

1 **Análise Subjetiva**

Esta seção é dedicada à comentários pessoais e acadêmicos relacionados ao trabalho realizado durante o ano e o curso de Ciência da Computação do IME-USP.

1.1 **A procura de um orientador**

Ao fazer a matrícula em MAC0499, nos deparamos com a necessidade de escolher um orientador que irá auxiliá-lo a realizar o trabalho ao longo do ano. Como foi a escolha do meu orientador?

Eu não tinha certeza do que poderia fazer como TCC, então conversei com professores de várias áreas que pudessem me interessar, como visão computacional, interação humano-computador, banco de dados, e acabei escolhendo uma área mais teórica: grafos.

MAC0328 foi lecionada pelo professor Coelho, para a turma de ingressos em 2009. Suas aulas eram ótimas, a matéria era interessante, o professor parecia muito aberto a qualquer tipo de diálogo. Por coincidência, um amigo meu que havia feito TCC no ano anterior iria tê-lo como orientador novamente, aí surgiu a ideia do trabalho em dupla. Nós já tínhamos feito um trabalho em dupla anteriormente, em MAC0331, e podemos dizer que foi um bom trabalho, embora tivéssemos ido para a Rec. Mas a Cris mandou todo mundo pra rec por causa do projeto, então não encanamos muito com isso.

Feito. Tinha um orientador legal, fazendo dupla de trabalho com um amigo próximo, com um tema legal (a princípio, estudar o PageRank e aplicar ele sobre os sites do IME), tinha tudo pra ser bom.

1.2 **O estágio e o primeiro semestre de 2012**

Durante o primeiro semestre de 2012, eu cursei quatro disciplinas e fiz estágio. Uma obrigatória do período ideal, duas trancadas do quinto semestre e uma optativa. Devido aos horários, eu estagiava de segunda, quarta e sexta. Era um estágio na área de desenvolvimento Java/Web. Por causa desse horário um pouco louco, por estar cursando MAC0315, e por não saber administrar muito bem meu tempo, o TCC ficou um quase parado durante o primeiro semestre.

Eu estudava algumas coisas necessária para a compreensão do trabalho. Mas botar a mão na massa mesmo veio só no segundo semestre. Muitas coisas eu aprendi com a minha dupla também. Nós pegamos o tema do TCC dele do ano anterior.

1.3 **O objetivo inicial e suas mudanças**

Ao longo do trabalho, nos perguntávamos qual era o nosso objetivo. Inicialmente, tínhamos que coletar as páginas com o domínio www.ime.usp.br e usar o método PageRank para calcular qual as páginas mais relevantes do instituto. A coleta das páginas seria

feita por meio de um crawler, um robô que navegaria pelo site www.ime.usp.br e sairia coletando as urls.

Logo nos deparamos com a grande dificuldade de manipular páginas html. A falta de padronização do código-fonte das páginas faziam com que tivéssemos mais exceções a tratar do que casos normais de referências. Isso nos desanimou muito. A cada exceção tratada, descobríamos umas 3 novas. Aos poucos, parecíamos abandonar a ideia de implementar um crawler para o trabalho. E o que seria dele? Pensamos então num estudo comparativo, talvez entre o PageRank e mais algum método importante que seja próximo dele.

Numa conversa com o nosso orientador, ele nos mostrou um livro que nos serviu muito bem como referência para esse trabalho, inclusive mostrando alguns outros métodos de classificação para buscas na web. Fomos caminhando a realizar um estudo comparado entre os métodos PageRank e HITS, embora nunca em nenhum momento alguém tivesse dito que o objetivo era esse. Ele simplesmente foi aparecendo e se mostrou mais viável do que implementar o crawler.

Ao longo do trabalho foi visível que o nosso orientador não ficou muito contente com a nossa gradativa mudança de objetivo.

1.4 A quantidade e a qualidade do TCC

Essa de objetivo nos tomou um pouco de tempo. Aproximando-se do final de outubro decidi largar o estágio e focar no TCC. Percebi que desenvolvimento não é a minha área, e que prefiro muito mais aprender algo relacionado ao núcleo de computação teórica do que algo do núcleo de sistemas. O tempo foi corrido, fizemos muito do trabalho nessas 6, 7 últimas semanas que nos restou. Acredito que fizemos uma apresentação boa, pelo menos em qualidade. Acredito que conseguimos realizar um estudo diferente, com uma boa qualidade e completo. O receio é o quanto. Será que é trabalho suficiente? Será que falta mais coisas?

Corremos um pouco com o trabalho por causa da necessidade do Caio entrar no mestrado. Nos planos dele, o mestrado começaria logo no primeiro semestre de 2013.

1.5 Lista das disciplinas cursadas que foram importantes para o trabalho

- MAC0110 - MAC0122 - MAC0323
Disciplinas de formação. Completamente essenciais para o nosso TCC. Utilizamos nossa linguagem de maior domínio, C, onde tive contato em MAC0122 e MAC0323.
- MAT0139 - Álgebra Linear para Computação
Esta disciplina foi a base para o conteúdo de autovalores e autovetores que utilizamos no trabalho. A necessidade de visão da modelagem e manipulação das estruturas utilizadas fizeram a importância desta disciplina.

- MAE0228 - Introdução à Probabilidade e Processos Estocásticos
Todo o método do PageRank é baseado na modelagem de uma matriz de adjacência da web até que ela vire primitiva para admitir uma distribuição estacionária de probabilidade. Isto é o coração do PageRank. A base para tudo isso nós obtivemos nessa disciplina.
- MAC0300 - Métodos Numéricos da Álgebra Linear
Para calcular o valor atribuído à cada página, tanto no PageRank como no HITS, foi necessário utilizar um algoritmo convergente que resulta no autovetor associado ao autovalor dominante de uma matriz. Isto já faz parte da parte de implementação do método. O método da potência, visto nessa disciplina, foi o algoritmo convergente utilizado.
- MAC0328 - Algoritmos em Grafos
A modelagem começa com um grafo. Tudo começa com o grafo da web. É ele a base que temos para moldar ambos os métodos de classificação de conteúdo que estudamos.

1.6 Agradecimentos

Agradeço à minha família, que sempre me proporcionou o melhor para que eu tivesse oportunidade de estudar numa universidade de excelência e sempre me apoiou na escolha do curso.

Para a realização deste trabalho, agradeço ao meu orientador e professor José Coelho de Pina, que teve muita paciência conosco e nos ofereceu muito do tempo dele para sanar dúvidas ou criar outras novas. Ele é o responsável por tudo o que temos hoje.

Agradeço também ao meu colega de trabalho Caio, com o qual dividi muitas alegrias e frustrações ao longo desse tempo de trabalho.