Seleção de Arquivo XML:**
 - O usuário pode facilmente selecionar o arquivo XML do currículo Lattes a ser processado usando um explorador de arquivos.
2. **Seleção da Pasta de Destino:**
 - O usuário define a pasta onde os arquivos gerados (planilhas Excel e JSON) serão salvos, garantindo uma organização clara dos resultados.
3. **Escolha do Formato de Saída:**
 - O usuário pode optar por gerar os resultados no formato **Excel** (planilhas com múltiplas abas) ou **JSON** (arquivo estruturado com os dados).
4. **Escolha do Tipo de Geração:**
 - **Detalhado**: Gera uma planilha ou arquivo JSON contendo informações completas sobre o pesquisador, organizadas em três abas:
 - **Dados Gerais**: Nome, nacionalidade e resumo do CV.

- **Artigos Publicados:** Informações detalhadas sobre os artigos publicados.
 - **Trabalhos em Eventos:** Informações detalhadas sobre os trabalhos apresentados em eventos.
 - **Resumo:** Atualiza a planilha "Dados Gerais" com o número de artigos publicados e trabalhos em eventos para cada pesquisador, verificando duplicidades e atualizando os dados, se necessário.
5. **Design Moderno e Estético:**
- A interface foi estilizada com o tema **clam** do Tkinter para oferecer uma aparência mais moderna e profissional, com cores personalizadas e espaçamento adequado para uma navegação confortável.
 - **Botões e componentes visuais** foram configurados para serem intuitivos e fáceis de usar, com feedback claro ao usuário sobre a execução das tarefas.

Benefícios da Interface Gráfica:

- **Facilidade de Uso:** A interface elimina a necessidade de executar comandos em linha de comando, facilitando o uso do sistema por qualquer pessoa, independentemente de sua experiência técnica.
- **Interatividade:** Com a seleção intuitiva de arquivos e pastas, o usuário tem controle completo sobre os dados processados e a localização dos resultados gerados.
- **Visualização e Personalização:** O usuário pode escolher entre gerar relatórios detalhados ou atualizar resumos quantitativos dos pesquisadores, tornando o sistema flexível e adaptável a diferentes necessidades.
- **Modularidade:** A interface foi projetada de forma modular, garantindo que seja fácil de manter e expandir com novas funcionalidades no futuro.

1. Planejamento de estratégias para o levantamento de requisitos do software junto ao cliente

Formulário:

➤ Objetivo do Sistema

Qual o objetivo principal do sistema?

- ☐ Consultar dados científicos de currículos Lattes.
- ☐ Gerar relatórios detalhados da produção acadêmica.
- ☐ Criar uma planilha de acompanhamento com dados numéricos de produção.
- ☐ Outro: _____

➤ Dados a Serem Consultados

Quais dados devem ser extraídos do currículo Lattes?

(Marque todos que se aplicam)

- ☐ Dados gerais (Nome, Nacionalidade, etc.)
- ☐ Artigos publicados
- ☐ Trabalhos em eventos
- ☐ Orientações concluídas
- ☐ Projetos de pesquisa
- ☐ Outro: _____

➤ Visualização dos Dados

Como você gostaria de visualizar as informações extraídas?

- ☐ Relatório detalhado com todos os dados.
- ☐ Planilha com um resumo numérico das produções.
- ☐ Ambos: relatório detalhado e resumo numérico.
- ☐ Outro: _____

Qual formato de saída você prefere?

- ☐ JSON
- ☐ Excel
- ☐ Outro: _____

➤ Atualização de Planilhas

Você gostaria que o sistema atualizasse uma planilha geral com os dados de todos os pesquisadores?

- ☐ Sim
- ☐ Não
- ☐ Outro: _____

➤ Frequência de Atualização

Com que frequência os dados precisam ser atualizados?

- ☐ Diário
- ☐ Semanal
- ☐ Mensal
- ☐ Sob demanda

☐ Outro: _____

➤ **Funcionalidades Extras**

Há algum recurso adicional que você gostaria de incluir no sistema?

➤ **Integração com Outros Sistemas**

Você necessita de integração com outros sistemas ou plataformas (ex.: Google Drive)?

☐ Sim, com: _____

☐ Não

➤ **Desafios Atuais**

Quais são os maiores desafios que você enfrenta ao lidar com esses dados atualmente?

➤ **Comentários Adicionais**

Por favor, insira qualquer outra sugestão ou requisito adicional que você considere relevante