Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Elias da Silva Barroso Soares

Monografia de Projeto Orientado em Computação II

The Science Tree System

Belo Horizonte
2016 / 2º semestre
Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Curso de Bacharelado em Ciência da Computação

The Science Tree System

por

Elias da Silva Barroso Soares

Monografia de Projeto Orientado em Computação II

Apresentado como requisito da disciplina de Projeto Orientado em Computação II do Curso de Bacharelado em Ciência da Computação da UFMG

Prof. Dr. ... *Fabrício Benevenuto*
Orientador
Prof. Dr. ... *Alberto Laender*
Co-orientador

Belo Horizonte
2016 / 2º semestre

À minha mãe,
à Carolina Márcia,
ao Fabrício Benevenuto,
aos colegas de curso,
e ao povo brasileiro
dedico este trabalho.

AGRADECIMENTOS

Inicialmente quero agradecer à minha mãe, por sempre acreditar em mim e pelo amor incondicional.

À minha namorada, Carolina, por me acompanhar por toda a trajetória acadêmica, sempre me incentivando.

Ao meu orientador, Fabrício, por acreditar em mim, quando nem eu o fazia.

Aos colegas de percurso, por terem tornado essa caminhada divertida e satisfatória.

Ao povo brasileiro, que com o seu suor diário, deram-me algo que eu nunca conseguirei retribuir.

À Fundação Universitária Mendes Pimentel, que possibilitou minha residência e permanência em Belo Horizonte, tornando possível eu me graduar.

E finalmente, à Deus ou à Entidade superior, a qual me me guiou para os caminhos certos que me levaram ao final dessa jornada

“Se eu vi mais longe, foi por estar sobre ombros de gigantes.”
Isaac Newton

RESUMO

Ao longo da história, muitos pesquisadores forneceram contribuições importantes à ciência, avançando tanto no conhecimento como em termos de orientação de novos cientistas. Identificar e estudar a formação de pesquisadores ao longo dos anos é uma tarefa desafiadora, pois os repositórios atuais de teses e dissertações são catalogados de forma descentralizada através de muitas bibliotecas digitais locais, com informações incompletas ou algumas vezes erradas. Essa monografia é parte de um projeto onde queremos construir um grande repositório de árvores acadêmicas, registrando a genealogia dos pesquisadores em todas as áreas e países. Foi coletado várias bases de dados e bibliotecas digitais, como a NDLTD, com elas foi realizado um processo de desambiguação e gerada uma nova base de dados. O sistema desenvolvido para essa monografia tem o objetivo de possibilitar a visualização das árvores e a correção dos algoritmos de desambiguação, onde os usuários editariam as informações. A longo prazo queremos realizar diversas análises sobre as árvores genealógicas e as interações sobre as mesmas.

Palavras-chave: *árvores acadêmicas, sistema web, biblioteca digital.*

ABSTRACT

Along the history, many researchers provided remarkable contributions to science, not only advancing knowledge but also in terms of mentoring new scientists. Identifying and studying the formation of researchers over the years is a challenging task as current repositories of theses and dissertations are cataloged in a decentralized way through many local digital libraries, with incomplete or sometimes wrong information. This monograph is part of a project where we want to building a large repository that records the academic genealogy of researchers across fields and countries. It was collected several databases and digital libraries, such as the Networked Digital Library of Theses and Dissertations, with these bases a disambiguation process was carried out and a new database was generated. The system developed for this monograph has the objective of enabling the visualization of the trees and the correction of the disambiguation algorithms, where the users would edit the information. In the future, we want to make a series of analyzes about the genealogical trees and the interactions about them.

Keywords*:* *academic tree, web system, digital library*

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS

ACM SIGACT	Association for Computing Machinery Special Interest Group on Algorithms and Computation Theory
AFT	The Academic Family Tree
DCC	Departamento de Ciência da Computação
IHC	Interação Humano-Computador
LOCUS	Laboratório de Computação Social
MGP	The Mathematics Genealogy Project
NDLTD	Networked Digital Library of Theses and Dissertations
NT	Neuro Tree
OATD	Open Access Theses and Dissertations
PHDCS	PhD students in computer science
POCII	Projeto Orientado em Computação II
STP	Science Tree Project
UFABC	Universidade Federal do ABC
USP	Universidade de São Paulo

SUMÁRIO

1 INTRODUÇÃO

A humanidade tenta registrar sua história, transmitindo-as em livros didáticos ou literários, pinturas, músicas etc. Na nossa história, várias pessoas serão sempre lembradas, seja por ter matado milhões de pessoas, como Adolf Hitler, por ter sido o primeiro a realizar uma determinada tarefa, como Neil Armstrong ou por terem realizado uma grande contribuição à humanidade e à ciência, como Marie Curie e Alan Turing.

No projeto Science Tree nós queremos reconstruir parte da história da humanidade. Nosso foco será na história da ciência, sua evolução, como as diferentes áreas se interagem para criar uma nova e como surgem os novos cientistas. É um projeto audacioso e promissor.

Esse projeto pode ser dividido em três etapas: coleta e tratamento dos dados, correção e visualização dos dados, e análise das árvores. É um projeto envolvendo dois professores, Prof. Fabrício Benevenuto e o Prof. Alberto Laender, e dois alunos, Wellington Dores, aluno de mestrado e Elias Soares, aluno de graduação.

Contribuí ao projeto na primeira etapa, criando um dos *crawlers* de uma das bases de dados e na segunda etapa, criando sozinho, o sistema *web* para visualização e edição das árvores, cujo é o assunto dessa monografia de aspecto técnico. Pretendo desenvolver a terceira etapa durante o meu mestrado.

2 CONTEXTUALIZAÇÃO E TRABALHOS RELACIONADOS

Tentar recriar a história da ciência não é uma novidade. Há diversas tentativas e trabalhos ainda em andamento.

Dentre eles temos a NT, que tenta construir a árvore da neurociência, onde os usuários inserem as orientações. Os criadores da NT resolveram ampliá-la e criaram a AFT, um sistema dividido em diversas áreas da ciência, áreas mais amplas como física até algumas bem específicas como robótica e neuro-oncologia, na AFT também são os usuários que inserem as informações.

A NDLTD é uma extensa biblioteca digital de teses e dissertações, não tem o intuito de criar árvores genealógicas da ciência, mas é rica para ser usada como tal, pois tem cerca de 4.5 milhões de teses e dissertações. Atualmente ela não está no sistema pela complexidade de tratar e desambiguar os dados da mesma. A quantidade de dados incompletos ou incorretos é extremamente expressiva e futuramente ela será inserida no sistema. A OATD é basicamente igual a NDLTD, contendo cerca de dois milhões de teses e dissertações.

No Brasil temos outras iniciativas de construção de árvores genealógicas, um projeto conjunto da UFABC e da USP pretende desenvolver um sistema web para visualização de árvores, usando dados do Lattes e da Capes. Inicialmente estão considerando apenas pesquisadores da matemática da UFABC.

Abaixo seguem as fontes de dados atualmente usadas no sistema.

2.1 PHDCS

O projeto PHDCS foi iniciado pela revista ACM-SIGACT com os pesquisadores que publicaram sobre computação teórica. A revista entrou em contato com os pesquisadores pedindo para que eles informassem quem eram seus orientadores e seus alunos. Posteriormente, foi pedido para que os alunos e orientadores dos pesquisadores anteriores também preenchessem o mesmo formulário. No final, obteve-se uma árvore com 1025 nodos e 1043 arestas. É uma árvore de cientistas da computação e matemáticos.

2.2 MGP

O *Mathematics Genealogy Project* é um dos projetos mais antigos, criado em 1997, pelo *North Dakota State University's Department of Mathematics*, nos Estados Unidos, tem atualmente 205.322 nodos com pessoas desde o século XIV. Das fontes que utilizamos, essa é a melhor, ela foi feita à mão pelo o pessoal do departamento, tendo poucos erros e os dados são bem estruturados. Apesar de ser focado em matemática, ela consegue abranger grandes nomes da engenharia e ciência da computação.

2.3 PLATAFORMA LATTES

Como o NDLTD, não é o objetivo do Lattes ser usado para análise de árvores genealógicas, mas dele podemos retirar tais informações analisando cada currículo e olhando quem foi os orientadores e quem foram os orientados. Ele nos oferece algumas informações úteis como a área de atuação de

cada pesquisador e o histórico deles. O Lattes nos possibilita entender a história da ciência nacional. Temos cerca de 2 milhões de nodos do Lattes, sendo que no total ele tem aproximadamente 4.5 milhões. Entretanto é alto a quantidade de nomes preenchidos erroneamentes, o nome do professor pode ser Philip e o aluno coloca Felipe, por exemplo. Esse é o tipo de caso em que somente o sistema crowdsourcing poderá resolver.

3 DESENVOLVIMENTO DO TRABALHO

Como dito anteriormente, esse projeto pode ser dividido em três etapas. Sendo a primeira a coleta dos dados juntamente com o tratamento, o sistema web para visualizar e corrigir os dados, e análise dos grafos e árvores. As subseções abaixo se referem às duas primeiras etapas.

3.1 COLETA E TRATAMENTO DOS DADOS

Alguns dados foram coletados e outros cedidos. O mestrando Wellington coletou a NDLTD e a OATD, ambas usam um protocolo de comunicação chamado OAI. Embora ambas tenham uma grande quantidade de dados, há muita informação erroneamente preenchida e algumas vezes não tem a informação. Diante da complexidade que o tratamento desses dados exigem, decidimos não trabalhar com eles por enquanto.

A base MGP foi coletada por mim. Fiz um *crawler* usando uma biblioteca em Python chamada Beautiful Soup, que é um parser de HTML. Coletei os mais de 200 mil nodos do site do MGP.

Já base da PHDCS, nós fizemos o *download* do site de datasets *Pajek*. Nossa base do Lattes foi cedida por um aluno do professor Laender.

Temos aproximadamente 2 milhões de nodos no Lattes, mais de 200 mil na MGP e 1025 na PHDCS. Uma pessoa pode aparecer na mesma base mais de uma vez e às vezes pode estar com nomes ligeiramente diferentes, como Clodoveu Augusto Davis Jr e Clodoveu Augusto Davis Junior. Além disso, pode repetir em uma mesma base ou estar em várias bases diferentes. O algoritmo de desambiguação deve ser capaz de identificar a maioria desses casos.

Primeiramente foi necessário tratar os nomes já que, em alguns casos, aparecem pronomes de tratamento como: doutor, phd, professor ou mestre. Retiramos, também, todas as acentuações dos nomes e os colocamos em *lowercase*.

Somando todas as bases temos mais de 2.2 milhões de nomes, para termos um processo de comparação e desambiguação eficiente tivemos que realizar um processo de blocagem dos nomes. Ao separá-los em blocos, não precisaríamos comparar cada um deles, o que diminui muito a complexidade computacional. Usamos o Ngram de tamanho 4 para criar os blocos.

Com os nomes normalizados e divididos em blocos, começamos o processo de desambiguação. Primeiramente foi realizado a desambiguação entre os nomes de uma mesma base, para depois realizar das que eram diferentes entre si. O algoritmo de desambiguação usado é uma variação do comparação por fragmento (Oliveira, 2005).

Ao final da desambiguação, temos uma nova base que é menor do que a soma inicial de todas. O último processo resultou em 1.078.767 nodos e 1.220.361 arestas. Segue na Tabela 1 as informações da atual base que já se encontra no sistema.

Quantidade de relações	1.220.361
Quantidade de nodos	1.078.767
Quantidade de mestres	490.335
Quantidade de doutores	730.026

Tabela 1: Informações sobre a base desambígua.

3.3 O SISTEMA

Quando o Professor Fabrício pediu-me para desenvolver o sistema, eu não tinha nenhum conhecimento das tecnologias necessárias para desenvolvê-lo, tive que realizar uma pesquisa sobre elas e criar um cronograma de estudo e desenvolvimento, por isso na minha proposta de POCII eu foquei no cronograma de estudos das tecnologias, enquanto isso o sistema seria desenvolvido.

O algoritmo de desambiguação e nossos *crawlers* eram desenvolvidos Python, para manter um padrão, decidi que usaria como tecnologia de back-end algum *framework* em Python. Encontrei duas opções, Flask e Django, porém pela variedade de exemplos e extensa documentação encontrada, optei pelo Django. Encontrei um curso (Filho, 2016) na plataforma Udemy que me ajudou a entender bem a tecnologia.

Conforme eu aprendia Django e desenvolvia o sistema, senti a necessidade de aprender HTML5. Pausei os estudos de Django e comecei a estudar HTML5, para isso assisti diversas *playlists* no YouTube, além de recorrer ao W3Schools. Porém houve a necessidade de aprender novas tecnologias, já que o HTML não era suficiente para o que precisava, então iniciei os estudos de Bootstrap e CSS3, recorri novamente a vídeos no Youtube e ao W3Schools.

Com domínio intermediário de Django, HTML, CSS e Bootstrap, já era possível ter um sistema básico funcionando. Entretanto, um diferencial da nossa ferramenta são as visualizações das árvores. Comecei a estudar um *framework* Javascript chamado D3, porém, apesar de ser poderoso, o aprendizado dessa tecnologia exigia muito tempo, por isso comecei a estudar o VisJS e percebi que era satisfatório para o que precisava. Atualmente ele é o *framework* de visualização usado.

Temos o front-end em HTML5, Bootstrap, CSS e VisJS; o back-end em Django; porém para termos um sistema funcional falta o banco de dados. O Wellington gerou a nova “base de dados” em um arquivo CSV. Conforme a orientação do Laender e do Wellington, o sistema usaria um banco de dados não relacional orientado a grafo chamado Neo4j. Estudei ele usando sua documentação e

converti o CSV para as entidades do Neo4j. Segue um exemplo de relacionamento entre um professor e aluno na Figura 1.

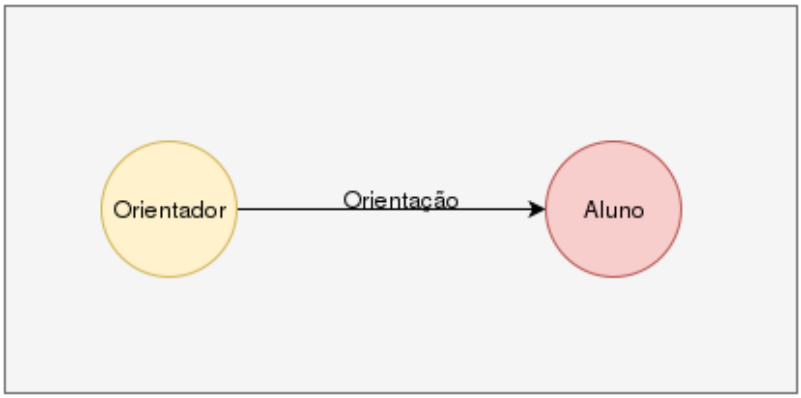

Figura 1: Relação no Neo4j.

Optamos por deixar a maioria das informações na aresta e nos nós apenas as informações referentes às pessoas. Um esquema de como os dados estão estruturados no Neo4j encontra-se na Tabela 2 e Tabela 3.

personID	full_name	first_name	middle_name	last_name	original_baseID	userID	email	webpage
ma69525	andrew chi chih yao	andrew	chi chih	yao	69525	Unknown	Unknown	Unknown

Tabela 2: Representação de um nó.

teacherID	**studentID**	type	university_name	title	year	research_area	research_sub_area
allnewma53269	**la9780438960315223**	Master	Unknown	Unknown	Unknown	Unknown	Unknown

Tabela 3: Representação de uma aresta.

O personID é o identificador criado na desambiguação, é único e é usado no relacionamento para identificar o professor e o aluno. O original_baseID é o identificador do elemento na base de origem, será útil se formos atualizar os dados futuramente, baixando-os novamente e verificando possíveis atualizações. O userID é o identificador criado na hora que o usuário cria uma conta no sistema, quando esse usuário buscar sua árvore pela primeira vez ele poderá anexá-la ao seu perfil e o identificador criado pelo sistema será salvo.

Na primeira vez que o Neo4j foi usado no sistema, eu usei expressão regular para buscar pelos nomes. Como são mais de 1 milhão de nomes, a busca demorava cerca de 10 segundos, o que não era ideal. Então, fiz um algoritmo básico que dividia o nome em três partes, indexei cada parte no Neo4j para agilizar a busca e assim elas passaram a ser quase que instantâneas.

A aresta além de armazenar informações básicas sobre a orientação, irá guardar a área e a subárea que a dissertação ou teses pertence. Isso ajudará na terceira etapa, analisando como as áreas se interagem e criam novas áreas.

4 RESULTADOS E DISCUSSÃO

Uma abstração do Science Tree System pode ser visualizada na Figura 2.

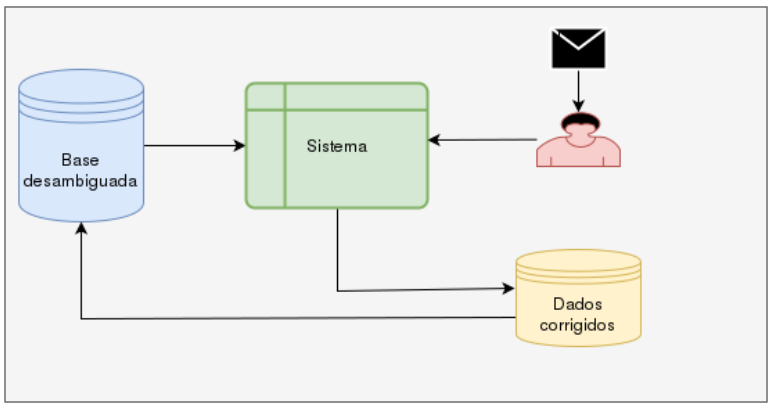

Figura 2: Funcionamento do Sistema.

Temos uma base de dados desambígua, o Neo4j, onde o sistema permitirá que o usuário visualize e edite as informações que ele considere incorretas; essas informações irão para um banco de dados secundário, em SQLite3; serão verificadas, de forma automática ou não, ainda a ser decidido; e então, verificado a veracidade da atualização, ela será inserida na base desambígua. É um ciclo que nos permitirá verificar a qualidade do algoritmo de desambiguação. Não podemos ir para a terceira etapa, a de análises de dados, se eles estiverem incoerentes.

Usuários não surgem instantemente. Portanto, inicialmente enviaremos e-mails à alguns pesquisadores chaves, seja para os que pesquisam sobre o assunto ou para àqueles que poderão de alguma forma melhorar a nossa base. Usaremos uma abordagem semelhante a da PHDCS, pediremos para que eles adicionem ou corrijam informações sobre seus orientadores e alunos.

Futuramente, o sistema será acessado pela url: http://sciencetree.net, mas para a avaliação desse trabalho, eu hospedei o sistema em uma máquina no laboratório LOCUS, que pode ser acessado

pelo endereço: http://150.164.10.208:1989/. A máquina ficará no ar até o final da disciplina. Vale ressaltar que, caso tente acessá-lo e esteja fora do ar, provavelmente foi devido a uma queda de energia.

Abaixo segue alguns prints tirados do sistema.

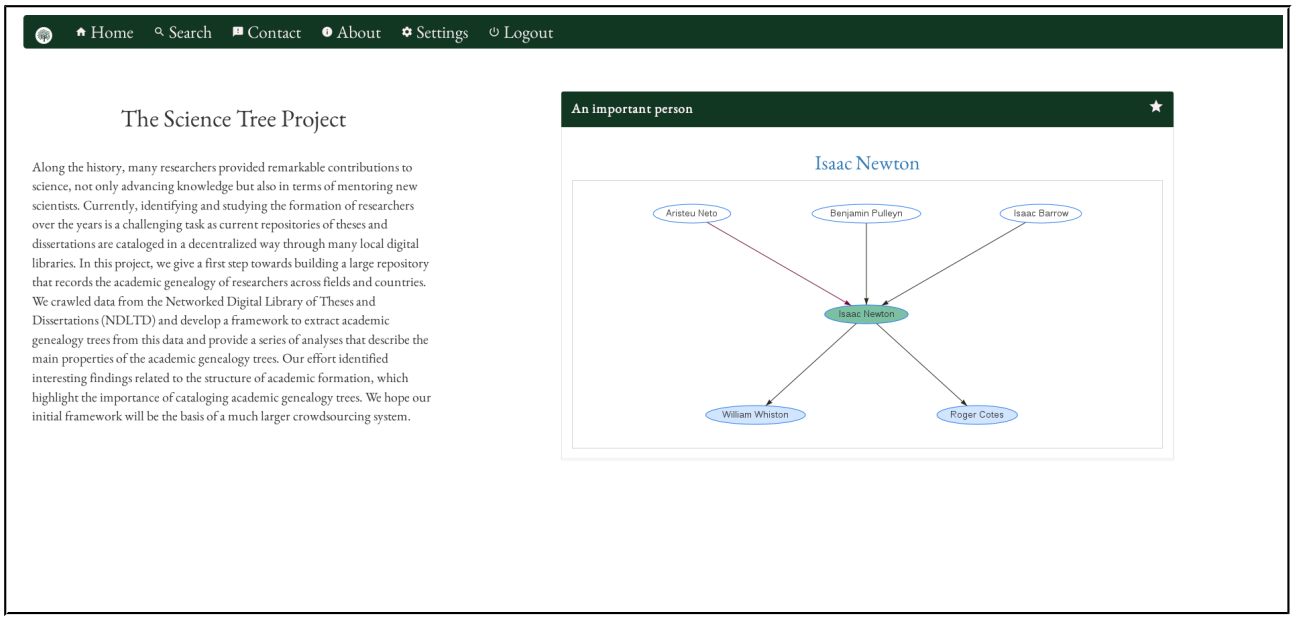

Figura 3: A Home do Science Tree

Escolhi entidades da ciência que tiveram alguma importância para a mesma, nomeio elas como “pessoas importantes”. Toda vez que o usuário acessar a *home*, uma dessas pessoas será aleatoriamente escolhida dentro de uma lista com várias outras “pessoas importantes”. Vale ressaltar que tais “pessoas importantes” podem ser ganhadores de Nobel, ou do prêmio Alan Turing, ou cientistas como Isaac Newton. A quantidade de “pessoas importantes” é 37.

Para buscar por alguém é bem simples, basta ir em *Search* e pesquisar pela a pessoa. Abaixo vemos uma busca por clodoveu.

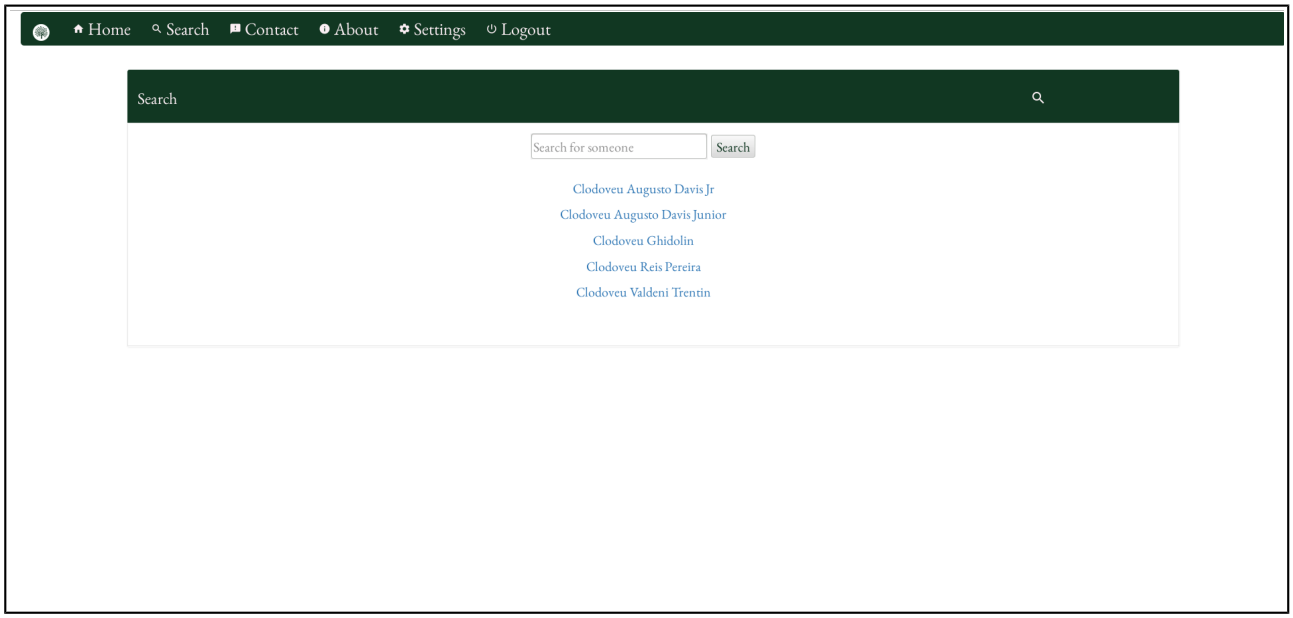

Figura 4: Resultado da busca por clodoveu

Nota-se que foram dois resultados para uma mesma pessoa, Clodoveu Augusto Davis Jr e Clodoveu Augusto Davis Junior. O algoritmo de desambiguação não conseguiu detectar que eram a mesma pessoa. Futuramente o professor Clodoveu Augusto Davis Junior será convidado a editar suas informações.

Ao clicar em um dos nomes dos resultados, temos acesso ao perfil da pessoa, como na Figura 5.

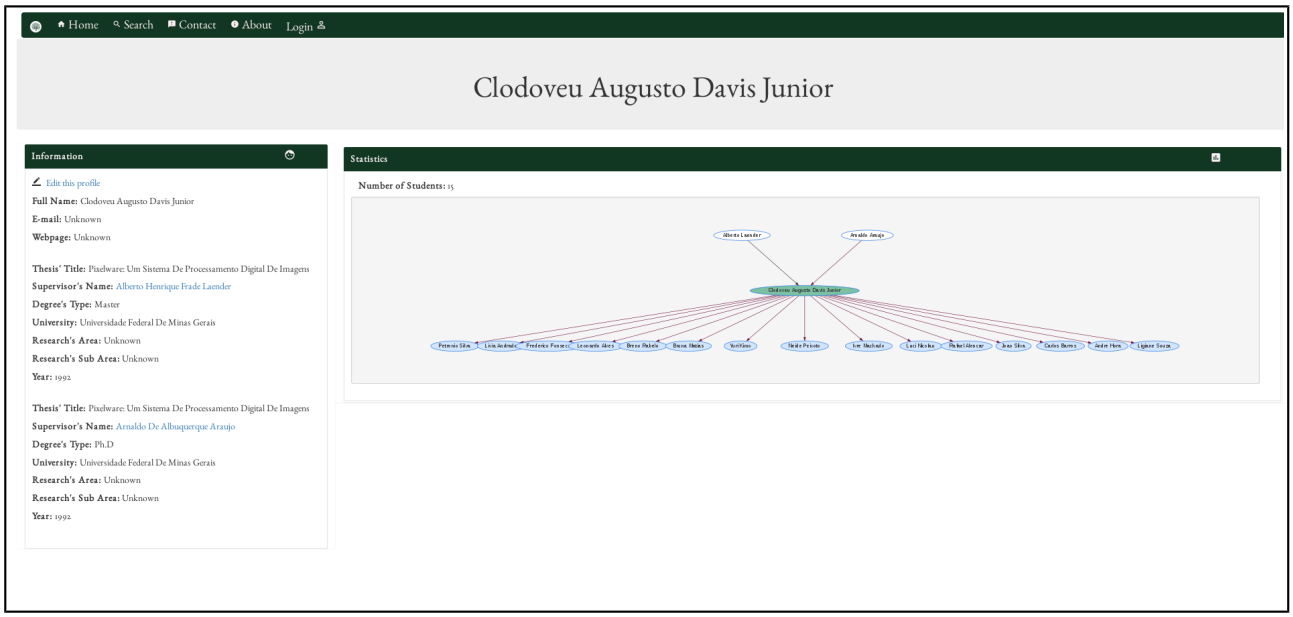

Figura 5: Perfil do Professor Clodoveu Augusto Davis Junior

No painel esquerdo podemos editar o perfil e temos as informações pessoais do perfil, juntamente com suas informações acadêmicas, como títulos das teses e dissertações, nome e link para os perfis dos orientadores e o ano que o título foi obtido. A implementação do sistema foi crucial para o avanço do projeto. Visualizando as árvores detectamos diversos erros do algoritmo de desambiguação. Algumas vezes ele mistura as informações de mestrado com as de doutorado, como exemplo o caso do professor Clodoveu (vide Figura 5). Esse erro já está sendo corrigido.

Além de possibilitar interação com a visualização, o VisJS fornece muitas configurações. Na Figura 6 vemos mais de perto a árvore do professor Clodoveu.

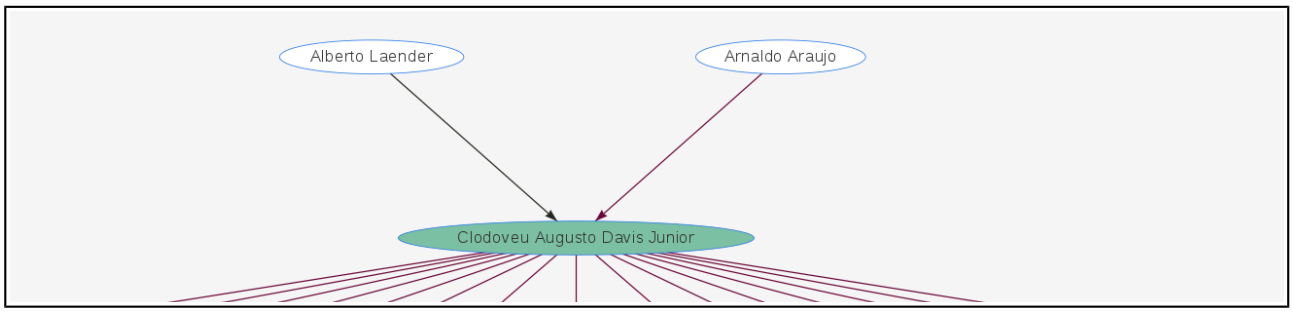

Figura 6: Zoom na árvore do professor Clodoveu.

Temos uma árvore hierárquica, os nodos superiores e de cor branca são dos orientadores, deles saem arestas apontando para o nó do perfil que está de cor verde, a aresta preta indica que àquela orientação é de mestrado e a roxa indica que é de doutorado. Do nodo verde saem arestas para todos os alunos daquele nó, que são os nodos azuis. Atualmente a árvore tem apenas três níveis, mas futuramente será possível navegar em mais níveis da árvore de uma pessoa. É algo que deverá ser limitado, pois um *browser* em um computador comum não suportaria receber toda a nossa base de uma vez.

Ao clicar em *Edit this profile* a página de edição será aberta. Entretanto, para fazer uma edição, é obrigatório que o usuário tenha uma conta. Assim, se a pessoa não estiver logada, antes de abrir a página de edição será aberta a página de login e criação de conta, apenas depois de fazer o login ou de criar a conta, que a pessoa será direcionada para a edição. É um mecanismo de segurança, queremos saber quem edita nossos dados, para termos um controle sobre o mesmo. Na Figura 7 vemos a página de edição do professor Clodoveu.

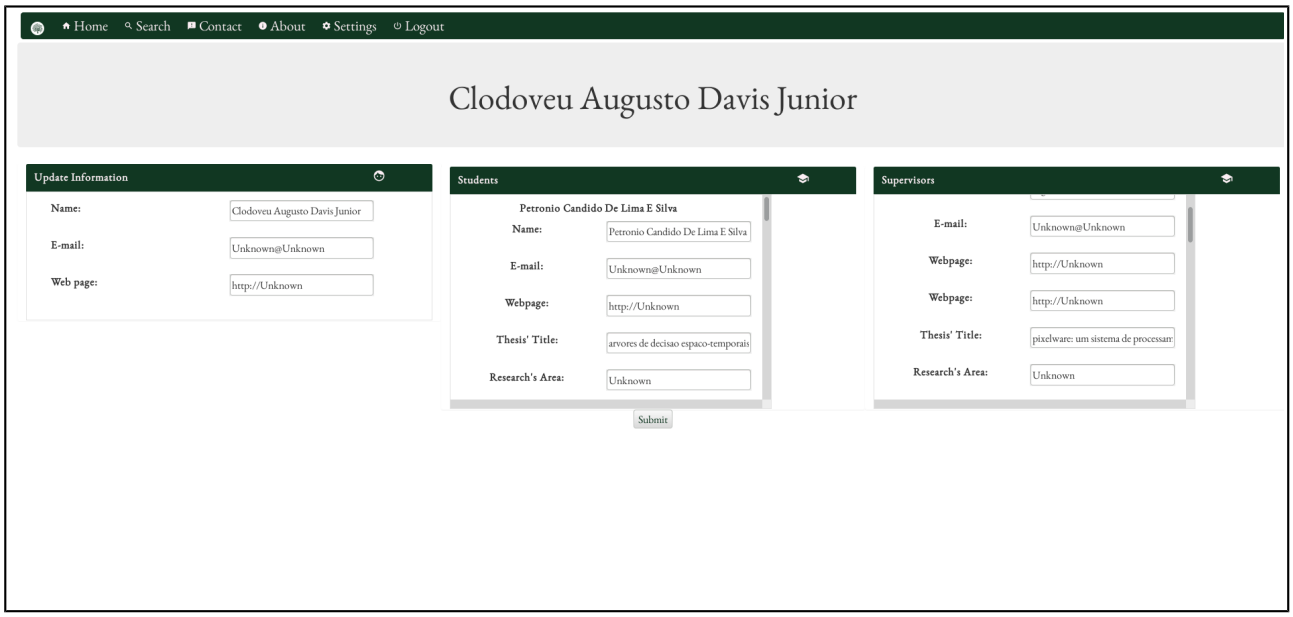

Figura 7: Edição do perfil do Professor Clodoveu.

Atualmente não é possível salvar as alterações, devido a uma exceção no Python/Django. Talvez seja pela a grande quantidade de informação que é enviada ao servidor quando se edita um perfil. Ainda estou investigando o problema mais profundamente e em breve consertarei esse bug. Nos campos de e-mail e webpage tive que formatar àqueles que eram desconhecidos para que passassem na verificação dos dados. Quando o usuário vai inserir uma webpage ou e-mail, verifico se são válidos.

Em *About* temos um texto falando do projeto, atualmente é o mesmo texto que se encontra na home. Com o intuito de que o usuário possa entrar em contato conosco, criamos a área *Contact* que permite que o usuário possa deixar suas informações e sua mensagem. Após logar, o usuário terá acesso a *Settings* que permite a ele fazer alterações em sua conta. A Figura 8 mostra o painel administrativo.

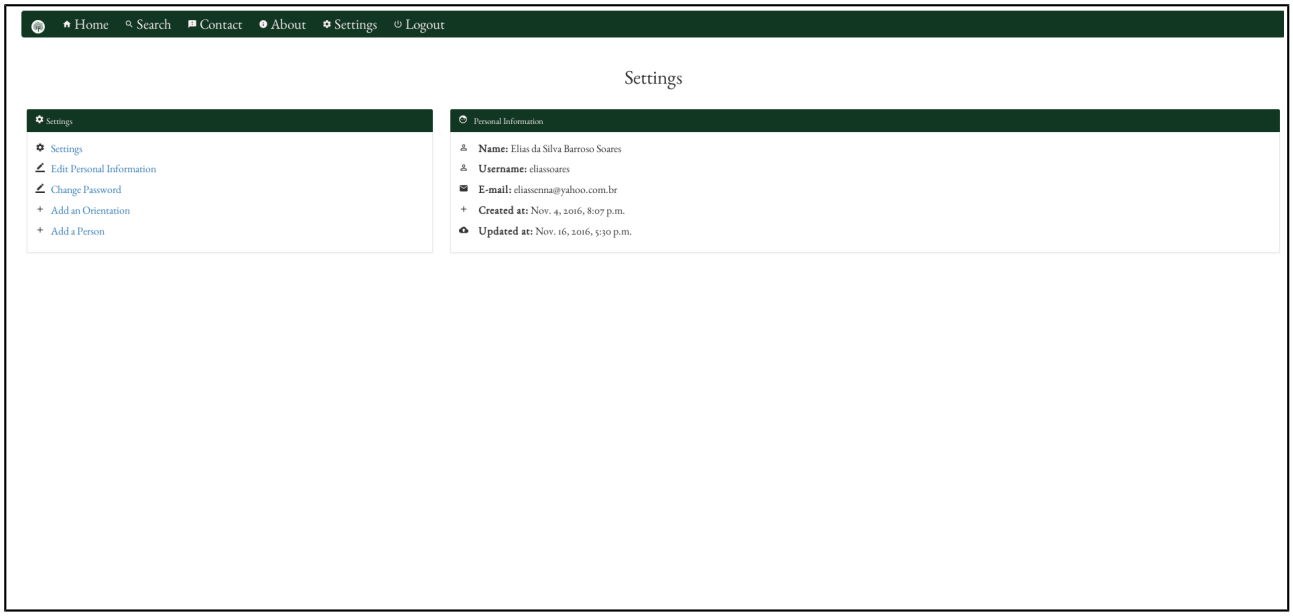

Figura 8: Settings.

Além de poder editar suas informações, nesse painel o usuário pode adicionar novos nodos e arestas. Todos os dados dos usuários, assim como suas alterações realizadas no sistema são armazenadas no banco de dados intermediário em SQLite3, como mostra a Figura 2. Isso nos dá mais flexibilidade, deixando o sistema desacoplado do banco de dados Neo4J. Como por exemplo, eles rodam em máquinas diferentes, o Neo4j é executado em um servidor mais potente, com 16 cores e 48 Gb de memória RAM, onde consome quase 10 Gb de memória RAM. Já o sistema é executado em um servidor QuadCore com 16 Gb de memória RAM, onde gasta menos 100 Mb para ser executado. Isso sem ter muitos acessos. Ainda não sabemos como ele se comportará quando se popularizar.

5 CONCLUSÕES E TRABALHO FUTUROS

Ainda há muito o que ser realizado e melhorado. Nas próximas semanas implementarei:

- Sistema de recuperação de senha. Ainda não é possível recuperar uma senha esquecida.
- Consertar um bug ao salvar as mudanças em um perfil. Após preencher as mudanças de um perfil e tentar salvá-las, acontece uma exceção no Django.
- Permitir anexar perfil. Não é possível que o usuário selecione algum perfil em nossa base como seu.
- Interação com redes sociais. Queremos que ao criar uma conta, o usuário possa logar-se usando o Facebook ou Google, podendo compartilhar sua árvore nas redes.
- Edição da árvore na própria visualização. Além de possibilitar a edição de um perfil via formulário, os usuários poderão adicionar novos nós e arestas na própria árvore.

Além das melhorias no sistema, nossa base de dados tem que ser melhorada assim como o algoritmo de desambiguação. O mestrando Wellington trabalha arduamente para que isso ocorra. Também devo decidir, juntamente com os professores, como os dados que o usuário corrigir serão colocados em nossa base no Neo4J. Assim como a alteração de algumas informações erradas, o usuário poderá adicionar novos nós e novas arestas, tais alterações poderão ser duplicadas se comparadas com a base do Neo4J, a inserção desses dados no Neo4J deverá ser realizada com determinado cuidado.

Uma vez que percebermos que a base está com um determinado nível de qualidade, nos baseando nas correções dos usuários, poderemos começar a analisar os dados e tentar realmente entender a história da ciência.

Criar um sistema desde o início para alguém que nunca o fez requer muito estudo. É um paradigma de programação diferente, passa-se a lidar com problemas diferentes e são muitas tecnologias envolvidas. Não há muitas disciplinas do DCC que abordam tais assuntos. Mas o DCC me forneceu algo mais valioso, a base e os conceitos fundamentais de computação. Com eles, eu fui capaz de buscar, sozinho, as informações necessárias para aprender, mesmo que no nível intermediário, as tecnologias que necessitei para o desenvolvimento do sistema. O maior aprendizado que tive no DCC é que com esforço e disciplina eu consigo aprender qualquer coisa.

Peço licença para terminar essa monografia de uma forma diferente dos demais. Gostaria citar uma parte do primeiro texto que li como aluno da UFMG. Um trecho do Guia acadêmico da UFMG (2014, p. 8):

> Por fim, não se esqueça de que a formação de qualidade que você irá receber nesta Instituição – a
> melhor que o País pode oferecer – será custeada pelo povo brasileiro.
> Portanto, você terá para com ele uma dívida de toda a vida. Não deixe, pois, de exercitar, desde
> agora, atitudes de solidariedade como forma de retribuir, em parte, o que, como estudante da
> UFMG, você está recebendo da sociedade.

À todos que batalham suas lutas diárias, pagando altos impostos, que nunca terão a real oportunidade de entrar nessa universidade, meus sinceros agradecimentos. Um dia, essa universidade será de todos.

6 REFERÊNCIAS

GUIA ACADÊMICO UFMG. Disponível em: <https://www.ufmg.br/online/arquivos/anexos/guia-academico.pdf>. Acesso em: 01 dez. 2016. p. 8.

OLIVEIRA, Jean W. A.; LAENDER, Alberto H. F.; GONÇALVES, Marcos André. Remoção de Ambigüidades na Identificação de Autoria de Objetos Bibliográficos. SBBD. 2005.

NETWORKED DIGITAL LIBRARY OF THESES AND DISSERTATIONS. Disponível em: <http://www.ndltd.org/> . Acesso em: 29 nov. 2016.

MATHEMATICS GENEALOGY PROJECT. Disponível em: <https://genealogy.math.ndsu.nodak.edu/index.php>. Acesso em: 29 nov. 2016.

NEUROTREE PROJECT. Disponível em: <http://neurotree.org>. Acesso em: 30 nov. 2016.

THE ACADEMIC FAMILY TREE PROJECT. Disponível em: <http://academictree.org/>. Acesso em: 30 nov. 2016.

OADT. Disponível em: <https://oatd.org/>. Acesso em: 30 nov. 2016.

GENEALOGIA UFABC. Disponível em: <http://hostel.ufabc.edu.br/~daniel.miranda/genealogia/>. Acesso em: 30 nov. 2016.

JOHNSON, D. S. The Genealogy of Theoretical Computer Science, SIGACT News, Vol. 16, No. 2, p. 36-44, 1984. Reprinted in Bulletin of the EATCS, No. 25, p. 198-211, 1985.

NOOY, W. de; MRVAR, A.; BATAGEL V.. Exploratory Social Network Analysis with Pajek. Cambridge University Press, c. 11. 2004.

PAJEK DATASETS. PhD students in computer science. Disponível em: <http://vlado.fmf.uni-lj.si/pub/networks/data./esna/CSPhD.htm> . Acesso em: 29 nov. 2016.

DORES, Wellington; BENEVENUTO, Fabrício; LAENDER, Alberto Laender. Extracting Academic Genealogy Trees from the Networked Digital Library of Theses and Dissertations. In Proceedings of the 16th ACM/IEEE-CS Joint Conference on Digital Libraries. Newark, USA. 2016.

FILHO, Gileno Alves Santa Cruz. Python 3 na Web com Django (Básico e Intermediário). Disponível em: <https://www.udemy.com/python-3-na-web-com-django-basico-intermediario>. Acesso em: 01 dez. 2016.