

Plataforma Lattes: extração e tratamento de dados de produção docente

Autor: ARNOTT RAMOS CAIADO

Email: arnotrcaiado@fac.pe.senac.br

Coautor: Danilo Miguel de Souza

Email: danielomsouza@fac.pe.senac.br

Introdução

Este relato traz um modelo de extração e tratamento de dados da Plataforma Lattes. A plataforma tem seu uso para além de uma rede de interação entre pesquisadores. É muito frequente seu uso em seleções de professores, tanto na esfera pública quanto na privada, e também para acompanhamento da produção científica docente por gestores de IES (Instituições de Ensino Superior). O modelo apresentado visa otimizar o tempo gasto na verificação individual e manual de Currículos Lattes para obtenção de informações precisas. Os experimentos e protótipos foram realizados em Python e bibliotecas complementares.

Métodos e técnicas

Trata-se de uma pesquisa aplicada, que segundo Gil (2010), “abrange estudos elaborados com a finalidade de resolver problemas no âmbito das sociedades em que os pesquisadores vivem”, bem como também pesquisas bibliográficas, com o apoio de artigos científicos, sites e artigos técnicos. A figura 1 mostra uma visão geral do método:



Figura 1 - Representação do Método
Fonte: os autores

Os estudos iniciaram com a plataforma Lattes, sua estrutura e opções de exportação de dados, seguido de revisões da literatura e aprofundando no entendimento da estrutura dos arquivos XML (*eXtensible Markup Language*) de currículos, utilizando a linguagem *Python*. O objetivo desta etapa foi explorar o funcionamento da plataforma Lattes, realizando experimentos para compreensão da estrutura dos dados. Em seguida, seguiram-se estudos experimentais para



XVII CONGRESSO
INTERNACIONAL
DE TECNOLOGIA
NA EDUCAÇÃO

Dilemas e desafios de um futuro presente: o que esperar da educação?

22 e 23 | setembro | 21

contagem das produções por categoria existentes nos currículos.

Por fim, procurou-se esboçar e testar um protótipo *web*, através de requisições *http* (forma comunicação *web*), linguagem *Python*, *framework Flask* e plataforma *Pythonanywhere*. O

desenvolvimento, estudos, reuniões e testes contou com a participação dos autores remotamente, utilizando principalmente as ferramentas google (google meet e google classroom).

Os experimentos

A partir da análise de currículos XML exportados da plataforma Lattes, compreendeu-se que o sistema distribui as informações dos usuários em várias seções, e que o conteúdo destas seções, por sua vez, são organizados por categorias e subcategorias. Sendo assim, o sistema Lattes armazena dados sobre diversos tipos de produção, como: produções bibliográficas, produções técnicas, artísticas/culturais, dentre outras. Também é possível identificar os produtos de inovação patenteados pelo usuário, os cursos de curta duração que foram realizados e as participações em eventos e congressos, de forma que, através do currículo Lattes é possível obter um acompanhamento geral, de tudo quanto o estudante/pesquisador estiver produzindo no ambiente acadêmico e profissional.

O XML (*eXtensible Markup Language*) é uma linguagem que, segundo Almeida (2002), é pautada em marcação. XML deriva do SGML (*Standard Generalized Markup Language*) e foi idealizada por Jon Bosak, engenheiro da Sun Microsystems. Embora XML tenha várias semelhanças com a linguagem HTML (*HyperText Markup Language*), possui propósitos diferentes. Geralmente, HTML é usado para exibir dados e estruturar páginas web, enquanto XML é usado para guardar e transportar dados. Os documentos XML são estruturados em tags e subtags que guardam chaves e valores (ALMEIDA, 2002). Seu componente principal é o "elemento", texto que vem entre os sinais "<" e ">". Como, por exemplo, "<raiz>" e "</raiz>". Dentro do elemento pode-se ter um texto bruto, outras marcações, ou até algo intercalado entre texto bruto e marcações (ALMEIDA, 2002). A tag raiz (ou root) é a principal, nela se encontram as sub tags (ou tags filhas), todas as tags tem sua abertura <> e seu fechamento </> como demonstrado no Quadro 1.

Quadro 1 - Estrutura de um arquivo XML



Fonte: os
autores

Uma funcionalidade disponível na plataforma Lattes é o *download* do Currículo em formato XML. Sua estrutura, representada na Figura 3, possui tags e subtags com algumas informações que aparecem apenas uma vez (CURRICULO VITAE, DADOS-GERAIS) e outras sub-tags que aparecem tantas vezes quantas forem as produções da pessoa. Por exemplo, na figura 3, ARTIGO-PUBLICADO aparece seis vezes, sendo cada uma delas com informações



específicas do artigo (data, periódico, título, etc.). A contagem destas ocorrências por sub-tag nos dá o quantitativo daquela categoria de produção.

Key	Type	Value
▼ CURRICULO-VITAE	Element	(5 Items)
SISTEMA-ORIGEM-XML	Attribute	LATTES_OFFLINE
NUMERO-IDENTIFICADOR	Attribute	6629441055656636
DATA-ATUALIZACAO	Attribute	20092020
HORA-ATUALIZACAO	Attribute	210320
▶ DADOS-GERAIS	Element	(6 Items)
▼ PRODUCAO-BIBLIOGRAFICA	Element	(2 Items)
▶ TRABALHOS-EM-EVENTOS	Element	(12 Items)
▼ ARTIGOS-PUBLICADOS	Element	(6 Items)
▶ ARTIGO-PUBLICADO	Element	(7 Items)
▶ ARTIGO-PUBLICADO	Element	(8 Items)
▶ ARTIGO-PUBLICADO	Element	(8 Items)
▶ ARTIGO-PUBLICADO	Element	(8 Items)
▶ ARTIGO-PUBLICADO	Element	(9 Items)
▶ ARTIGO-PUBLICADO	Element	(15 Items)
▼ PRODUCAO-TECNICA	Element	(6 Items)
▶ SOFTWARE	Element	(7 Items)
▶ SOFTWARE	Element	(7 Items)
▶ PRODUTO-TECNOLOGICO	Element	(5 Items)
▶ TRABALHO-TECNICO	Element	(6 Items)
▶ TRABALHO-TECNICO	Element	(7 Items)
▶ DEMAIS-TIPOS-DE-PRODUCAO-TECNICA	Element	(46 Items)
▼ OUTRA-PRODUCAO	Element	(1 Item)

Figura 3 - Estrutura de Arquivo XML de Currículo Fonte: os autores

Assim, após diversas simulações e compreensão do modelo, chegou-se a um algoritmo que percorre os dados e os quantifica por categoria. O algoritmo, expresso através de um fluxograma (Diagrama BPMN), exibido na figura 4, consiste basicamente de estruturas de decisão usadas para que seja feita a escolha das categorias, seguido de laços de repetição que são usados para percorrer as subcategorias, realizar a contagem das produções e exibir os principais

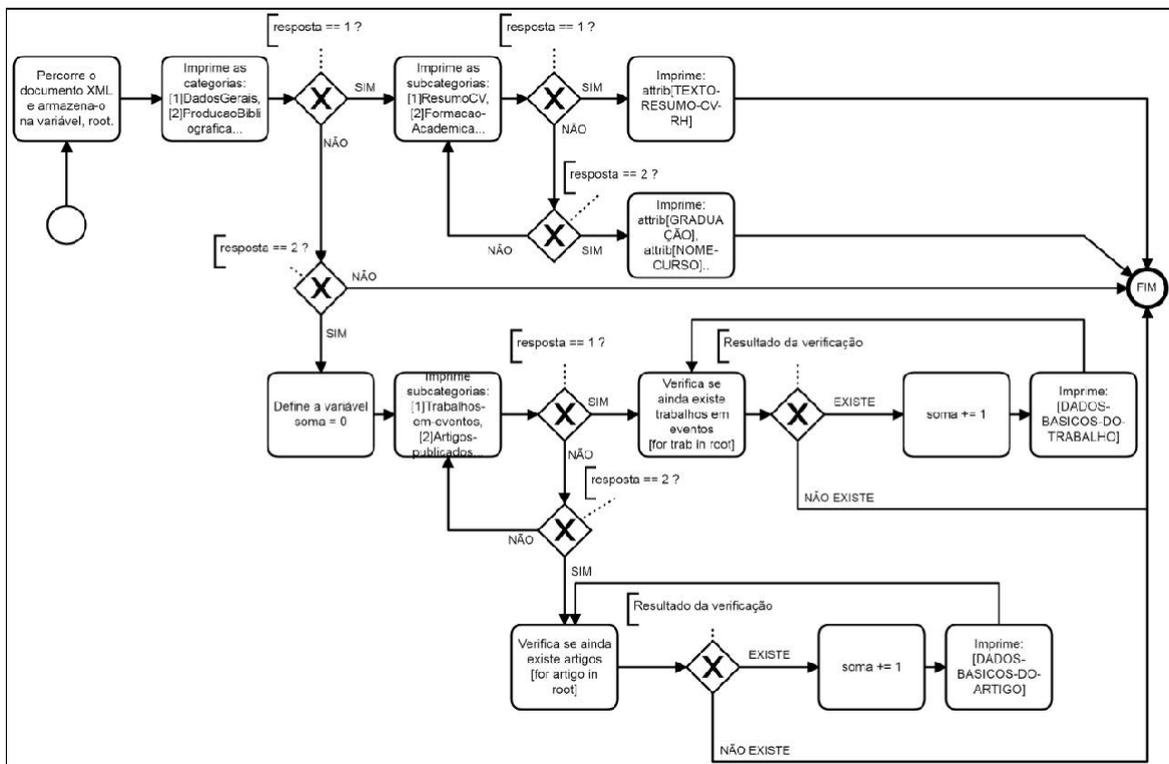


Figura 4 - Exemplo de algoritmo para tratar os dados XML Fonte: os autores

dados. Havendo mais produções, o mesmo ciclo é iterado, caso contrário, encerra-se a leitura.

Também foi possível esboçar um modelo conceitual de banco de dados relacional, através de um Diagrama Entidade-Relacionamento (Figura 5):

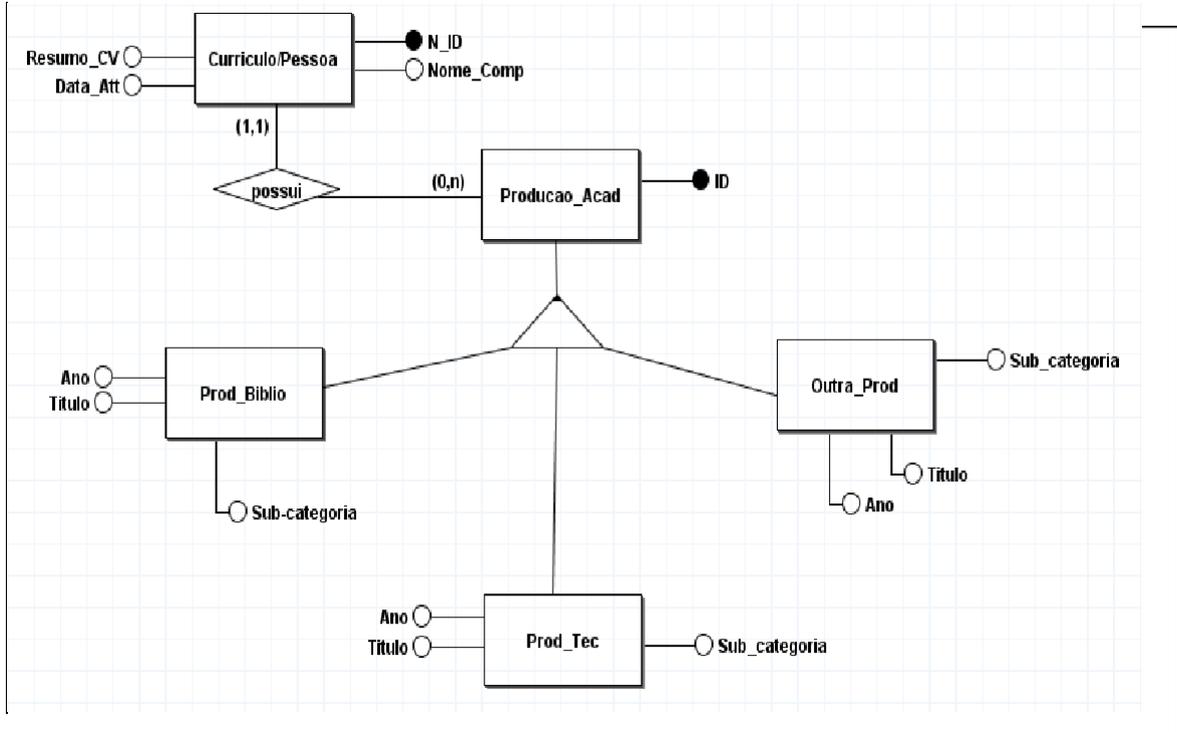


Figura 5 - Diagrama Entidade-Relacionamento (Modelo Conceitual).
Fonte: os autores

Na figura 5, pode-se observar um esboço de como os dados podem ser armazenados. Na entidade “Currículo/Pessoa”, encontram-se as informações relacionadas aos dados gerais, tais como Número de Identificação do Currículo (N_ID), Nome Completo (Nome_Comp), Data da última atualização (Data_Att) e um resumo do currículo (Resumo_CV). Em seguida, encontra-se a entidade “Producao_Acad”, ligada a entidade “Currículo/Pessoa” por um relacionamento (1,n). Em “Producao_Acad” encontra-se um atributo (ID) que identifica cada produção de forma única, podendo essa produção ser armazenada em alguma das entidades herdeiras, sendo elas a representação das três principais categorias de produção científica, que são Produções Bibliográficas (Prod_Biblio), Produções Técnicas (Prod_Tec) e Outras Produções (Outra_Prod). Cada entidade herdeira, possui os atributos de ano de produção (Ano), título da produção (Título) e Sub Categoria (Sub_Categoria). No atributo de sub categoria, as produções poderão ser identificadas de maneira mais específica (Ex.: Produção Bibliográfica > Sub Categoria: “Trabalhos em Eventos”, “Artigos Publicados” ou “Livros e Capítulos”).

Prototipagem

Em um protótipo inicial, as telas desenvolvidas utilizam *dashboards* como no exemplo da figura 6. Estão visíveis no exemplo categorias de produção e o período de tempo. Neste exemplo, pode-se identificar que o profissional realizou 6 produções entre os anos de 2017 e 2019, sendo 83,3% equivalentes a artigos e 16,7%, a livros publicados ou organizados.

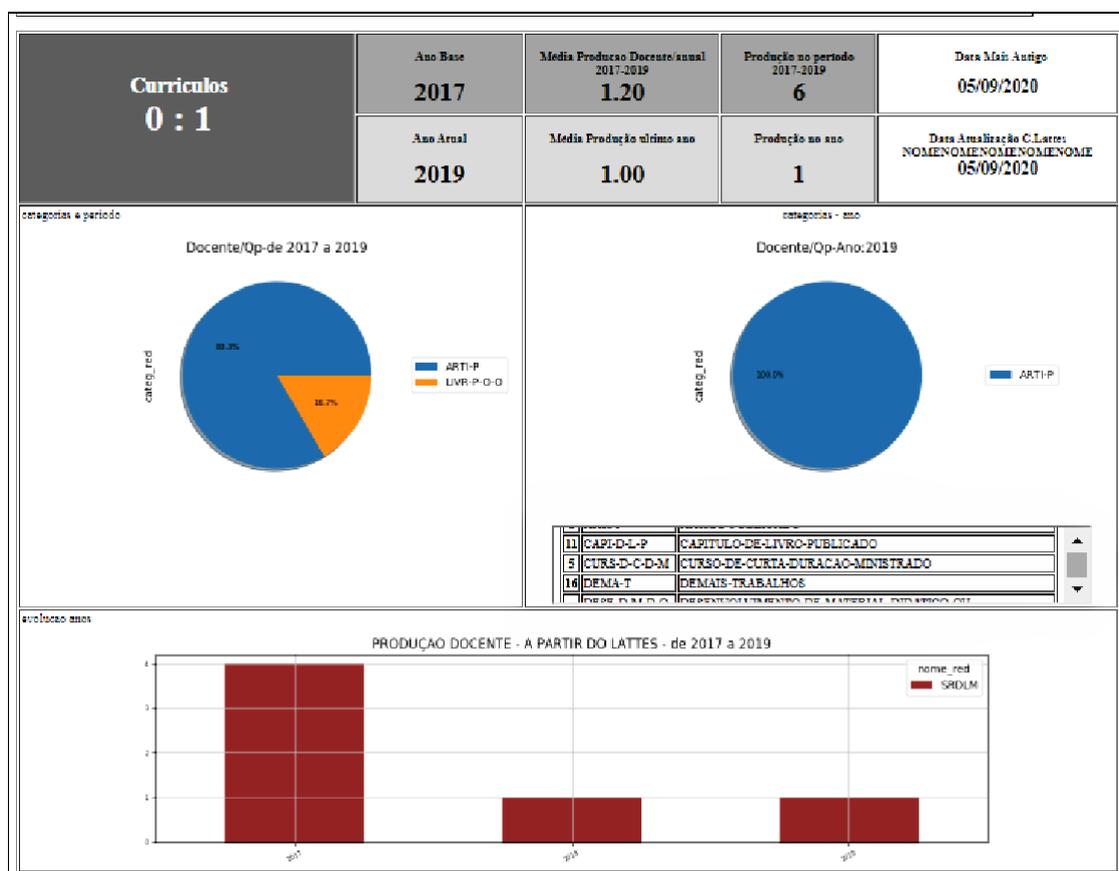


Figura 6 - Exemplo de funcionamento do protótipo (Dashboard).

Fonte: os autores. Disponível em <https://arnotti9i.wixsite.com/i9i-app-bla/prod-lattes> acesso em agosto de 2021

Resultados, impacto das ações

A proposta de solução a ser desenvolvida visa facilitar a extração e exploração de dados de currículos cadastrados na base Lattes. Há uma forte convicção que esta solução poderá contribuir diretamente nos processos de gestão acadêmica, com ênfase no acompanhamento da produção científica do corpo docente e na seleção de novos professores. A princípio, a solução / *software proposto* poderá ser testada pela Faculdade SENAC. Os benefícios proporcionados pelo sistema serão muitos, como a agilidade para a análise de currículos Lattes, visão panorâmica da quantidade da produção realizada pela instituição e por cada Professor, possibilidade de visualização dos dados em *dashboards* e filtragem das informações por categorias.

Foi desenvolvido um tutorial sobre “Desenvolvimento, hospedagem e publicação de um



site, usando Pythonanywhere, Python, Flask Framework e MySQL” que proporcionou entender de

forma prática o funcionamento do ambiente de desenvolvimento e hospedagem, Pythonanywhere. Foram explorados os recursos da plataforma Lattes, com o objetivo de conhecer mais o funcionamento e as possibilidades oferecidas pela plataforma. Também foi possível analisar a estrutura do currículo Lattes, e entender o formato de arquivo XML.

Superações e Aprendizados

Durante a fase inicial do estudo, as maiores dificuldades encontradas foram, a ausência de conhecimento prévio na linguagem Python e as restrições ocasionadas pela pandemia. Também houve dificuldade na realização do cadastro na plataforma Lattes, mas que foram superadas com experimentações, pesquisas e orientações, sendo possível realizar o cadastramento e também entender de modo mais aprofundado o funcionamento do Sistema Lattes. Através das pesquisas realizadas em vários sites de tecnologia, foi possível obter um conhecimento introdutório na linguagem Python e no ambiente de desenvolvimento e hospedagem, PythonAnywhere. Foi possível aprender a construir artigos científicos seguindo as normas da ABNT e em concílio com as atividades regulares do curso, inquiriu-se a utilização do *framework* Flask. Portanto, pode-se afirmar que a experiência tem sido exitosa e muito produtiva para o desenvolvimento da carreira acadêmica e profissional dos participantes.

Referências

ALMEIDA, M. B. **Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares**. Ci. Inf., Brasília, v. 31, n. 2, p. 5-13, Aug. 2002. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652002000200001&Ing=en&nrm=iso>. acesso em 19 de fevereiro de 2021.

I9I.APP - PRODLATTES . Protótipo em funcionamento prévio. Disponível em <https://arnotti9i.wixsite.com/i9i-app-bla/prod-lattes> acesso em agosto de 2021.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010.

MACHADO, G. B., **Desenvolvimento, hospedagem e publicação de um site, usando Pythonanywhere, Python, Flask Framework e MySQL**. Disponível em: <<https://medium.com/@guilhermebertonimachado/desenvolvimento-hospedagem-e-publica%C3%A7%C3%A3o-de-um-site-usando-pythonanywhere-python-flask-framework-10effe5e7d1e>> Acesso em 12 de novembro de 2020.