



IME-USP

# Genealogia Acadêmica Lattes

Universidade de São Paulo  
Instituto de Matemática e Estatística

aluno: Edson Kiyohiro Miyahara  
orientador: Roberto Marcondes Cesar Junior  
coorientador: Jesús Pascual Mena-Chalco

## Introdução

Atualmente, com o grande crescimento de informações disponíveis na internet, torna-se um campo fértil para o desenvolvimento de diversas áreas da ciência da computação, dentre elas a mineração de dados que vem ganhando grande destaque e importância nesse cenário.

Este trabalho foca sobre as informações contidas nos currículos Lattes presente na Plataforma Lattes<sup>1</sup> para a geração automática da árvore de genealogia dos acadêmicos, sendo então possível realizar a sua visualização por meio de ferramentas complementares ou externas.

## Motivação

Esta é uma proposta nova no Brasil e está inserida no contexto da prospecção de dados para análise de dados da genealogia acadêmica. Esta informação é de importante interesse para a análise da inter-relação de orientações.

## Objetivos

Desenvolver uma ferramenta para geração automática da árvore genealógica para cientistas e/ou acadêmicos cadastrados na Plataforma Lattes, através de suas relações de orientação ou supervisão concluídas.

No intuito do desenvolvimento desse projeto, tem-se como principais objetivos:

-Desenvolver um sistema de armazenamento dinâmico do Currícula Vitae (CVs), com o intuito de reduzir o número de acessos à rede.

-Tratamento de similaridades entre nomes de uma mesma pessoa escritos de formas diferentes para evitar possíveis ambiguidades.

-Desenvolver um algoritmo para geração da árvore de genealogia acadêmica descendente e ascendente.

-Exportação das informações para arquivos que permitam a visualização e análise por ferramentas externas e/ou complementares.

-Geração da árvore em um mapa geográfico utilizando a API do Google Maps<sup>2</sup>.

## O Projeto

O projeto está sendo desenvolvido em Python<sup>3</sup> no ambiente Gnu/Linux, constitui de 5 módulos.

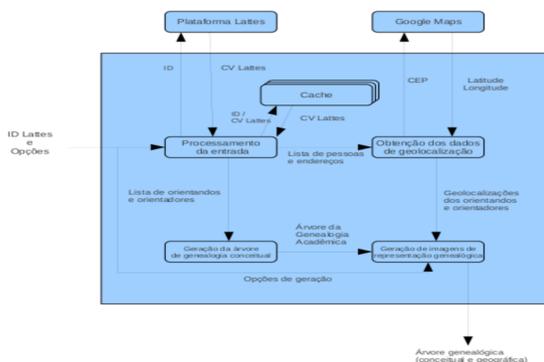


Figura1: Fluxograma dos processos envolvidos

## Cache

É um módulo bastante simples, ele recebe um ID Lattes e verifica se o currículo Lattes correspondente está armazenado, armazena e recupera os currículos.

## Processamento da entrada

Esse módulo é responsável tanto pelo processamento da entrada fornecida pela linha de comando como pelo pré-processamento dos currículos que é realizado pelo método parser do scriptLattes, os currículos podem ser obtidos do cache ou da Plataforma Lattes. Para todo CV são extraídas tanto a lista de orientações concluídas quanto a lista de orientadores/supervisores cadastrados na Plataforma Lattes. Aqui ocorre também, a verificação de similaridade entre nomes evitando possíveis ambiguidades e o reprocessamento dos currículos.

```
/* recebe 2 nomes completos */
ComparaNomes (nome1, nome2)
  n1 ← array de palavras de nome1,
  n2 ← array de palavras de nome2
  se sobrenome de nome1 == sobrenome de nome2
    então se nome de nome1 == nome de nome2
      então se ComparaAux (n1, n2)
        retorna True
    senão se nome de nome1 ou nome2 está abreviado
      então se a 1º letra de nome1 == à de nome2
        então se ComparaAux(n1, n2)
          retorna True
    retorna False
/* recebe 2 arrays de nomes */
ComparaAux (nome1, nome2)
  se o nº de palavras em nome1 > à do nome2
    então nome1 ↔ nome2
  se o nº de palavras em nome1-2 == 0
    então retorna True
  k ← 1, corretos ← 0, n ← nº de palavras de nome1
  m ← nº de palavras de nome2
  para i em [1 ... n-2] faça
    para j em [k ... m-2] faça
      se nome1[i][0] == nome2[j][0]
        então k ← j+1, corretos ← corretos+ 1, break
  se corretos/(n-2) >= 0.9
    então retorna True
  senão retorna False
```

## Geração da árvore de genealogia conceitual

Neste módulo é utilizada a ferramenta Graph-Tool<sup>4</sup> para criar e manipular o grafo, e também é aqui que a árvore de genealogia é exportada para o formato PNG, GraphML e JSON.

## Obtenção dos dados de geolocalização

Este módulo utiliza métodos do scriptLattes<sup>5</sup> para a obtenção das coordenadas geográficas: Latitude e Longitude.

## Geração de imagens de representação geográfica

Neste módulo é utilizado a API do Google Maps e as informações salvas em formato JSON para a geração da representação que nem sempre representa a árvore toda, pois um indivíduo pode

não ter um currículo registrado na Plataforma Lattes.

## Resultados

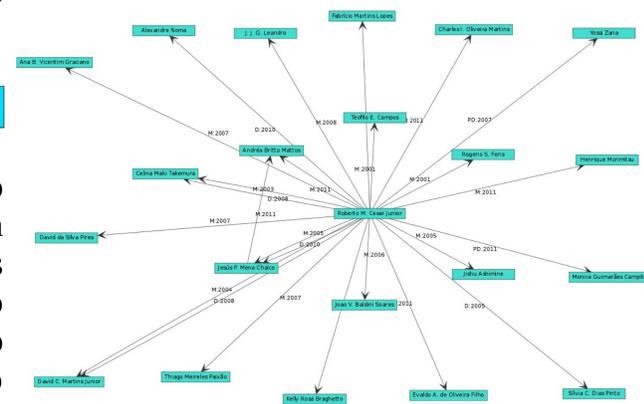


Figura 2: Árvore de genealogia acadêmica do prof. Dr. Roberto M. Cesar Jr

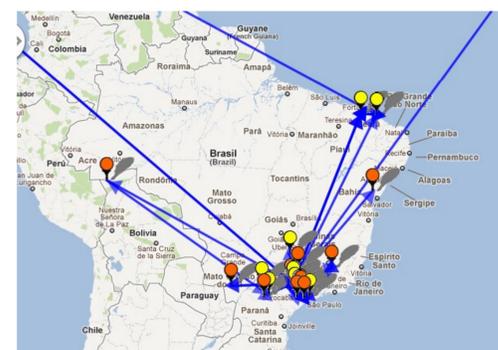


Figura 3: Representação geográfica da árvore de genealogia acadêmica do Prof. Dr. Luciano da Fontoura Costa

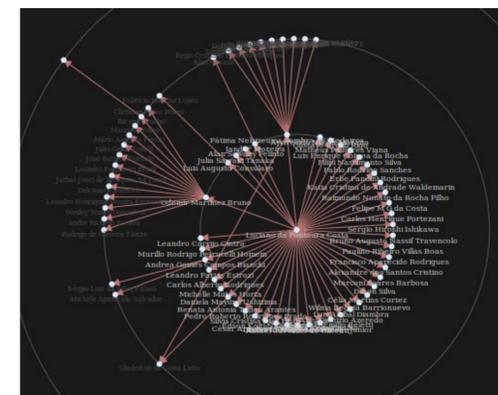


Figura 4: Representação interativa da árvore de genealogia acadêmica do Prof. Dr. Luciano da Fontoura Costa, gerado por uma ferramenta externa (InfoVis Toolkit) usando os dados salvo em formato JSON

## Limitações

Dada a falta de padronização no preenchimento de dados dos CVs Lattes, alguns nós correspondentes a orientandos/orientadores não são expandidos devido à inexistência dos ID Lattes. Atualmente deve se informar o nível de expansão da árvores de genealogia a ser obtida.

## Referências

[1] <http://lattes.cnpq.br/>  
[2] <http://code.google.com/apis/maps/index.html>  
[3] <http://www.python.org/>  
[4] <http://projects.skewed.de/graph-tool/>  
[5] <http://scriptlattes.sourceforge.net/>

