

PESQUISA E DESENVOLVIMENTO DE SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÕES COM CONTEXTO SEMÂNTICO

 **Daniel Schreiber Guimarães**  **Sergio Lifschitz**

1/4

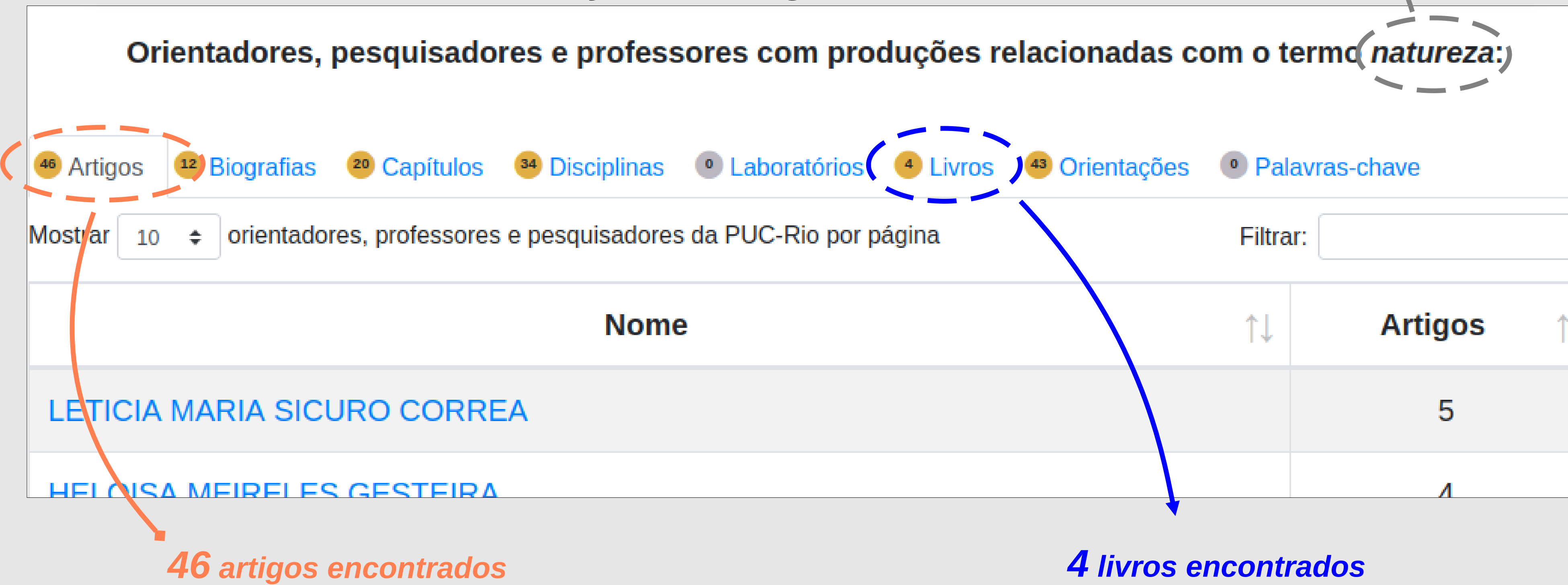
Resumo

Essa pesquisa teve como objetivo desenvolver um filtro semântico para o sistema de recuperação de informação intitulado Quem@PUC. Este sistema permite encontrar professores e pesquisadores da PUC-Rio através de suas produções acadêmicas cadastradas nos seus currículos Lattes e suas disciplinas lecionadas na universidade.

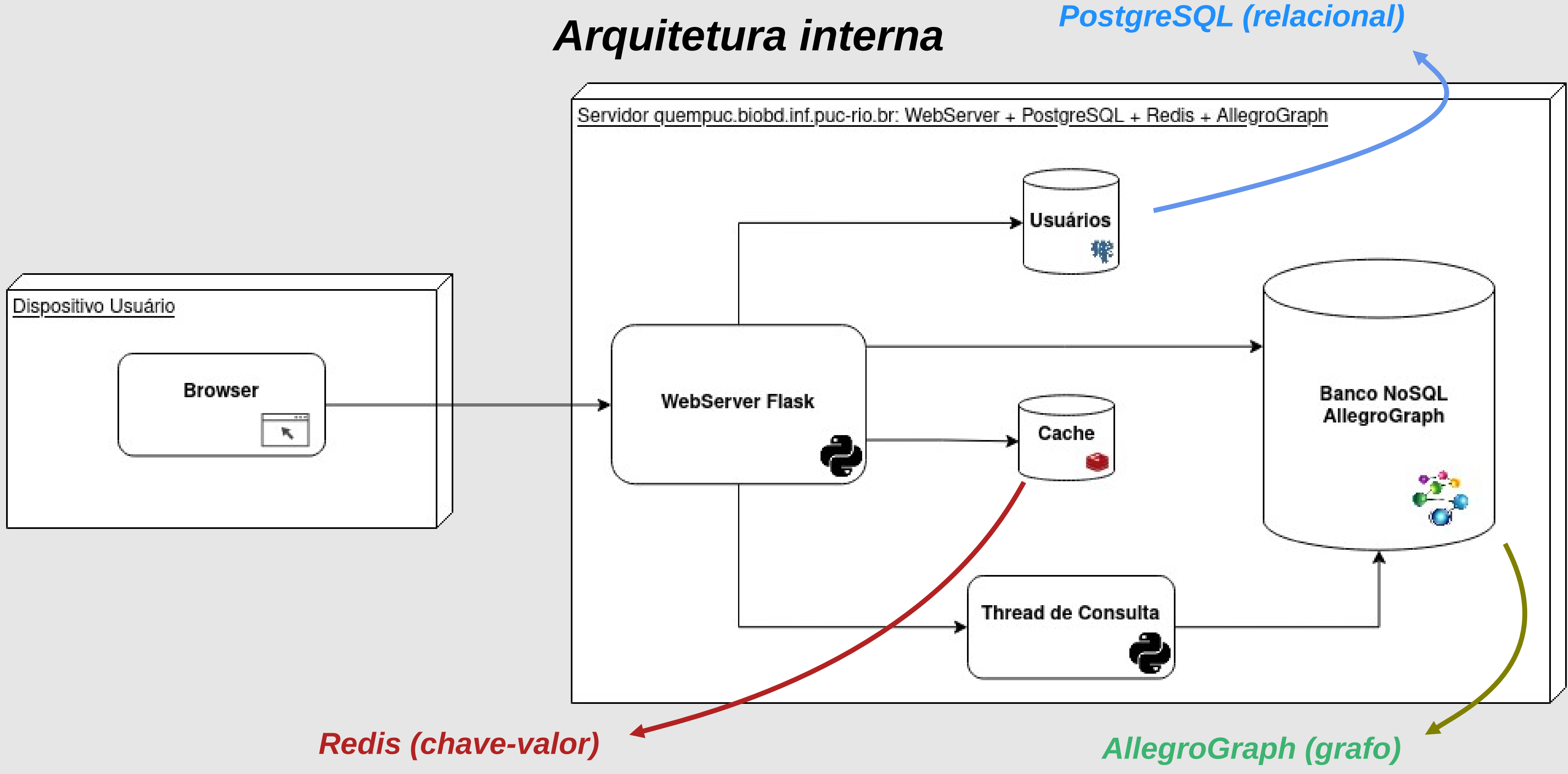
Foi criada uma versão desse sistema com as mesmas funcionalidades originais, mas com um filtro semântico que exhibe somente professores e produções com temática de meio ambiente, e foi disponibilizado para o Núcleo Interdisciplinar de Meio Ambiente (NIMA), sendo intitulado Busca@NIMA.

Introdução

Visualização da Página Quem@PUC



Arquitetura interna



PESQUISA E DESENVOLVIMENTO DE SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÕES COM CONTEXTO SEMÂNTICO

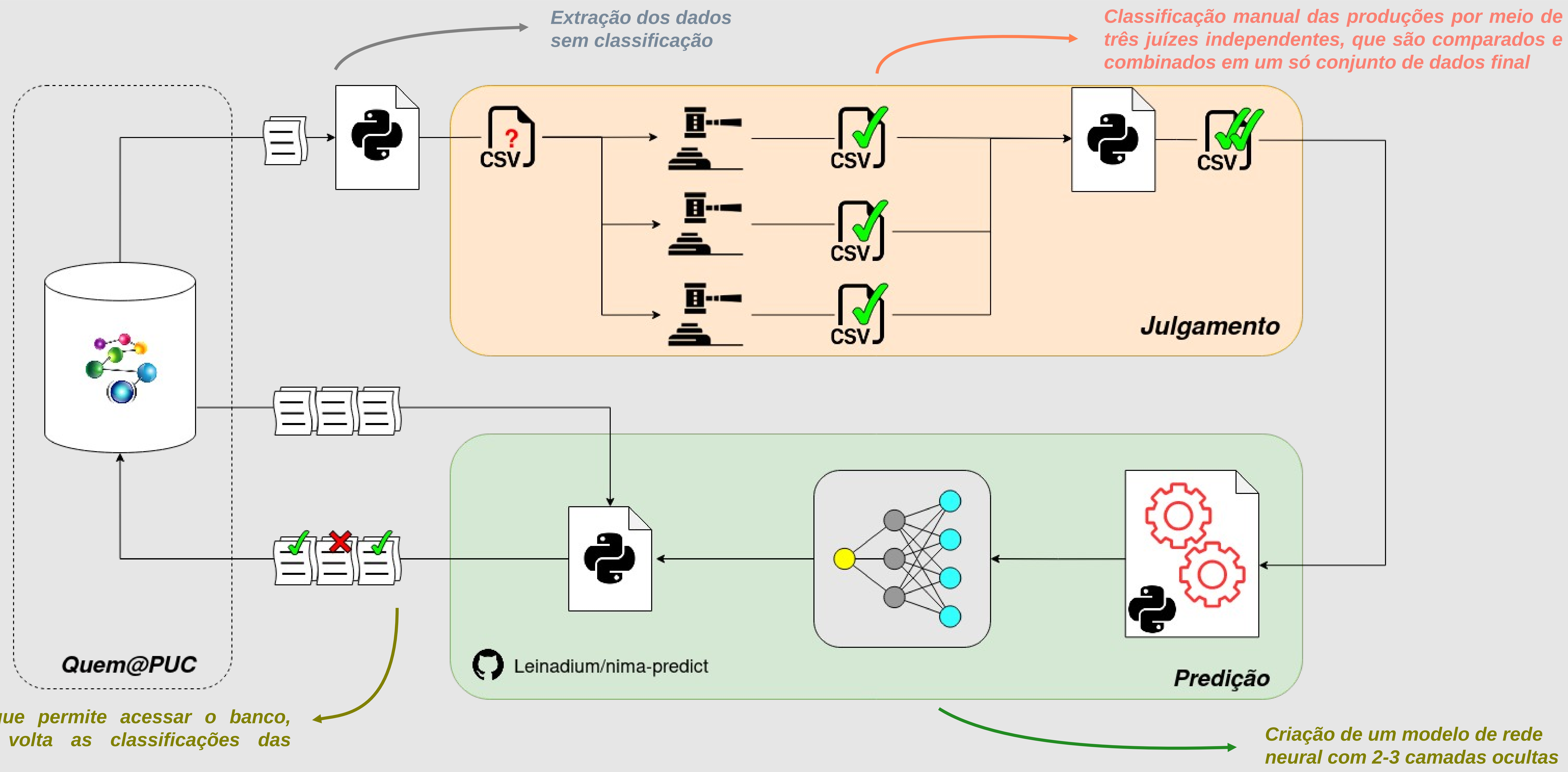
🎓 **Daniel Schreiber Guimarães** 🧑‍🔬 **Sergio Lifschitz**

2/4

Metodologia

Foi extraída uma grande quantidade de produções da base de dados existente sem nenhuma classificação para passar por uma fase de classificação manual (“Julgamento”), em que três indivíduos (“juízes”) fazem suas classificações e os seus resultados são comparados, de forma a gerar um arquivo final com todas essas produções classificadas.

Esse arquivo é utilizado para treinar um modelo de rede neural profunda com 3 camadas ocultas, que após treinado classifica automaticamente todas as outras produções do banco, armazenando os resultados no próprio banco.



PESQUISA E DESENVOLVIMENTO DE SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÕES COM CONTEXTO SEMÂNTICO

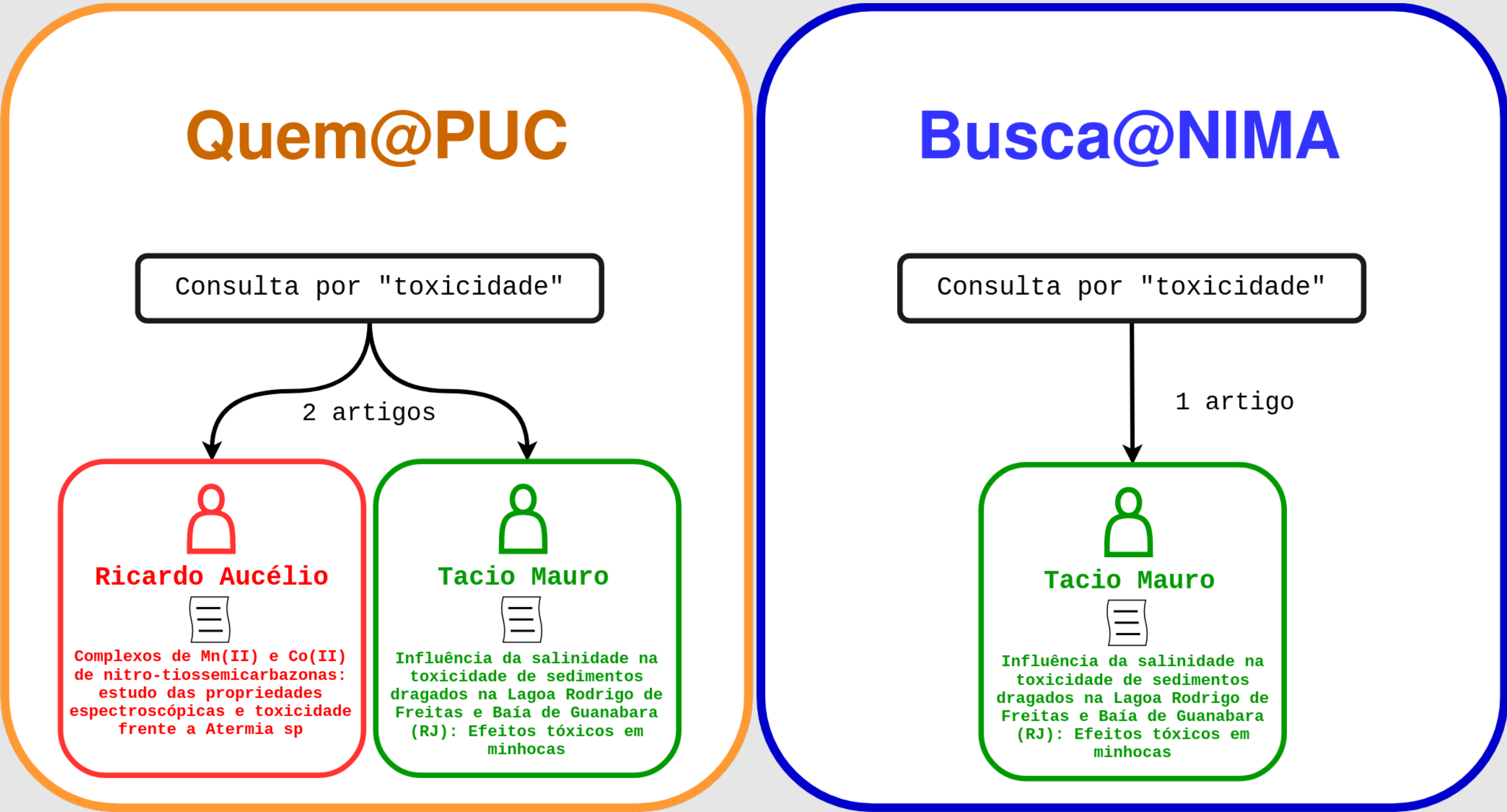
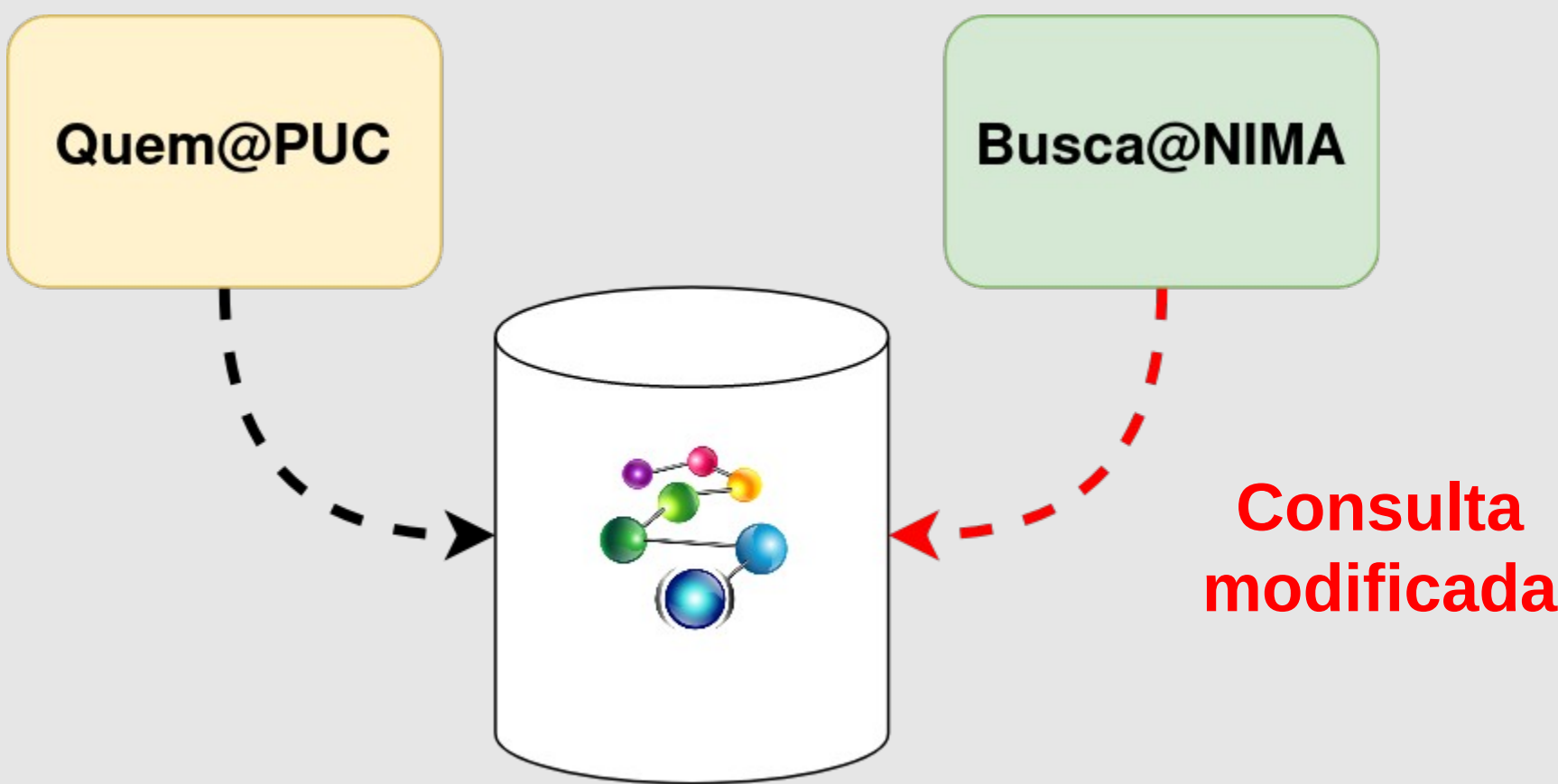
 **Daniel Schreiber Guimarães**  **Sergio Lifschitz**

3/4

Resultados

Ao lado encontra-se um exemplo de uma mesma consulta nos dois sistemas. No Quem@PUC, a consulta exibiu dois professores, cada um com dois artigos. Porém, no Busca@NIMA, somente um dos professores foi exibido, pois o artigo do *Ricardo Aucélio* não havia uma temática de meio ambiente, ao contrário do artigo do *Tacio Mauro*.

Ambos os sistemas acessam o banco de dados AllegroGraph, porém o Busca@NIMA possui a diferença de utilizar os valores de classificação do modelo de rede neural armazenados.



PESQUISA E DESENVOLVIMENTO DE SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÕES COM CONTEXTO SEMÂNTICO

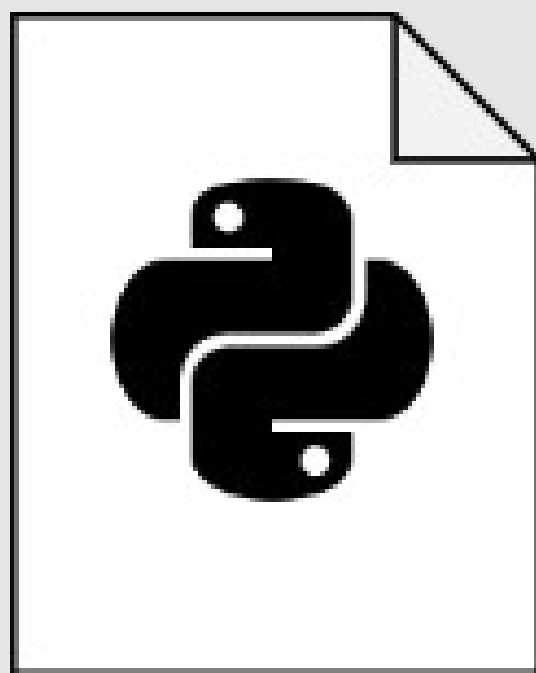
 **Daniel Schreiber Guimarães**  **Sergio Lifschitz**

4/4

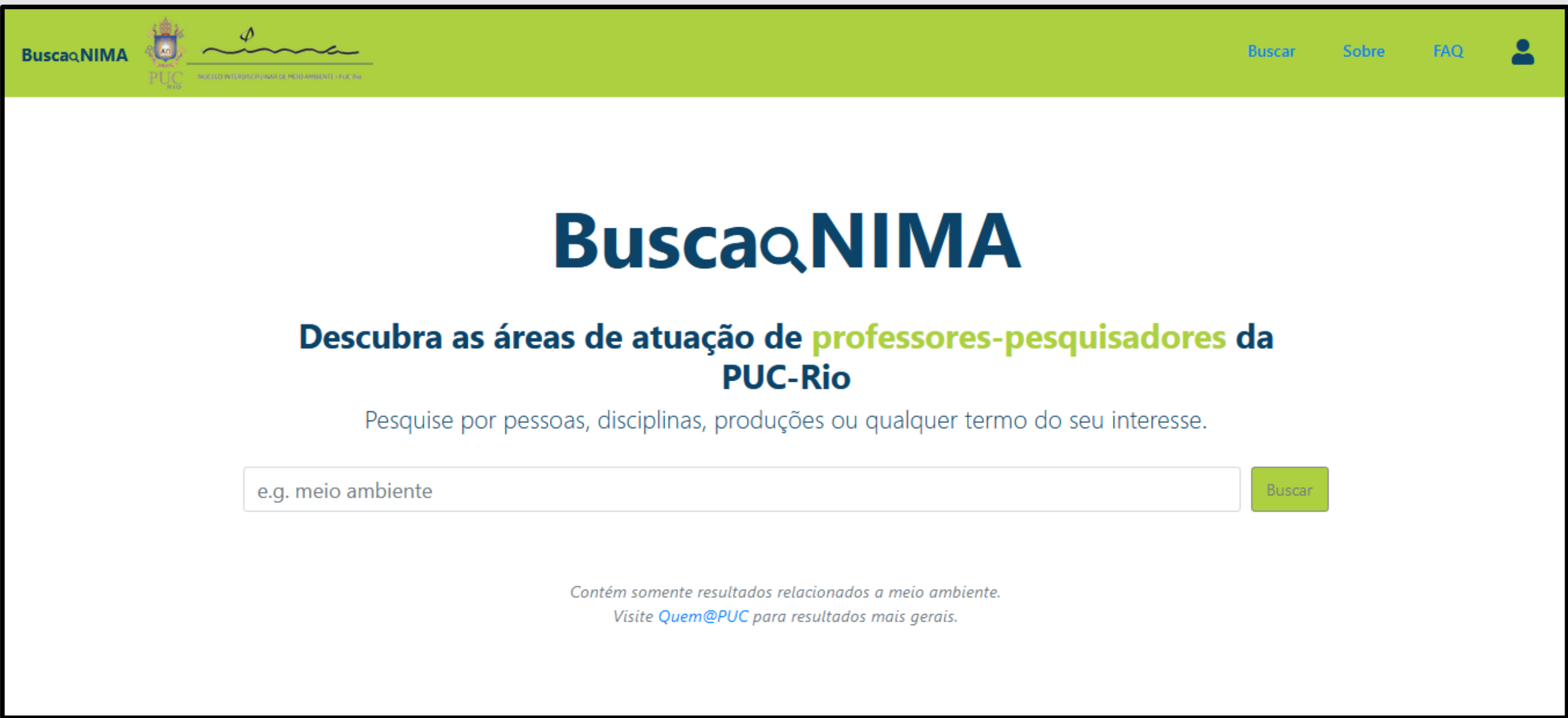
Resultados

O sistema Busca@NIMA pode ser acessado hoje através da plataforma do NIMA, e o Quem@PUC pode ser acessado pela página inicial do Ciclo Básico (CBCTC) da PUC-Rio, ou através das URLs ao lado.

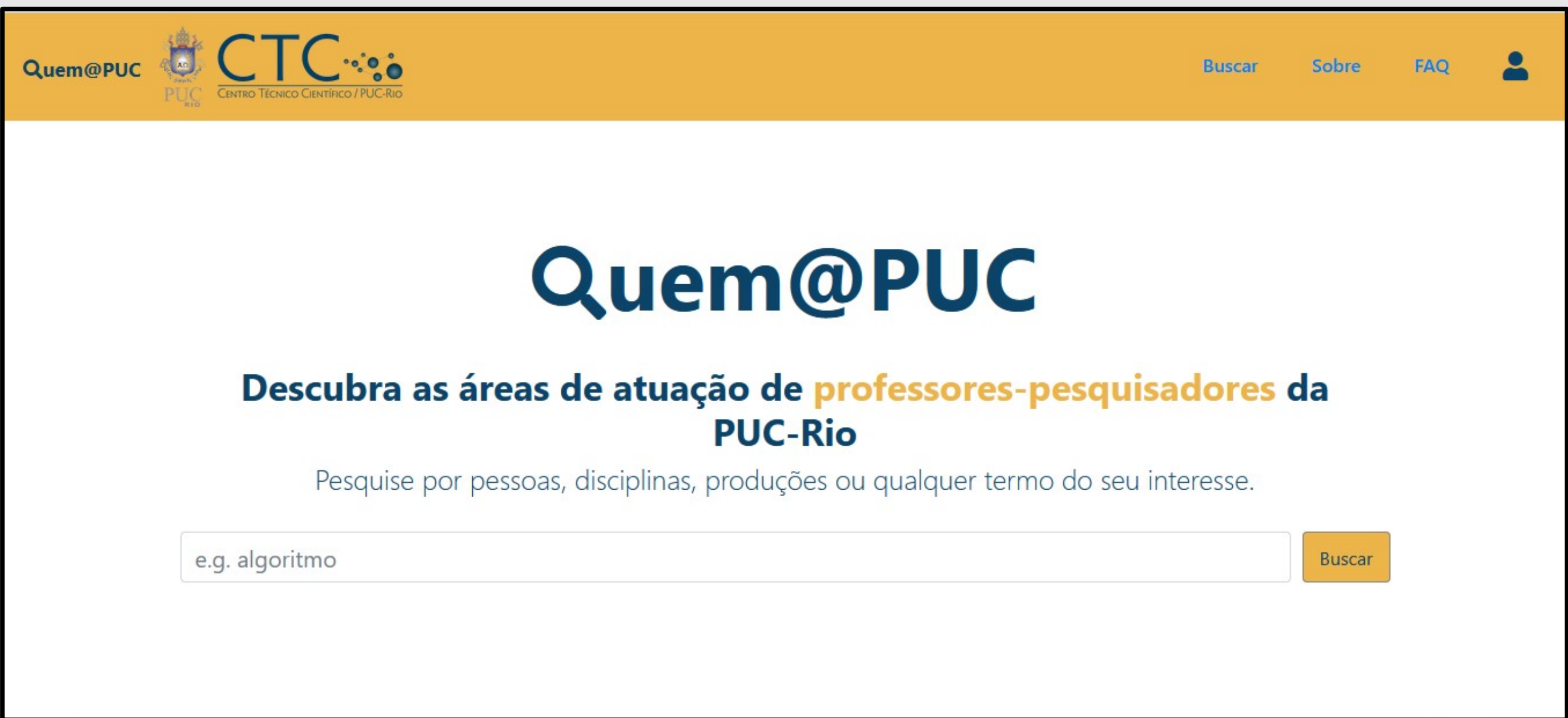
O modelo neural, assim como o script de coleta e inserção das produções classificadas no banco está publicada no GitHub, disponível na URL abaixo.



 github.com/Leinadium/nima-predict



 buscanima.biobd.inf.puc-rio.br



 quempuc.biobd.inf.puc-rio.br

Referências

M. Salgueiro *et al.* "Sistemas de Recuperação de Informações Aplicados à Produções Acadêmicas". *Anais Estendidos do XXXVI Simpósio Brasileiro de Bancos de Dados*, Rio de Janeiro, 2021, pág. 43-48, Disponível em: https://doi.org/10.5753/sbbd_estendido.2021.18161.

A. Rêgo, D. Guimarães, M. Villas, S. Lifschitz. "Busc@NIMA: um Sistema de Recuperação de Informações com Semântica de Meio Ambiente". *Anais Estendidos do XXXVII Simpósio Brasileiro de Bancos de Dados*, Búzios, 2022, pág. 47-52. Disponível em: https://doi.org/10.5753/sbbd_estendido.2022.21842.

Plataforma Lattes. CNPq. Disponível em: lattes.cnpq.br.

AllegroGraph. Franz Inc. Disponível em: allegrograph.com