

Laboratório Nacional de  
Computação Científica

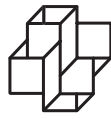
# PIBIC PIBITI

## XII Jornada de iniciação científica e tecnológica

Livro de Resumos

07 de Agosto de 2017  
Av. Getúlio Vargas 333  
Petrópolis, RJ.





Laboratório  
Nacional de  
Computação  
Científica

XII Jornada de iniciação  
científica e tecnológica

# Livro de Resumos

Local de realização  
Hall do Auditório do  
Laboratório Nacional de Computação Científica

Endereço  
Av. Getúlio Vargas 333 - CEP 25651-075  
Tel.: (24) 2233-6150  
[www.lncc.br](http://www.lncc.br)

Quitandinha-Petrópolis, RJ  
07 de Agosto de 2017

# **Jornada de Iniciação Científica e Tecnológica do LNCC**

Petrópolis, 07 de agosto de 20176.

## **Laboratório Nacional de Computação Científica – LNCC**

Diretor  
Augusto Cesar Gadelha

Coordenação de Gestão e Administração - COGEA  
Anmily Paula dos Santos Martins

Coordenação de Métodos Matemáticos e Computacionais - COMAC  
Frédéric Gerard Christian Valentin

Coordenação de Modelagem Computacional - COMOD  
Márcio Arab Murad

Coordenação de Pós-Graduação e Aperfeiçoamento - COPGA  
Abimael Fernando Dourado Loula

Coordenação de Tecnologia da Informação e Comunicação - COTIC  
Wagner Vieira Léo

Programa Institucional de Bolsas de Iniciação Científica &  
Programa Institucional de Bolsas de Iniciação em Desenvolvimento Tecnológico e Inovação  
Regina Célia Cerqueira de Almeida

## **Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq**

Presidente  
**Glaucius Oliva**

Coordenadora Geral do PIBIC/PIBITI  
**Lucimar Batista de Almeida**

# **Jornada de Iniciação Científica e Tecnológica do LNCC**

## **Comissão Interna do PIBIC/PIBITI-LNCC**

Regina Célia Cerqueira de Almeida  
Eduardo Lucio Mendes Garcia  
Helio José Correia Barbosa  
Jack Baczynski

## **Avaliadores Externos**

José Luis Drummond Alves - COPPE/UFRJ  
Priscila Vanessa Zabala Capriles Goliat - UFJF  
Regina Maria Maciel Braga Villela - UFJF

# Apresentação

O LNCC realiza este ano a XII Edição da Jornada de Iniciação Científica e Tecnológica, que é um fórum de divulgação das pesquisas desenvolvidas no contexto dos Programas Institucionais de Bolsas de Iniciação Científica (PIBIC) e de Bolsas de Iniciação Tecnológica (PIBITI) fomentados pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). No período de agosto de 2016 a julho de 2017, o PIBIC e PIBITI congregaram alunos de várias instituições de ensino e de diversas áreas do conhecimento. Este volume apresenta os resumos dos trabalhos desenvolvidos pelos bolsistas no período. Durante a Jornada, os trabalhos são apresentados pelos bolsistas oralmente e em sessão de pôsteres e avaliados por um comitê científico externo e multidisciplinar.

Nesta XII Edição da Jornada, o Comitê Externo de Avaliação do PIBIC/PIBITI tem a seguinte composição:

Prof. José Luis Drummond Alves - COPPE/UFRJ

Prof<sup>ª</sup>. Regina Maria Maciel Braga Villela - UFJF

Prof<sup>ª</sup>. Priscila Vanessa Zabala Capriles Goliat - UFJF

Destacamos o papel relevante do PIBIC/PIBITI do LNCC no desenvolvimento das pesquisas no LNCC e, principalmente, na formação complementar dos bolsistas, promovendo o aprimoramento do conhecimento, espírito criativo, reflexão crítica e ética. Estas características têm contribuído para suas inserções no mercado de trabalho e em programas de pós-graduação, como o PPG em Modelagem Computacional do LNCC. Este é o resultado do esforço e dedicação de todos os participantes.

# Agradecimentos

Agradecemos ao CNPq pelas bolsas concedidas, à Direção do LNCC pelo apoio e à Comissão Interna do PIBIC e PIBITI no LNCC.

Agradecemos a disponibilidade e contribuição dos membros do Comitê Externo de Avaliação.

O sucesso desta Jornada, e do Programa como um todo, é o resultado da dedicação e do esforço de toda a comunidade do LNCC. Expressamos em particular nosso reconhecimento ao apoio concedido pela secretaria do PPG-LNCC e, em particular, à Sra. Roberta Machado.

Regina Célia Cerqueira de Almeida  
*Coordenadora do PIBIC/PIBITI - LNCC*

## Índice

Construção de bancos de ligantes e alvos moleculares para o Portal de Triagem Virtual DockThor- VS.....	01
Bolsista: Aldo Patrick Corrêa Orientador: Laurent Emmanuel Dardenne Coorientador: Isabella Alvim Guedes	
Compilação e formatação de informação médica relevante para a elaboração de um atlas 3D de Anatomia, com posterior teste do sistema na comunidade médica, nível médio e demais áreas da saúde.....	06
Bolsista: Anna Caroline Wayand Martins Orientador: Jauvane Cavalcanti	
Modelagem geométrica e otimização de objeto 3D para ambientes virtuais colaborativos.....	11
Bolsista: Arthur Schulze Bittar Orientador: Jauvane C. de Oliveira	
Modelagem da Malha Aérea Doméstica Brasileira com Grafos Multi-Aspecto.....	19
Bolsista: Bernardo Botelho Orientador: Artur Ziviani	
Controle ótimo para sistemas Elásticos .....	23
Bolsista: Debora de Oliveira Orientador: Jaime Rivera	
Derivada Topológica na Mecânica da Fratura.....	24
Bolsista: Filipe Alves da Silva de Moura Orientador: Antonio André Novotny Coorientador: Marcel Xavier	
Avaliação de sistemas de computação distribuída de alto desempenho: um estudo sobre estratégias para monitoramento de desempenho.....	28
Bolsista: Gabrieli Dutra Silva Orientador: Bruno Schulze	
Simulação numérica e computacional de tráfego viário.....	34
Bolsista: Guilherme Villar Piccinini de Magalhães Orientador: Elson Toledo Coorientador: Regina C.P. Leal Toledo	
Aplicações de Filtro de Kalman em Robótica.....	40
Bolsista: Hiago Riba Guedes Orientador: Marcelo Fragoso	
Métodos de Discretização para sistemas físicos dispersivos e/ou dissipativos.....	42
Bolsista: Igor Pereira dos Santos Pereira Orientador: Sandra Mara Cardoso Malta Coorientador: Cristiane O. Faria	

Explorando Sistemas de Gerência de <i>Workflows</i> Científicos para Experimentos de Bioinformática em Ambientes HPC.....	48
Bolsista: Joice Cristina G. A. Pedro	
Orientador: Kary Ann del Carmen Ocaña Gautherot	
Sensoriamento autônomo com inteligência baseada em nuvens computacionais.....	54
Bolsista: Jonathas Rodrigues Almeida	
Orientador Bruno Schulze	
Ambiente integrado para execução de workflows científicos.....	62
Bolsista: João Guilherme Rittmeyer	
Orientador: Fábio Porto	
Análise comparativa baseada em redes complexas dos efeitos da crise econômica na malha aérea doméstica brasileira.....	67
Bolsista: João Victor Marinho Bechara	
Orientador: Artur Ziviani	
Caracterização de texturas utilizando redes neurais e espaços fractais com aplicações a imagens radiológicas.....	72
Bolsista: Lucas Borsatto Simão	
Orientador: Gilson Antonio Giraldi	
Uma arquitetura baseada em banco de dados e visualizaçãono apoio à análise da biodiversidade marinha brasileira.....	75
Bolsista: Matheus Machado da Rosa Albuquerque	
Orientador: Luiz Manoel Rocha Gadelha Júnior	
Coorientador: Kary Ann del Carmen Ocaña Gautherot	
Simulação Computacional de Algoritmos Quânticos.....	81
Bolsista: Matheus Roberto Araújo Borsato	
Orientador: Renato Portugal	
Aplicações e técnicas de aprendizagem de máquinas à análise de dados de domínio e de tempo de execução de workflows de bioinformática.....	85
Bolsista: Matheus Tonelli de Souza	
Orientador: Luiz Manoel Rocha Gadelha Júnior	
Coorientador: Kary Ann del Carmen Ocaña Gautherot	
Controle ótimo de vigas de Euler-Bernoulli.....	91
Bolsista: Monique Ribeiro da Costa	
Orientador: Jaime Rivera	
Introdução á computação Quântica.....	94
Bolsista: Paulo Ricardo Borré Reis	
Orientador: Renato Portugal	
Análise de dados de atendimento da área de saúde com base em redes complexas.....	98
Bolsista: Rayssa Barreto Miranda da Rosa	
Orientador: Artur Ziviani	



Paralelização de análises sobre dados de simulação utilizando o paradigma MapReduce.....100  
Bolsista: Rodrigo Botelho  
Orientador: Fabio Porto

Investigação de polimorfismos de nucleotídeos únicos induzidos pela ação da Polimixina B e efeitos  
abióticos em *Klebsiella pneumoniae*.....102  
Bolsista: Thiago Cardoso Pereira Carneiro  
Orientador: Marisa Fabiana Nicolás

### **Bolsistas PIBITI**

Aplicação das ferramentas Intel Parallel Studio para modernização de código para métodos numéricos  
de diferenças finitas para solução de equações diferenciais parciais em arquitetura Intel  
Haswel/Broadwell.....108  
Bolsista: Gabriel Pinheiro da Costa  
Orientador: Carla Osthoff Barros  
Coorientador: Frederico Luís Cabral

Introdução aos métodos numéricos para solução de equações diferenciais parciais.....105  
Bolsista: Maria Eduarda Araújo  
Orientador: Orientador: Sandra Mara Cardoso Malta

Processamento e visualização de imagens de microscopia confocal.....118  
Bolsista: Matheus Alves de Moura  
Orientador: Gilson Antonio Giraldi

Estudo de estratégias auxiliares para predição de estruturas de proteínas baseada em moldes: Análise de  
preditores de atracamento molecular proteína-proteína.....121  
Bolsista: Matheus Miloski  
Orientador: Fábio Lima Custódio

RELATÓRIO DAS ATIVIDADES REALIZADAS NO PROGRAMA  
INSTITUCIONAL DE BOLSA DE INICIAÇÃO CIENTÍFICA

**Construção de Bancos de Ligantes e Alvos Moleculares  
para o Portal de Triagem Virtual DockThor-VS**

*Período 08/2016 – 07/2017*

Aldo Patrick A. Corrêa<sup>a,1</sup>, Isabella A. Guedes<sup>b,1</sup>, Laurent E. Dardenne<sup>c,1</sup>

<sup>a</sup> patrickredfield@gmail.com, (24) 22336179

<sup>b</sup> isabella@lncc.br, (24) 22336121

<sup>c</sup> dardenne@lncc.br, (24) 22336009, (24) 992072941

<sup>1</sup> Laboratório Nacional de Computação Científica, Avenida Getúlio Vargas,  
333, Quitandinha, Petrópolis - Rio de Janeiro, CEP 25651-075 – Brasil.  
[posgrad@lncc.br](mailto:posgrad@lncc.br), (24) 2233-6000

## 1. INTRODUÇÃO

A redução de custos e tempo no desenvolvimento de novos fármacos mostra-se indispensável em todos os aspectos. Nesta busca por novos candidatos a fármacos, as técnicas de triagem virtual em larga escala tem sido cada vez mais utilizadas, possibilitando a análise computacional de um grande número de moléculas e a seleção das mais promissoras<sup>1</sup>. O programa DockThor, desenvolvido pelo Grupo de Modelagem Molecular de Sistemas Biológicos (GMMSB),<sup>2-4</sup> tem obtido excelentes resultados em estudos de *docking molecular*. O sucesso neste desenvolvimento viabiliza o uso do programa DockThor em estudos de triagem virtual em larga escala, principalmente no portal de triagem virtual DockThor-VS (em desenvolvimento). O presente projeto deu segmento à construção de novos bancos de ligantes, a partir da pesquisa de conjuntos de compostos ativos relacionados a doenças negligenciadas, em especial às tripanossomíases.

O uso do termo "doenças negligenciadas" mostra-se comumente presente em relatos provenientes de países subdesenvolvidos e em desenvolvimento, referindo-se a quadros infecciosos que não possuem tratamento eficaz, financeiramente acessível e/ou de fácil administração. Os exemplos mais importantes desse grupo de doenças são Dengue, Leishmaniose, Malária, Doença de Chagas, Esquistossomose e Hanseníase. Todas estas doenças possuem expressiva prevalência no território brasileiro, atingindo uma parcela considerável da população. Mundialmente, estima-se que um sexto da população possua pelo menos uma destas doenças.<sup>5-6</sup>

A construção de novos bancos de ligantes amplia o potencial do método de triagem virtual como ferramenta na busca do melhor entendimento e aplicação de moléculas no desenvolvimento de novos medicamentos e no reposicionamento de fármacos já consolidados. O conceito de "reposicionamento de fármacos", descrito por Ashburn e Thor, resume-se na identificação e desenvolvimento de novos usos para fármacos já existentes, ou seja, seu uso além da indicação médica original<sup>7</sup>. Esta aplicação reduz expressivamente o custo e o tempo de desenvolvimento e aplicabilidade dos fármacos. A triagem virtual de bibliotecas de reposicionamento de fármacos encontra-se entre as abordagens mais utilizadas na busca destas novas aplicações de fármacos, ganhando destaque não apenas em setores acadêmicos, como também na indústria farmacêutica.<sup>7-8</sup>

A construção de novos bancos de ligantes envolve a pesquisa de possíveis compostos a serem explorados, desenho, otimização e parametrização das suas estruturas 3D, sendo uma etapa fundamental no desenvolvimento do portal de triagem virtual em larga escala

DockThor-VS.

## 2. OBJETIVOS

Construir os bancos de ligantes com atividade conhecida experimentalmente contra organismos causadores de doenças negligenciadas.

## 3. METODOLOGIA

O trabalho envolveu a análise de 4 bancos de compostos com dados de atividade obtidos experimentalmente no combate aos agentes etiológicos causadores de tripanossomíases, *T. brucei* e *T. cruzi*. Os bancos foram disponibilizados pelo DNDi - *Drugs for Neglected Diseases Initiative* (<https://www.dndi.org/>), através de publicações realizadas no banco de dados da ChEMBL-NDT<sup>9-11</sup> (Neglected Tropical Diseases) (<https://www.ebi.ac.uk/chemblntd/download>), totalizando 2072 compostos.

Nome do Conjunto	Número de Compostos
DNDi Chagas Dataset	751
DNDi <i>T.b. brucei</i> Dataset	1149
DNDi <i>T. cruzi</i> Dataset	67
DNDi Dataset: Antiprotozoal activity profiling of approved drugs	105

Os compostos disponibilizados nos bancos possuem deficiência de informação estrutural acurada, como por exemplo ausência de coordenadas tridimensionais e diferentes estados de protonação em diferentes faixas de pH. Desta forma, primeiramente foram desenhadas as estruturas 2D dos compostos com o programa MarvinSketch (<https://www.chemaxon.com/products/marvin/marvinsketch/>). Isômeros, estados de protonação e tautômeros dos ligantes estão sendo preditos com a ferramenta Ligprep/Epik da suíte de aplicativos Maestro<sup>12</sup>, utilizando diferentes faixas de pH e a presença de metais. Todos os compostos, que anteriormente não possuíam estrutura tridimensional disponível, terão sua estrutura 3D otimizada por meio da minimização de energia no vácuo utilizando o campo de força OPLS2005.

## 4. RESULTADOS

Os bancos de ligantes abrangem estruturas já conhecidas e utilizadas no tratamento de doenças não relacionadas, assim como moléculas pouco pesquisadas, possuindo um total de 2072 estruturas (disponível em: <https://www.ebi.ac.uk/chemblntd/download>) que mostram

possível atividade antiprotozoária sob as espécies *T. cruzi* e *T. brucei*. Entretanto, cada um destes bancos possui informações estruturais com complexidade variada, impossibilitando uma catalogação padronizada de seus dados sem a intervenção realizada.

Em seguida, as estruturas tridimensionais serão geradas e otimizadas. As estruturas serão disponibilizadas em formatos amplamente utilizados, como .mol2 e .sdf (estrutura tridimensional), .SMILE (estrutura unidimensional) e como ilustração 2D. Os arquivos serão disponibilizados no portal de triagem virtual em larga escala DockThor-VS.

## 5. CONCLUSÃO

A construção de bancos de moléculas com estrutura tridimensional curada é de grande importância para estudos de triagem virtual em larga escala. Com este projeto, será disponibilizado para a comunidade científica um banco de compostos com estruturas tridimensionais validadas, característica essencial para o sucesso de experimentos de triagem virtual baseada em estrutura. A aplicação dos mesmos na pesquisa de tratamentos de doenças negligenciadas mostra-se um fator chave na busca pelo interesse de grandes empresas farmacêuticas, diminuindo custos e tempo de pesquisa, aumentando as chances de novas descobertas voltadas a tais doenças. Além disso, estes compostos possuem atividade obtida experimentalmente contra organismos importantes em doenças negligenciadas, incluindo compostos já comercializados e que podem estar envolvidos em outras aplicações através do reposicionamento de fármacos.

## 6. REFERÊNCIAS

- Grey, J. L. & Thompson, D. H. Challenges and opportunities for new protein crystallization strategies in structure-based drug design. *Expert Opin. Drug Discov.* 5, 1039–1045 (2010).
- Magalhães, C. S. de. Algoritmos genéticos para o problema de Docking proteína-ligante. (2006).
- Magalhães, C. S., Barbosa, H. J. C. & Dardenne, L. E. in *Genetic and Evolutionary Computation – GECCO 2004* (ed. Deb, K.) 3102, 368–379 (Springer Berlin Heidelberg, 2004).
- Magalhães, C. S., Almeida, D. M., Barbosa, H. J. C. & Dardenne, L. E. A dynamic niching genetic algorithm strategy for docking highly flexible ligands. *Inf. Sci.* 289, 206–224 (2014).
- Yamey, G. The world's most neglected diseases: Ignored by the pharmaceutical industry and by public-private partnerships. *BJM* 325,176-177 (2002)
- Decit - Departamento de Ciência e Tecnologia do Ministério da Saúde. Doenças negligenciadas: estratégias do Ministério da Saúde. *Rev. Saúde Pública* 44(1),200-202 (2010).
- Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov.* 2004;3:673-683. doi: 10.1038/nrd1468

Bajorath, J. Integration of virtual and high-throughput screening. *Nat Rev Drug Discovery*. 882-894 (2002)

Melissa L. Sykes, Jonathan B. Baell, Marcel Kaiser, Eric Chatelain, Danny Ganame, Jean-Robert Ioset, Vicky M Avery. Identification of compounds with anti-proliferative activity against *Trypanosoma brucei brucei* BS427 by a whole cell viability based HTS campaign. *PLoS Negl Trop Dis*. 6 (11) e1896 (2012)

Martine Keenan et al. Selection and optimization of hits from a high-throughput phenotypic screen against *Trypanosoma cruzi*. *Future Med Chem*. 5 (15) 1733-1753 (2013)

Kaiser M, Mäser P, Tadoori LP, Ioset JR, Brun R. Antiprotozoal Activity Profiling of Approved Drugs: A Starting Point toward Drug Repositioning. *PLoS One*. 2015 10(8): e0135556

Shelley, J. C. *et al.* Epik: a software program for pK( a ) prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* 21, 681–691 (2007).

# **Relatório de Atividades de Iniciação Científica**

## **LNCC - Laboratório Nacional de Computação Científica**

Período: maio de 2016 a julho de 2017

Bolsista: Anna Caroline Wayand Martins

Orientador: Jauvane C. de Oliveira

Instituição: Laboratório Nacional de Computação Científica (LNCC)

Área: Medicina assistida por Computação Científica

Plano de Trabalho: Compilação e formatação de informação médica relevante para a elaboração de um atlas 3D de Anatomia, com posterior teste do sistema na comunidade médica, nível médio e demais áreas da saúde.

Projetos: ACiMA – Ambientes Colaborativos e Multimídia Aplicada  
INCT – MACC – Inst. Nac. de C& T em Medicina Assistida por  
Computação Científica

### **1. Introdução**

Ambientes Virtuais Colaborativos (AVCs) permitem que usuários localizados em posições geográficas distintas colaborem através de uma simulação de um mundo sintético controlado por computadores, utilizando uma infra-estrutura de comunicação tal como a Internet.

AVCs tem sido, historicamente, aplicados a diversas áreas do conhecimento, tais como treinamento, Engenharia de Software, Design e Engenharia, Medicina, simulação militar de combate para treinamento de tropas, entre outros.

Tais aplicações, se imersivas, podem incluir um nível a mais de realismo em uma simulação, permitindo que o usuário tenha uma experiência similar àquela que o mesmo teria em similar situação no mundo real. Aplicações de grande escala, que dependem da transferência, em tempo real, de uma massa de dados considerável para visualização colaborativa pelos participantes da simulação impõem uma carga considerável no subsistema de comunicação, o que as torna inviáveis para uso com redes convencionais.

Ambientes virtuais Colaborativos podem ser pouco-imersivos, quando se utiliza apenas um computador de mesa ou equivalente para interface com o usuário, semi-imersivos, no qual existe um nível de imersão no ambiente, e imersivos, onde um usuário tem maior sensação de estar de fato imerso no mundo sintético. A sensação de imersão é implementada através da exploração

dos sentidos humanos, através de dispositivos especiais de entrada e saída ou o uso de técnicas que explorem o modo como os sentidos humanos funcionam. O nível de realismo pode ter um aumento significativo se o ambiente oferecer suporte a dispositivos de interação com o usuário. Esses dispositivos podem ser utilizados para navegação pelo Ambiente Virtual e manipulação de objetos virtuais, entre outras funcionalidades, o que resulta em uma maior sensação de imersão.

As atividades de pesquisa descritas são realizadas no laboratório ACiMA – Laboratório de Ambientes Colaborativos e Multimídia Aplicada – da Coordenação de Ciência da Computação do LNCC e consiste do desenvolvimento de um Atlas 3D de Anatomia.

## **2. Delimitação do problema**

O estudo da Anatomia do corpo humano (suas estruturas e sistemas) consiste na base para o entendimento da Fisiologia do organismo, ou seja, como ele se comporta em indivíduos hígidos, e, susceptivamente, para a compreensão das doenças, seus mecanismos fisiopatológicos, diagnóstico e tratamento. Portanto, torna-se indispensável no currículo dos cursos de Ensino Superior das Ciências da Saúde, entre elas a Medicina, na qual está presente de forma geral na grade curricular do primeiro ano do curso.

Os acadêmicos das áreas da Saúde têm acesso ao estudo da Anatomia Humana através de livros, que proporcionam embasamento teórico, e através de cadáveres e peças anatômicas, para o embasamento prático. No entanto, para o estudo prático, é necessário que haja peças no ambiente acadêmico, que as mesmas estejam disponíveis, e sejam dissecadas de forma a apresentarem as estruturas nas quais o aluno necessita. Em grande parte das instituições, isso não ocorre, dificultando o estudo prático da disciplina.

É possível, no entanto, substituir um cadáver por um sistema de realidade virtual que apresente as estruturas anatômicas em 3D, em um atlas de Anatomia em 3D, o que exige a agregação de uma grande quantidade de informação para que o mesmo tenha relevância médica e para as demais áreas da saúde.

## **3. Resultados**

Tendo iniciado minhas atividades como bolsista de iniciação científica em Maio de 2016, venho desenvolvendo tarefas que envolvem a compilação do conteúdo teórico a ser exposto no atlas virtual. O presente trabalho apresenta-



se em andamento. Utilizando o framework XTK (The X Toolkit) baseado em JavaScript através de WebGL, foram compiladas algumas estruturas, que serão aperfeiçoadas de acordo com as necessidades expressas no conteúdo teórico e no modelo instrucional dos sistemas. A partir do recolhimento das informações, partiu-se para o desenvolvimento de um modelo instrucional a ser utilizado no Ambiente Virtual, além de caracteres técnicos com inserção das informações coletadas e dos processos de interação desenhados no modelo instrucional. O programa foi dividido, a princípio, de acordo com o nível desejado pelo usuário (Médio, Superior Medicina e Superior demais áreas da saúde) e por sistemas do corpo humano (músculo-esquelético, nervoso, cardiovascular). Em cada sistema, há inicialmente uma visão geral, seguida da visão aprofundada, de acordo com a categoria selecionada, com opções para leitura ou narrativa do conteúdo programado. Por ser um modelo 3D, o usuário poderá girá-lo livremente. Utilizou-se, como modelo e padrão para os demais a serem compilados, o Sistema Ósseo, nos 3 níveis descritos acima.

Após a elaboração do conteúdo teórico e instrucional do projeto, os mesmos foram enviados ao Laboratório Nacional de Computação Científica (LNCC) (referentes ao sistema músculo-esquelético) para a realização da compilação, formatação e inclusão no sistema de navegação e ambiente em 3D.

Com a finalização do modelo anatômico virtual, objetiva-se a realização de teste do sistema com membros da comunidade médica e demais áreas da saúde, em parceria com a FMP/Fase, para que o mesmo seja relevante ao aprendizado da Anatomia Humana.

#### **4. Referências**

[1] Oliveira, J. C., Shen, X., & Georganas, N. D. (2000). Collaborative Virtual Environment for Industrial Training and e-Commerce. *IEEE VRTS'2000 (Globecom'2000 Conference's Workshop on Application of Virtual Reality Technologies for Future Telecommunication Systems), San Francisco, CA, USA.*

[2] Oliveira, J. C., Hosseini, M., Shirmohammadi, S., Cordea, M., Petriu, E., Petriu, D., & Georganas, N. D. (2000). VIRTUAL THEATER for Industrial Training: A Collaborative Virtual Environment. *Proc. 4th World Multi-Conference on Circuits, Systems, Communications & Computers (CSCC 2000), Vouliagmeni, Greece.*

- [3] Kirner, T. G.; Kirner, C.; Kawamoto, A. L. S.; Cantão, J.; Pinto, A. & Wazlawick, R. S. (2001) Development of a Collaborative Virtual Environment for Educational Applications, *ACM Web3D, Paderbon, Germany*, 61-68.
- [4] Fernando, T.; Murray, N.; Tan, K. & Wilmalartne, P. (1999). Software Architecture for a Constraint-based Virtual Environments. *ACM VRST'99, London, UK*, 147-154.
- [5] Doppke, J. C.; Heimbigner, D. & Wolf, A. L. (1998) Software Process Modeling and Execution within Virtual Environments, *ACM Trans. on Software Engineering and Methodology*, 7(1), 1–40.
- [6] Daily, M.; Howard, M.; Jerald, J.; Lee, C.; Martin, K.; McInnes, D. & Tinker, P. (2000). Distributed Design Review in Virtual Environments. *ACM Collaborative Virtual Environments, San Francisco, CA, USA*, 57-63.
- [7] Hindmarsh, J.; Fraser, M.; Heath, C.; Benford, S. & Greenhalgh, C. (2000) Object-Focused Interaction in Collaborative Virtual Environments, *ACM Transactions on Computer-Human Interaction*, (7)4, 477–509.
- [8] Seo, J. & Kim, G.J. (2002) Design for Presence: A Structured Approach to Virtual Reality System Design, *Presence: Teleoperators and Virtual Environments*, MIT Press, 11(4), 378-403.
- [9] Greenish, S.; Hayward, V.; Chial, V.; Okamura, A. & Steffen, T. (2002) Measurement, Analysis and Display of Haptic Signals During Surgical Cutting, *Presence: Teleoperators and Virtual Environments*, MIT Press, 11(6), 626-651.
- [10] Macedonia, M., Zyda, M., Pratt, D., Barham, P., & Zeswitz, S. (1994) NPSNET: A Network Software Architecture for Large Scale Virtual Environments. *Presence: Teleoperators and Virtual Environments*, MIT Press, 3(4), 265-287.
- [11] D. Arendash (2004) The unreal editor as a Web 3D authoring environment, *Proceedings of the ninth ACM International Conference on 3D Web Technology*, Monterey, California, EUA, 119-126.
- [12] J. Wang, M. Lewis & J. Gennari (2003) Emerging areas: urban operations and UCAVs: a game engine based simulation of the NIST urban search and rescue arenas, *Proceedings of the 35th ACM/IEEE conference on Winter simulation: driving innovation*, New Orleans, Louisiana, EUA, 1039-1045.

[13] J. Jacobson, Z. Hwang (2002) Unreal tournament for immersive interactive theater, *Communications of the ACM*, 45(1), 39-42.

[14] A. Bierbaum, C. Just, P. Hartling, K. Meinert, A. Baker & C. Cruz-Neira (2001) VR Juggler: A Virtual Platform for Virtual Reality Application Development, *IEEE VR 2001*, Yokohama, Japan.

## **LABORATÓRIO NACIONAL DE COMPUTAÇÃO CIENTÍFICA**

**BOLSISTA:** Arthur S. Bittar

**COORDENADOR:** Jauvane C. de Oliveira

**PERÍODO DE VIGÊNCIA:** Abril de 2016 a Julho de 2017

**MODALIDADE DA BOLSA:** IC

**TÍTULO DO PROJETO CIENTÍFICO:** Modelagem Geométrica e Otimização de Objeto 3D para Ambientes Virtuais Colaborativos

### **OBJETIVO**

A tarefa original do bolsista consiste na geração de modelos geométricos de objetos 3D necessários aos projetos do Laboratório ACiMA, incluindo a necessária otimização das malhas e exportação nos formatos necessários a cada sistema. Tais objetos serão posteriormente importados dentro dos vários AVCs (Ambientes Virtuais Colaborativos).

### **INTRODUÇÃO**

Desenvolvimento e reiteração de modelos tridimensionais de acordo com a necessidade dos projetos em curso no Laboratório ACiMA. São utilizadas técnicas comuns de modelagem poligonal (low poly) para desenvolver assets que possam ser usados em ambientes virtuais. Aos modelos são aplicados shaders básicos (Phone, Blinn, Lambert) e eventualmente diferentes mapas de texturas (Albedo, Normal, Height) quando necessário para simular materiais com maior detalhe. Muitos dos modelos desenvolvidos têm ou tiveram várias versões ao longo de sua elaboração, mudando de acordo com as alterações de briefing dos projetos, passando por retopologia ou decimação de malha para melhoria de desempenho nas máquinas. Quando finalizados, os modelos são então repassados para os respectivos projetos para serem implementados.

### **METODOLOGIA**

O trabalho realizado foi dedicado ao desenvolvimento de modelos geométricos conforme a demanda do Laboratório ACiMA. Tal modelagem 3D inclui também a otimização de malhas com o intuito de permitir um uso mais eficiente nos sistemas desenvolvidos, assim como quaisquer alterações requisitadas pelos integrantes do laboratório, como decimação e deformação de malhas, cortes de geometria, recolocação de pivots, animações simples, etc. Certos modelos desenvolvidos também foram preparados na plataforma Unity3D e exportados em pacotes prontos para uso, diminuindo o ruído entre desenvolvimento e implementação por parte da equipe. O programa principal utilizado durante o trabalho foi o Autodesk Maya.

### **RESULTADOS E DISCUSSÃO**

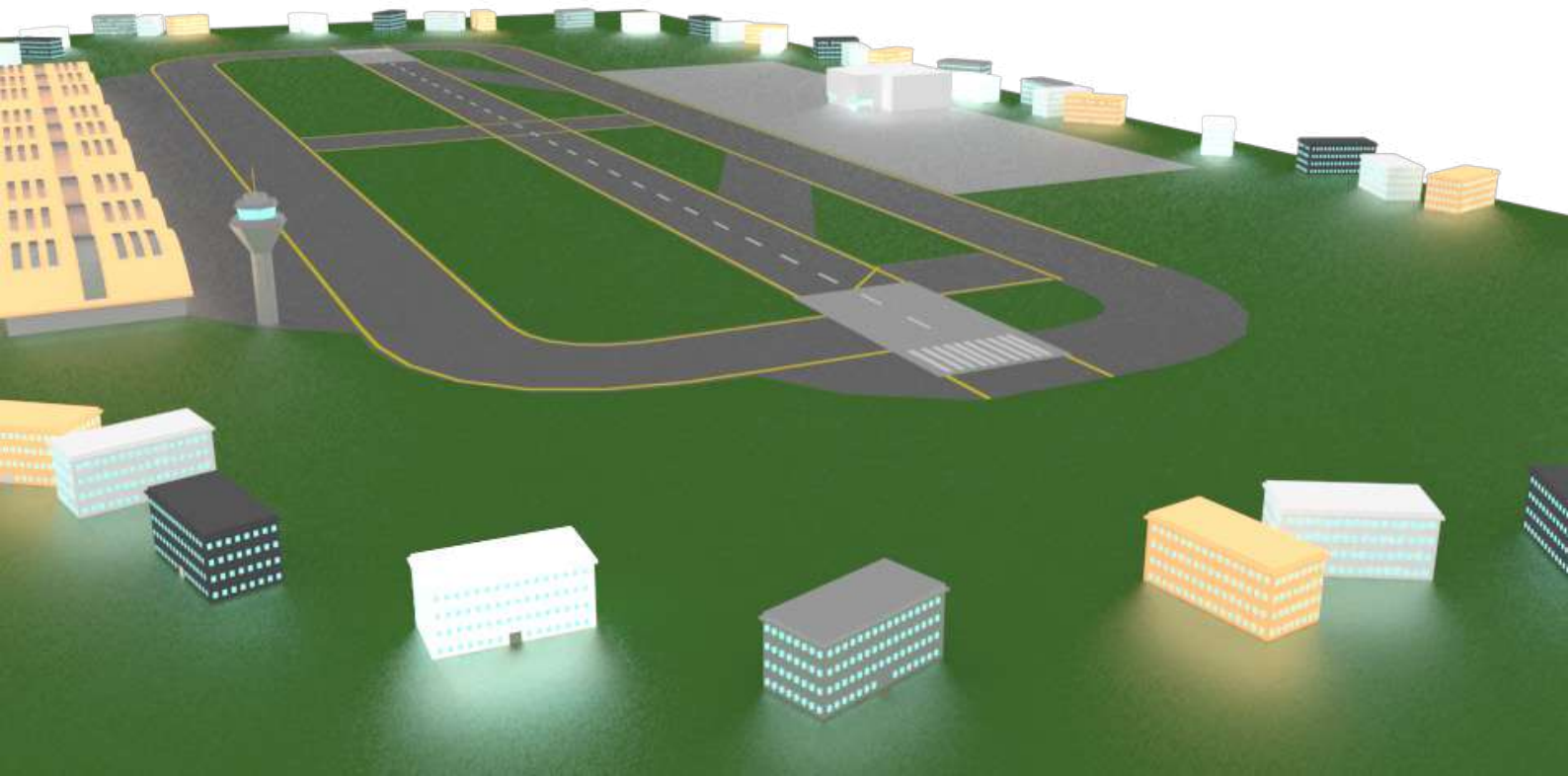
Durante o processo de construção e modelagem — etapas usualmente desenvolvidas por uma equipe — foram realizadas pelo bolsista: busca de referências e concepts; modelagem; texturização; deformações, blendshapes ou outras animações simples; preparação de assets prontos para a plataforma Unity3D; redimensionamento e fechamento de arquivos. Os resultados variaram de acordo com a demanda dos projetos individuais, à medida do feedback recebido por cada peça. Certos projetos requeriam detalhes em certos modelos, outros priorizavam a compreensão visual de ambientes. No geral os modelos foram bem recebidos. Alguns problemas surgiram durante o desenvolvimento de certos trabalhos. Limitações de hardware, especificamente o Phantom, não sustentaram todo o potencial da geometria oferecido pela plataforma Unity3D (observado no projeto de cirurgia de catarata de Letícia Fonseca). Como sugestão seria interessante saber mais do projeto de antemão para diminuir a recorrência de remodelagem, ou receber um briefing fechado com referências e desenhos/clippings de material que deve ser desenvolvido (ajuda visual), visto que a maioria dos pedidos é repassado verbalmente.

### **CONCLUSÕES**

Durante a vigência inteira da bolsa foram desenvolvidos vários assets, ilustrados com renders/screenshots a seguir.

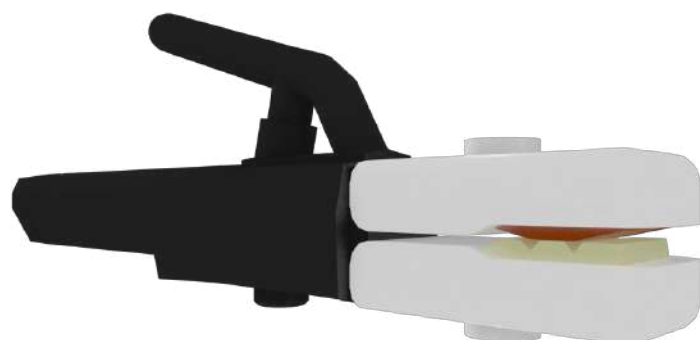
- Foi reimportado e manipulado um modelo de aeroporto e diversos de seus elementos, criada uma pista de decolagem, prédios ao entorno, hangares e torre de comando. O objetivo do projeto era simular a situação de embarque para pessoas que têm pavor de viajar de avião.
- Foi criado um porta-eletródo e uma sala simples (não incluída nesse documento) para uma simulação de aprendizado de métodos de solda com catodo.
- Foi modelado um olho para simulação de cirurgia de catarata, com várias camadas e elementos de marcação de corte, lente fragmentada e uma lente IOL de reposição do cristalino. Posteriormente foram modelados várias instrumentos para serem utilizadas durante a simulação das etapas da cirurgia.
- Foi encontrado um modelo freeware de uma cidade básica, ao qual foi adicionado um prédio, um elevador panorâmico, sinalização de rua, postes, lixeiras e uma praia. O objetivo desse projeto era simular uma situação para pessoas com pavor de altura.
- Um modelo do esqueleto humano foi reimportado, reorganizado e exportado em novo formato, tanto nas dimensões originais quanto pós decimação. O novo formato atendia à limitações de contagem de triângulos e o seu processamento para visualização em tempo real em um navegador.
- Foram modeladas diversas ferramentas específicas para uma simulação de um procedimento de punção lombar.
- Foi modelado um sistema de fixação de um boneco de treino a uma Wii-Board sensível à pressão assim como elementos para demonstração de instrução do seu uso em uma simulação de massagem cardíaca.

# AEROPORTO (CHECK-IN, GATE) & PISTA

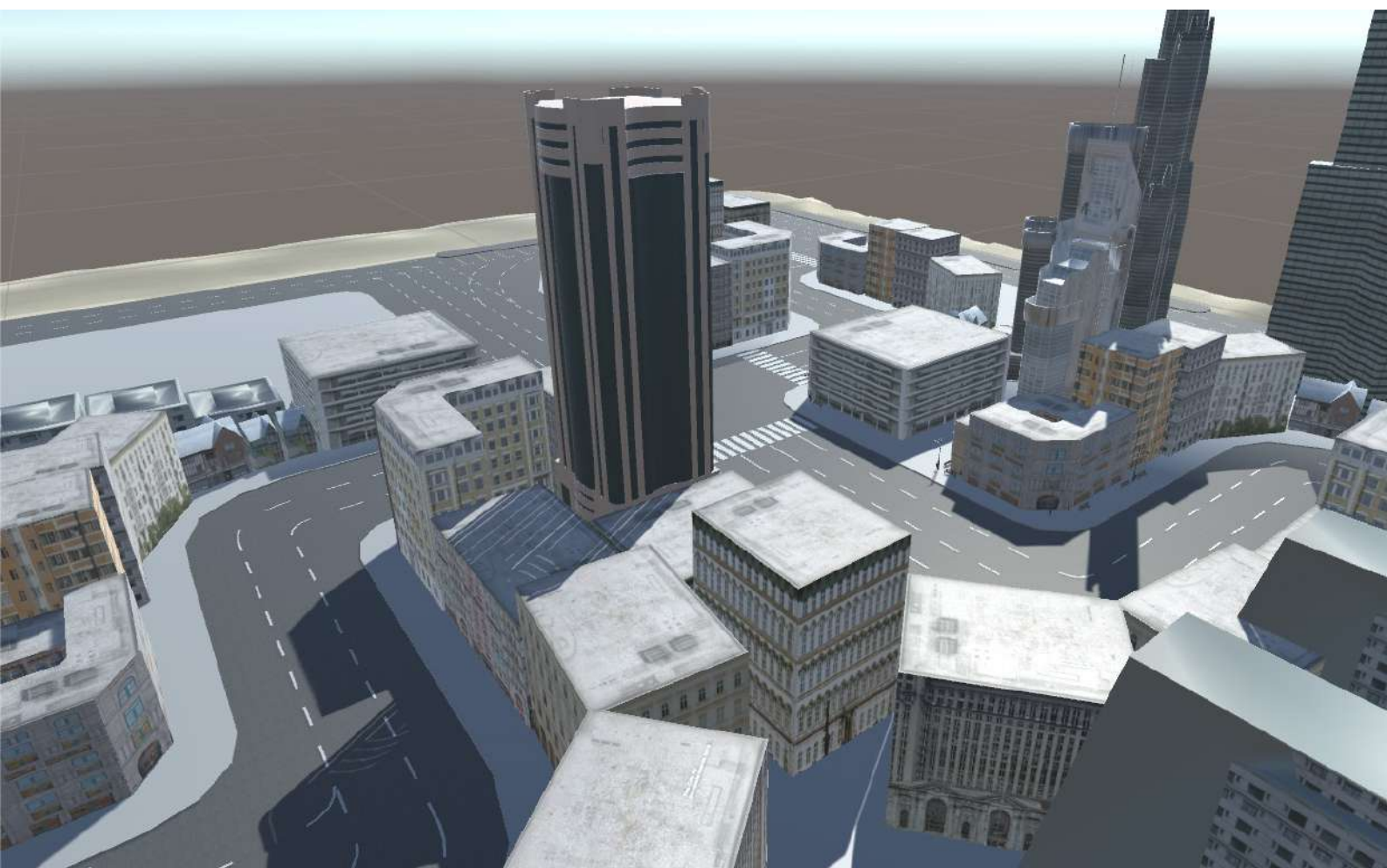




## PORTA-ELETRODO

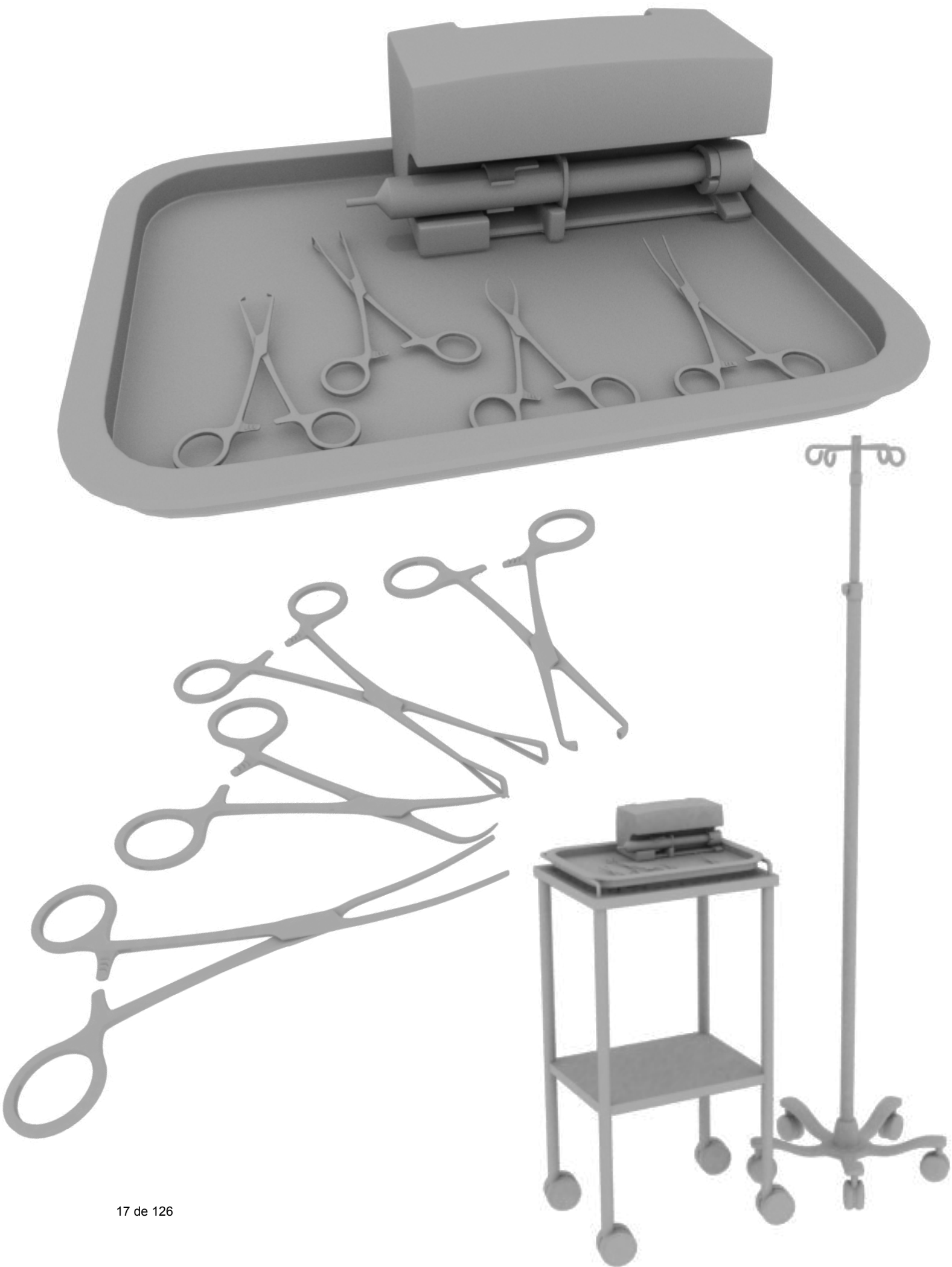




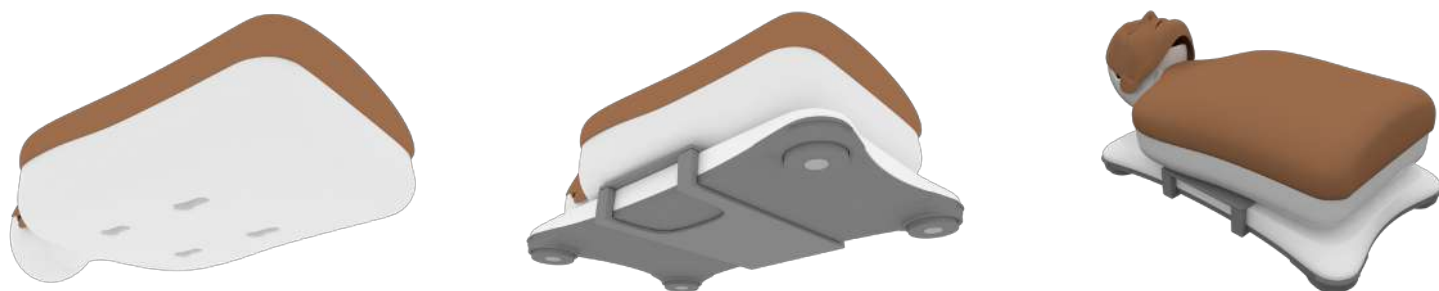
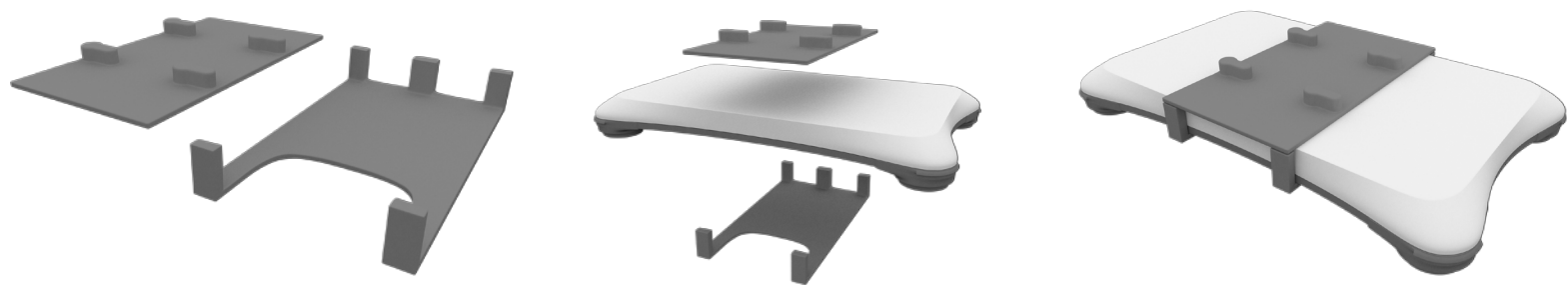








# FIXAÇÃO DE BONECO À WII-BOARD E ELEMENTOS PARA INSTRUÇÃO MASSAGEM CARDÍACA



# Relatório de Iniciação Científica - LNCC

Bernardo Botelho Antunes da Costa

## 1. Dados gerais

- Título do projeto: Modelagem da Malha Aérea Doméstica Brasileira com Grafos Multi-Aspecto
- Nome do bolsista: Bernardo Botelho Antunes da Costa
- Nome do orientador: Artur Ziviani (LNCC)
- Tipo de bolsa e período do relatório: Iniciação Científica (07/2016-06/2017)

## 2. Objetivos

Este trabalho se propõe a modelar a malha aérea doméstica brasileira com Grafos Multi-Aspecto (MAG) [1], capaz de representar redes complexas variantes no tempo, com múltiplas camadas, ou com ambas as características. Um conjunto de vértices, camadas, instantes de tempo ou qualquer outra característica independente presente na rede complexa pode ser considerado como um aspecto do MAG. A malha aérea foi modelada com quatro aspectos: aeroportos, companhias aéreas, tempo (escala de horários), e época da base de dados para comparação estrutural entre dois períodos de tempo distintos.

## 3. Introdução

Com o aumento da tecnologia e dos meios de transporte, a aviação civil tem papel cada vez mais central na economia de um país, visto seu poder de interligar pessoas de diferentes regiões, de forma acessível e rápida. Trabalhos anteriores [2,3] tratam a modelagem da malha aérea como um grafo simples, o que sonega diversas informações relevantes sobre a rede real, tais como: número de voos em cada rota e possibilidade de conexão entre os voos.

Este trabalho tem como objetivo a modelagem da malha aérea doméstica brasileira usando Grafos Multi-Aspecto (MAG), bem como o desenvolvimento de algoritmos que permitem a modelagem e análise das perspectivas de multicamada e variante no tempo.



Figura 1: Aresta do modelo proposto representando uma relação binária entre os múltiplos aspectos considerados usando MAG para modelagem da malha aérea.

## 5. Resultados e Discussão

Com o modelo desenvolvido foi permitido que com um único objeto matemático, o MAG, fossem analisadas diferentes características da malha aérea brasileira, relacionadas a diferentes perspectivas relacionadas a seguir:

- **Multi-camada:** onde os voos de cada empresa aérea formam suas próprias redes complexas, que juntas formam a malha aérea total.
- **Variante no tempo:** onde cada dia da semana pode ser analisado de forma separada, discriminando cada ocorrência dos voos, que permitem calcular com maior exatidão a possibilidade de conexões.
- **Multi-escala:** onde dois períodos de tempo podem ser analisados no mesmo objeto, nesse caso, analisamos os dias da semana e a época que os dados foram coletados.
- **Visão rota x vôos:** definimos a diferença entre rota e vôo como sendo rota o caminho possível entre dois aeroportos e vôo a quantidade de vezes que uma empresa usa essa rota para realizar uma viagem. Realizando operações com o MAG é possível transformá-lo em um dígrafo, onde extraímos informações sobre as rotas, ou seja, a conexão possível entre os aeroportos. Também extraímos a visão multi-dígrafo, onde é possível quantificar os voos que utilizam uma rota entre dois aeroportos.

Resultados preliminares deste trabalho foram publicados em [4] junto aos resultados alcançados pelo trabalho do também bolsista de IC João Victor M. Bechara.

## 6. Conclusões

Este trabalho de modelagem da malha aérea doméstica brasileira possibilitou a construção de um único objeto matemático através do uso de MAGs, contendo todas as informações relevantes para as análises posteriores. O MAG permite o uso de algoritmos clássicos da literatura para grafos direcionados na análise de redes complexas representando redes de alta ordem (variantes no tempo, multicamada, etc). Logo, foi possível transportar tais algoritmos para obter informações que sejam consideradas relevantes para a análise da malha aérea brasileira sob diferentes perspectivas.

Em resumo, este trabalho proveu a base para permitir a modelagem e análise da malha aérea brasileira como uma rede multicamada variante no tempo através de um único objeto matemático pelo uso de MAGs. A visão multicamada da análise permitiu revelar as estratégias de cada companhia aérea na estruturação de sua malha aérea e a visão da rede variante no tempo permitiu a análise multi-escala e a análise comparativa das estruturas das companhias aéreas considerando rotas e voos. Além disso, dada a análise comparativa entre períodos de tempo distintos também foi possível avaliar-se o impacto da crise econômica na operação das diferentes companhias aéreas.

Com base nos resultados preliminares publicados em [4], um artigo mais completo encontra-se em fase final de redação visando sua submissão a um periódico indexado.

## **7 Referências bibliográficas**

- [1] K. Wehmuth, E. Fleury, A. Ziviani, On MultiAspect graphs, Theoretical Computer Science (TCS), Elsevier, vol. 651, pp. 50-61, outubro de 2016.
- [2] G. S. Couto, A. P. C. Silva, L. B. Ruiz, F. Benevenuto, Structural Properties of the Brazilian Air Transportation Network, Anais da Academia Brasileira de Ciências (AABC), vol. 87, no. 3, pp. 1653-1674, setembro de 2015.
- [3] T. Verma, N. A. M. Araújo, H. J. Herrmann, Revealing the structure of the world airline network, Scientific Reports, vol. 4, 2014.
- [4] B. B. A. Costa, J. V. M. Bechara, K. Wehmuth, A. Ziviani, Modelagem e Análise da Malha Aérea Doméstica Brasileira com Grafos MultiAspectos (MAGs), Anais do X Encontro Acadêmico de Modelagem Computacional - EAMC 2017, Petrópolis, RJ, fevereiro de 2017.

## **Controle Ótimo Para Sistemas Elásticos**

**Débora de Oliveira Souza**

**Jaime Edilberto Munoz Rivera**

**PIBIC – 03/2017 até 07/2017**

O objetivo inicial da pesquisa consistiu em investigar e conhecer os principais conceitos acerca da teoria de minimização e controle ótimo e suas aplicações. Contudo, nossa meta, objetivos finais, é desenvolver técnicas de modelagem para problemas de elasticidade e epidemiologia, assim como, as técnicas numéricas e computacionais direcionadas a modelagem computacional destes problemas.

A teoria e os problemas estudados nessa pesquisa, dispõe de um sistema dinâmico, que evolui no tempo, na qual pode-se agir por meio de um comando ou controle. Além disso, a teoria do controle analisa as propriedades dos sistemas com o intuito de conduzir de um determinado estado inicial para um estado final, respeitando suas restrições. Com base nas anotações realizadas por meio de leituras acerca do problema de controle ótimo, fez-se uma descrição dos pontos destacados, como a teoria do controle ótimo (ou teoria do controle) e a otimização.

O modelo específico que estamos tratando é o modelo de Timoshenko para vigas estacionárias. O primeiro ponto é deduzir o problema desde o ponto de vista físico. Logo desenvolver o método de diferenças finitas para encontrar soluções aproximadas do modelo.

Desenvolvemos o modelo aproximado e estamos tratando as aproximações numérica, os métodos computacionais serão desenvolvidos usando Matlab.

### **Referencias**

SILVA, C. J.; TORRES, D. F. M.; TRÉLAT, E. O controle ótimo e as suas múltiplas aplicações. Bol. Soc. Port. Mat. N. 61, p. 11–37, 2009.

PEREIRA, C. M.; CECCONELLO, M. S.; BASSANEZI, R. C. Controle ótimo em sistemas baseados em regras fuzzy. Biomatemática, v. 23, p. 147-164, 2013.

SARAMAGO, S. P.; STEFFEN JR, V. Introdução às técnicas de otimização em engenharia. Horizonte científico, v. 2, n. 2, p. 1-28. 2008.



# PRIMEIRO RELATÓRIO DE INICIAÇÃO CIENTÍFICA

ALUNO: FILIPE ALVES DA SILVA DE MOURA

## SUMÁRIO

1. Dados Gerais	1
2. Objetivos do trabalho	1
3. Introdução	1
4. Método dos Elementos Finitos	2
4.1. Conceitos fundamentais: um problema de valor de contorno unidimensional	2
4.2. Formulação forte do problema	2
4.3. Formulação Fraca ou variacional do problema	3
5. Conclusão	3
Referências	4

## 1. DADOS GERAIS

**Título do projeto:** Derivada Topológica na Mecânica da Fratura

**Nome do bolsista:** Filipe Alves da Silva de Moura

**Nome do orientador:** Antonio André Novotny

**Nome do co-orientador:** Marcel Xavier

**Tipo de bolsa e período do relatório:** Iniciação científica - Março/Julho

## 2. OBJETIVOS DO TRABALHO

O objetivo do presente trabalho é apresentar os primeiros aspectos que são fundamentais para o desenvolvimento do método dos elementos finitos. Este método será empregado para resolver numericamente problemas envolvidos no tratamento de questões no contexto da mecânica da fratura.

## 3. INTRODUÇÃO

A finalidade deste primeiro relatório consiste em apresentar os aspectos iniciais e fundamentais para o desenvolvimento do método dos elementos finitos como apresentado em [1]. Inicialmente, será proposto o problema da linha elástica. A ideia é utilizar um problema simples, cuja solução exata é conhecida, para apresentar o método de elementos finitos. O trabalho está organizado da seguinte forma. Na Seção 4.1 é apresentado o problema em questão. Depois, na Seção 4.2 são introduzidas as equações que caracterizam o problema. Em seguida, na Seção 4.3, é apresentada a formulação fraca associada. Por fim, na Seção 5 são feitas as considerações finais.

---

*Date:* 10 de julho de 2017.

*Key words and phrases.* Elementos finitos, formulação forte, formulação fraca.

## 4. MÉTODO DOS ELEMENTOS FINITOS

**4.1. Conceitos fundamentais: um problema de valor de contorno unidimensional .** Nesta seção serão introduzidos os principais pontos para apresentar o método dos elementos finitos.

### Observações introdutórias e preliminares

Para obter a solução de um problema de valor de contorno utilizando o método de elementos finitos são necessários:

- A forma variacional ou fraca do problema;
- A solução aproximada da equação variacional através do uso de funções de elementos finitos.

Para esclarecer alguns conceitos básicos vamos começar com um exemplo simples. Suponha que deseja-se resolver a seguinte equação para  $u$ :

$$u_{,xx} + f = 0 , \tag{4.1}$$

onde a vírgula representa uma diferenciação (i.e.,  $u_{,xx} = d^2u/dx^2$ ) e  $f$  é uma dada função suave de valor escalar definida sobre o intervalo unitário. Então, escreve-se

$$f : [0, 1] \rightarrow \mathbb{R} , \tag{4.2}$$

onde  $[0, 1]$  representa o intervalo unitário (i.e., o conjunto dos pontos  $x$  tal que  $0 \leq x \leq 1$ ) e  $\mathbb{R}$  representa os números reais. Em palavras, (4.2) afirma que para um dado  $x \in [0, 1]$   $f(x)$  é real. (Frequentemente usaremos a notação  $\in$  para indicar "em" ou "um membro de". Portanto, para cada  $x \in [0, 1]$ ,  $f(x) \in \mathbb{R}$ . Também  $[0, 1]$  é dito ser o domínio de  $f$  e  $\mathbb{R}$  é sua imagem.

A função  $f$  é descrita como sendo suave. Isso significa, a grosso modo, que quando desenharmos o gráfico da função  $f$  a mesma será uma curva suave sem descontinuidades ou torções. Isso é feito para evitar dificuldades técnicas.

A equação (4.1) é conhecida por governar o deslocamento transversal de uma corda tensionada e também o deslocamento longitudinal de uma haste elástica. Nestes casos, parâmetros físicos tais como a magnitude da tensão na corda; ou o módulo de elasticidade no caso da haste, aparecem na equação (4.1). No entanto, esses parâmetros são omitidos com o objetivo de simplificar futuros desenvolvimentos.

Antes de prosseguir, serão introduzidas algumas notações e terminologias adicionais. Seja  $]0, 1[$  o intervalo unitário aberto (i.e., o conjunto de pontos tal que  $0 < x < 1$ ).  $]0, 1[$  e  $[0, 1]$  são referidos como intervalos unitários aberto e fechado, respectivamente. Para simplificar o desenvolvimento, vamos introduzir a seguinte notação:  $\Omega = ]0, 1[$  para denotar o conjunto unitário aberto e  $\bar{\Omega} = [0, 1]$  para denotar o conjunto unitário fechado.

**4.2. Formulação forte do problema .** Um problema de valor de contorno, tal como o problema (4.1), envolve a imposição de condições contorno na função  $u$ . Existe uma variedade de possibilidades. Vamos supor que  $u$  é obrigada a satisfazer as seguintes condições

$$\begin{cases} u(1) & = g ; \\ -u_{,x}(0) & = h , \end{cases} \tag{4.3}$$

onde  $g$  e  $h$  são constantes dadas. As equações em (4.3) requerem que  $u$  tome o valor  $g$  em  $x = 1$  e a derivada de  $u$  (isto é inclinação) assume o valor  $-h$  em  $x = 0$ , respectivamente.

A forma forte do problema de valor de contorno, (S), é dado por: dado  $f : \bar{\Omega} \rightarrow \mathbb{R}$  e constantes  $g$  e  $h$ , encontrar  $u : \Omega \rightarrow \mathbb{R}$  tal que:

$$\begin{cases} u_{,xx} + f = 0 ; \\ u(1) = g ; \\ -u_{,x}(0) = h , \end{cases} \quad (4.4)$$

Quando escrevemos  $u_{,xx} + f = 0$  em  $\Omega$  queremos dizer que  $u_{,xx} + f(x) = 0$  para todo  $x \in \Omega$ . A solução exata do problema  $S$  é trivialmente obtida, à saber,

$$u(x) = g + (1-x)h + \int_x^1 \left\{ \int_0^y f(z)dz \right\} \quad (4.5)$$

**4.3. Formulação Fraca ou variacional do problema .** Para definir a forma fraca ou variacional do problema ( $S$ ) é preciso caracterizar duas classes de funções. A primeira delas é composta pelas chamadas soluções admissíveis. Tais funções devem satisfazer a condição de contorno  $u(1) = g$ . A outra condição de contorno não será exigida na definição. além disso, para que certas expressões sejam empregadas e façam sentido, será exigido que as derivadas das soluções admissíveis sejam quadrado integráveis. Ou seja, se  $u$  é uma solução admissível, então

$$\int_0^1 (u_{,x}^2)dx < \infty \quad (4.6)$$

Funções que satisfazem o problema (4.6) são chamadas de funções  $H^1$ . Denota-se  $u \in H^1$ . Algumas vezes o domínio é explicitamente incluído (i.e.,  $u \in H^1 [0, 1]$ ).

Portanto, a coleção de funções admissíveis, denotada por  $\mathcal{S}$ , consiste de todas as funções que são quadrado integráveis deriváveis e tomam o valor  $g$  em  $x = 1$ . Desta forma, o conjunto  $\mathcal{S}$  é escrito da seguinte forma

$$\mathcal{S} \{u; u \in H^1, u(1) = g\} \quad (4.7)$$

O fato de  $\mathcal{S}$  ser um conjunto de objetos é indicado pela presença das chaves na equação acima.

A segunda coleção de funções é chamada de funções teste ou variações. Esta coleção é muito semelhante as soluções admissíveis exceto pelo fato de que sobre elas é exigida a homogeneidade na condição de contorno, ou seja,  $w(1) = 0$  on  $x = 1$ . Esta coleção é denotado por  $\mathcal{V}$  e definida da seguinte forma:

$$\mathcal{V} = \{w; w \in H^1, w(1) = 0\} \quad (4.8)$$

Portanto, a formulação fraca, ou variacional, do problema (4.4) é definida da seguinte forma: Dadas  $f$ ,  $g$ , e  $h$ , como antes, encontrar  $u \in \mathcal{S}$  tal que para todo  $w \in \mathcal{V}$

$$\int_0^1 w_{,x}u_{,x}dx = \int_0^1 wfdx + w(0)h \quad (4.9)$$

## 5. CONCLUSÃO

No presente trabalho foi proposto o problema da linha elástica com o objetivo de utilizá-lo para apresentar o método dos elementos finitos. Em particular, foram apresentadas as equações que governam o problema tanto na sua forma forte quanto na forma variacional. A ideia é, a partir destes elementos, apresentar o método Galerkin para resolvê-lo numericamente.

## REFERÊNCIAS

- [1] T. J. R. Hughes. *The finite element method: linear static and dynamic finite element analysis*. General Publishing Company, 2000.

(F. Moura) LABORATÓRIO NACIONAL DE COMPUTAÇÃO CIENTÍFICA LNCC/MCT, COORDENAÇÃO DE MATEMÁTICA APLICADA E COMPUTACIONAL, AV. GETÚLIO VARGAS 333, 25651-075 PETRÓPOLIS - RJ, BRASIL

*E-mail address:* `filipemec.ucp@gmail.com`



---

Laboratório Nacional de Computação Científica  
Coordenação de Ciência da Computação (CCC)

## Relatório de atividades

Programa Institucional de Bolsas de Iniciação Científica  
PIBIC

Avaliação de Sistemas de Computação Distribuída de Alto  
Desempenho: Um Estudo sobre Estratégias para  
Monitoramento de Desempenho

Gabrieli Dutra Silva

Orientador - Prof. Bruno Schulze

Período Agosto/2016 a Julho/2017

Julho 2017

---



## Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Objetivo</b>	<b>2</b>
<b>3</b>	<b>Metodologia</b>	<b>2</b>
<b>4</b>	<b>Resultados e Discussão</b>	<b>3</b>
<b>5</b>	<b>Conclusões</b>	<b>3</b>
5.1	Atividades futuras . . . . .	4



## Dados Referentes à Bolsa

- **Instituição:** Laboratório Nacional de Computação Científica
- **Bolsista:** Gabrieli Dutra Silva
- **Coordenador/Orientador:** Bruno Schulze (CCC-LNCC)
- **Coorientadora:** Mariza Ferro (LNCC)
- **Modalidade:** Bolsa de Iniciação Científica
- **Período de Vigência da Bolsa:** Agosto de 2016 a Julho de 2017
- **Projetos Relacionados:** Computação Científica Distribuída (ComCiDis) e *High Performance Computing for Energy* (HPC4E)

## 1 Introdução

A Computação Científica Distribuída de Alto Desempenho é um segmento da ciência da computação que tem como objetivo a melhoria do desempenho de aplicações distribuídas e paralelas, utilizando complexas infraestruturas computacionais. Esse tipo de infraestrutura tem se tornado crucial para a pesquisa científica em muitos domínios de investigação. Um desses domínios é a indústria de energia, cujas aplicações envolvem intensivas simulações numéricas, fundamentais na exploração de novas fontes de energia, tais como, energia eólica e petróleo e gás. Dentro deste contexto, o projeto *High Performance Computing for Energy* (HPC4E) <sup>1</sup>, no qual este trabalho se insere, visa o desenvolvimento de simulações altamente escaláveis que viabilizem a geração de energia de maneira mais eficiente. Entre os objetivos técnicos do projeto está a análise de desempenho dessas aplicações, a identificação de possíveis gargalos na sua execução, bem como a eficiência no consumo de energia elétrica das aplicações.

Para alcançar os objetivos do projeto, monitorar aplicações, ambiente e degradação de desempenho são tarefas indispensáveis. Um ponto é o monitoramento e a avaliação de desempenho das aplicações científicas, sendo importante caracterizar, para os diferentes modelos, quais seus principais requisitos em diferentes arquiteturas computacionais. Outro fator de grande importância é a relação que existe entre esses modelos de aplicações científicas e o consumo de energia elétrica necessários para sua execução e a avaliação continuada dos ambientes computacionais de HPC, verificando seus níveis de desempenho, possíveis falhas e necessidades de melhorias. Porém, realizar todas essas avaliações não é tarefa trivial, pois cada um desses aspectos envolvem um conjunto de parâmetros diferentes, obtidos por meio de um grande conjunto experimental. Por exemplo, na avaliação das aplicações individuais, para caracterizar os seus requisitos computacionais é necessário medir os percentuais e tempos de utilização de recursos (CPU, E/S, Memória, etc.), necessários para a sua execução. Além disso, é preciso relacionar cada um desses fatores com a quantidade de energia elétrica consumida. Porém, esse tipo de avaliação, envolvendo esse conjunto de métricas, não é comum de ser encontrada em uma única abordagem ou ferramenta [Silva et al. 2016].

<sup>1</sup><https://hpc4e.eu/>



Portanto, é necessário realizar um estudo sobre as diferentes abordagens de monitoramento, encontrando soluções que viabilizem a coleta de medidas referentes aos diferentes aspectos mencionados. Além disso, é essencial que esses dados coletados sejam adequadamente analisados, pois a avaliação e a interpretação dos mesmos é o que viabilizará encontrar as soluções que poderão contribuir para o projeto.

## 2 Objetivo

A coleta e observação dos parâmetros relevantes a serem avaliados, permite não somente avaliar se uma aplicação obteve bom desempenho, mas também fornece os subsídios necessários para a compreensão dos requisitos computacionais das aplicações científicas, seus gargalos de execução e a relação desses requisitos com o consumo de energia. O principal objetivo é realizar este tipo de análise, em diferentes arquiteturas de alto desempenho e com diferentes modelos de aplicação e tamanhos de problema. Para viabilizar essa coleta é fundamental a utilização de ferramentas de análise de desempenho. Porém as ferramentas existentes possuem diferentes técnicas para realizar o registro de informações comportamentais [Reed 1994]. Além disso, como mencionado, não é comum encontrar uma única abordagem nem ferramenta de monitoramento que permita coletar todos estes aspectos sobre a execução da aplicação.

Portanto, com base no que foi exposto, o objetivo é investigar as diferentes técnicas existentes para o monitoramento de aplicações, que viabilizem a coleta e análise dos parâmetros considerados relevantes neste projeto. Ainda se faz necessário realizar um estudo sobre as diferentes ferramentas de monitoramento, suas vantagens, limitações e viabilidade de uso para coletar os parâmetros definidos.

Além disso, a partir das definições e estudos, caso nenhuma ferramenta se mostre adequada, auxiliar no desenvolvimento de uma ferramenta de coleta no grupo ComCiDis <sup>2</sup> utilizando as abordagens de monitoramento estudadas. Como resultado final deve-se integrar à essa ferramenta de monitoramento um mecanismo de análise e visualização de dados.

## 3 Metodologia

Para cumprir os objetivos acima descritos foi adotada uma metodologia para o desenvolvimento da pesquisa.

- Estudo bibliográfico que possibilitou levantar o estado da arte das abordagens e ferramentas de monitoramento de desempenho e consumo de energia.
- Análise técnica de algumas ferramentas selecionadas, como por exemplo PowerAPI [Bourdon et al. 2013], Intel VTune <sup>3</sup> e Paraver <sup>4</sup>, verificando requisitos de instalação e uso bem como suas vantagens e limitações.
- Colaboração, a partir dos estudos realizados, para o desenvolvimento de uma ferramenta de monitoramento.
- Implementação de uma ferramenta web, ainda em desenvolvimento, para análise e visualização dos dados.

<sup>2</sup><http://comcidis.lncc.br/>

<sup>3</sup><https://software.intel.com/en-us/get-started-with-vtune-for-systems>

<sup>4</sup><https://tools.bsc.es/paraver>



- Estudo para integrar a ferramenta web à ferramenta de monitoramento, ainda em andamento.

## 4 Resultados e Discussão

Seguindo os passos de pesquisa definidos na Seção 3, foi possível identificar que embora existam diversas ferramentas para análise de desempenho, cada uma delas coletam um conjunto específico de parâmetros. Porém, muitos dos parâmetros que se deseja analisar neste projeto não são coletados em uma mesma ferramenta, o que torna muito difícil a utilização das ferramentas já existentes. Além disso, muitas ferramentas obtêm parâmetros que não são foco de análise neste trabalho, como por exemplo a análise aprofundada referente a cada função existente nos códigos das aplicações, os quais são mais adequados quando se deseja otimizar tais códigos, o que não é o objetivo. Já em outras ferramentas o foco é analisar o desempenho do sistema e não da aplicação que está em execução. Logo, uma única ferramenta que monitore desempenho sobre a execução de aplicações científicas, mostrando seus consumos de recursos computacionais (tempos e percentuais de uso de CPU, I/O, Memória, rede) e consumo de energia, relacionando esses aspectos não foi encontrada. Assim, foi possível contribuir para o desenvolvimento de uma ferramenta de monitoramento no grupo ComCiDis.

Além de coletar é preciso visualizar e analisar essas informações do monitoramento realizado. Porém, é necessário organizar todos esses dados em planilhas e gerar novos gráficos a cada novo conjunto de experimentos executados. Mas, esse tipo de abordagem de análise e visualização é altamente custoso, pois envolve um grande conjunto experimental com diferentes aplicações, tamanhos de problema e em diferentes arquiteturas. Assim, existe a necessidade do desenvolvimento de uma ferramenta para análise e visualização desses dados. No entanto, tanto a implementação dessa ferramenta quanto o estudo de sua integração com a ferramenta de monitoramento ainda estão em desenvolvimento.

Mostrando a importância dos resultados obtidos até o momento, foram realizadas as seguintes publicações:

- [Silva et al. 2016] - Silva, G. D., Klôh, V. P., Ferro, M., and Schulze, B. (2016). Abordagens de monitoramento de desempenho em apoio a pesquisa científica. Anais do XVII Simpósio em Sistemas Computacionais de Alto Desempenho (Wscad-WIC 2016), pages 74–79, Aracaju-SE. Sociedade Brasileira de Computação, SBC.
- [Klôh et al. 2016] - Klôh, V. P., Ferro, M., Silva, G. D., and Schulze, B. (2016). Performance monitoring using nagios core. Relatórios de Pesquisa e Desenvolvimento do LNCC 03/2016, Laboratório Nacional de Computação Científica, Petrópolis - RJ.
- [Ferro et al. 2017] - Ferro, M., Yokoyama, A., Klôh, V. P., Silva, G. D., and Schulze, B. (2017). Analysis of GPU power consumption using internal sensors. Workshop em Desempenho de Sistemas Computacionais e de Comunicação (WPerformance 2017). Este trabalho foi premiado entre os três melhores trabalhos do evento.

## 5 Conclusões

Como mencionado, a fim de contribuir com os objetivos do projeto HPC4E, é necessário compreender os requisitos computacionais das aplicações científicas, seus gargalos de execução e a relação desses requisitos com o consumo de energia. Para isso é preciso coletar

os parâmetros considerados relevantes no projeto. No entanto, não foi possível encontrar uma única ferramenta que coletasse todos os aspectos estudados e ainda fosse de uso livre. Portanto, o que se concluiu até o momento é que para coletar todos os parâmetros relevantes, houve a necessidade da implementação de uma nova ferramenta. O estudo bibliográfico sobre as abordagens de monitoramento e de várias ferramentas existentes, apontando suas vantagens, limitações e viabilidade de uso para pesquisa científica foi usada como base para o desenvolvimento da ferramenta. Além disso, para visualizar e analisar estes parâmetros, há necessidade de integrar à essa ferramenta de coleta, um mecanismo de análise e visualização dos dados, o qual encontra-se em desenvolvimento.

## 5.1 Atividades futuras

Diante do que foi exposto na Seção 4, as etapas de implementação e integração da ferramenta ainda estão em andamento e assim são previstas as seguintes atividades futuras:

- Desenvolvimento e publicação de relatórios técnicos e artigos científicos.
- Desenvolvimento da ferramenta de análise e visualização de dados.
- Testes e integração no ambiente de análise desenvolvido.

## Referências

- [Bourdon et al. 2013] Bourdon, A., Nouredine, A., Rouvoy, R., and Seinturier, L. (2013). Powerapi: A software library to monitor the energy consumed at the process-level. *ERCIM News*, 2013(92). 2
- [Ferro et al. 2017] Ferro, M., Yokoyama, A., Klôh, V. P., Silva, G. D., and Schulze, B. (2017). Analysis of gpu power consumption using internal sensors. In de Computação (SBC), S. B., editor, *Workshop em Desempenho de Sistemas Computacionais e de Comunicação (WPerformance 2017)*. 3
- [Klôh et al. 2016] Klôh, V. P., Ferro, M., Silva, G. D., and Schulze, B. (2016). Performance monitoring using nagios core. Relatórios de Pesquisa e Desenvolvimento do LNCC 03/2016, Laboratório Nacional de Computação Científica, Petropolis - RJ. 3
- [Reed 1994] Reed, D. A. (1994). Experimental analysis of parallel systems: techniques and open problems. In *International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pages 25–51. Springer. 2
- [Silva et al. 2016] Silva, G. D., Klôh, V. P., Ferro, M., and Schulze, B. (2016). Abordagens de monitoramento de desempenho em apoio a pesquisa científica. In de Computação (SBC), S. B., editor, *Anais do XVII Simposio em Sistemas Computacionais de Alto Desempenho (Wscad-WIC 2016)*, pages 74–79, Aracaju-SE. Sociedade Brasileira de Computação, SBC. <http://www.lbd.dcc.ufmg.br/colecoes/wscad-wic/2016/010.pdf>. 1, 3

Bolsista:  
*Guilherme Villar Piccinini de Magalhães*

Orientadores:  
*Elson M. Toledo(LNCC)*  
*Regina Célia P. Leal Toledo (UFF)*

**Relatório do projeto:**  
**SIMULAÇÃO NUMÉRICA E COMPUTACIONAL DO TRÁFEGO VIÁRIO**

LNCC  
PROGRAMA PIBIC/PIBITI  
2016-2017

Julho de 2017

## 1- Objetivo

O aumento do fluxo de veículos em vias urbanas ou não, tem afetado diretamente a mobilidade e a qualidade de vida dos cidadãos, particularmente daqueles que vivem em grandes centros. O entendimento da dinâmica do tráfego viário, através da modelagem de seu comportamento, poderá permitir a realização de simulações computacionais que auxiliem na tomada de decisões, planejamento e controle do tráfego viário. Esse projeto, que vem sendo desenvolvido em parceria com a Universidade Federal Fluminense (UFF), pretende fornecer um modelo numérico e computacional capaz de representar, qualitativamente e quantitativamente, a dinâmica do fluxo do tráfego considerando a influência que diferentes perfis de comportamentos de motoristas exercem sobre sua dinâmica. A modelagem desses diferentes perfis de comportamento é a principal característica que o diferencia dos trabalhos encontrados na literatura. Resultados importantes e encorajadores já foram obtidos propondo um modelo de antecipação que simula inicialmente a percepção do motorista em relação ao movimento do veículo que está à sua frente para posteriormente, ajustar sua velocidade. Uma Função de Densidade de Probabilidade (FDP) não uniforme tem sido utilizada para melhor representar a incerteza no comportamento desses motoristas. Esse conjunto de ações torna o modelo inovador e permite que melhor se caracterize os diferentes perfis de comportamento.

Na presente etapa do projeto trabalhou-se na estruturação de um novo código computacional, mais flexível e robusto, para a realização das simulações e em modelar a existência de situações perigosas (DS – Dangerous Situations) no tráfego, que podem conduzir à existência de acidentes nas vias. No enfoque aqui utilizado tenta-se avaliar, além da velocidade máxima da via, se e como os diferentes perfis de motoristas podem afetar na possibilidade de existência dessas situações perigosas.

## 2- Introdução

A sociedade moderna exige mais, a cada dia, condições de mobilidade o que tem levado a graves problemas de engarrafamento na maioria das grandes cidades ao redor do mundo. A possibilidade de se fazer simulações do tráfego, utilizando-se modelos matemáticos, pode contribuir de forma fundamental, para previsão e planejamento de seu fluxo.

Nesse sentido, a descrição matemática do fluxo de tráfego, nas mais diversas situações, tem sido um tema amplamente estudado utilizando-se, principalmente modelos microscópicos. Conhecidos softwares existentes, como o AIMSUN (<https://www.aimsun.com>) e o PTV VISSIM (<http://vision-traffic.ptvgroup.com/en-us/products/ptv-vissim/>), utilizam esse tipo de modelagem. Entre os modelos microscópicos, o de Autômatos Celulares (AC) têm sido amplamente utilizados devido à sua flexibilidade para modelar situações complexas observadas no tráfego real. O primeiro modelo probabilístico para descrever o tráfego viário foi o modelo Nagel-Schreckenberg (NaSch) [1]. Com base nesse modelo muitos outros modelos de CA foram propostos, tanto para melhorar os resultados obtidos pelo modelo NaSch, como para investigar várias outras situações de interesse.

Este projeto teve o intuito de dar continuidade ao estudo de modelos de simulação numérica e computacional do tráfego viário que vem sendo desenvolvido ao longo desse projeto, tanto trabalhando-se no modelo proposto pelo grupo de pesquisa [2], como avaliando situações de interesse. Neste relatório será apresentada a descrição do trabalho que foi realizado até o momento. Quanto ao modelo proposto pelo grupo, descreveremos o CATS, simulador que vem sendo desenvolvido, e também descreveremos a análise que foi feita sobre a possibilidade de

acidentes no tráfego, avaliando a influência dos diversos perfis de direção nessa situação. Resultados desse trabalho geraram um artigo em congresso internacional, a ser realizado no presente mês. Finalmente concluiremos fazendo comentários sobre a continuidade do presente projeto.

### 3- Metodologia

Para dar seguimento ao projeto, concluímos que o programa que vinha sendo utilizado, resultado de uma tese no tema, não estava cumprindo com os novos desafios que se pretendia trabalhar. Dessa forma, deixamos de lado o programa que estava sendo utilizado para as simulações e foi desenvolvido um novo, com nova arquitetura e funcionalidades, o CATS.

Com relação a situações de interesse, foi proposta uma modificação no modelo NaSch, incluindo diferentes políticas de aceleração, e foi realizado um estudo referente a acidentes e novas métricas foram implementadas na simulação para verificar o comportamento do modelo proposto com relação a diversos tipos de acidentes.

#### 2.1 – Quanto ao CATS- Cellular Automata Traffic Simulator

O problema do simulador legado e a criação de um novo simulador: Durante o período do projeto anterior, todas as simulações foram rodadas utilizando o programa desenvolvido por Marcelo Zamith em C++ em sua tese de doutorado. Como era necessário realizar várias mudanças nos modelos e extrair cada vez mais informações, o programa não estava atendendo pois não possuía muita documentação dificultando a implementação e melhorias e novas funcionalidades. Diante desse problema optou-se por se desenvolver o CATS (Cellular Automata Traffic Simulator).

Inicialmente chamado de TUFFEngine, tinha o intuito de simular o modelo TUFF, modelo proposto pelo grupo e foco da pesquisa. O programa foi desenvolvido com base nos conceitos do programa original e nos conceitos do CAEngine (<https://github.com/gvpm/caengine>), uma biblioteca java genérica para simular autômatos celulares, desenvolvido pelo bolsista como projeto final do intercâmbio pelo CSF (Ciência sem Fronteiras) na Goldsmiths University Of London (09/2015-09/2016).

O sistema desenvolvido foi baseado em programação orientada a objetos, bastante modular, programado de forma desacoplada e genérica, tornando fácil a implementação de novos modelos propostos. Assim, o nome do programa deixou de ser TUFFEngine e passou a se chamar CATS. O programa também é *multithreaded* e paraleliza a sua execução de acordo com o número de núcleos no processador da máquina em que está rodando, tornando a execução mais rápida.

A funcionalidade mais recente que ainda não foi documentada, pois está em testes, é a possibilidade de rodar o programa em um servidor remoto. Este servidor ficará aguardando novas mensagens em um canal de um servidor Redis [2]. Redis é um banco de dados popular gratuito e de código aberto que armazena conteúdo no formato Chave-Valor e possui esta funcionalidade de entrega de mensagens via canal que utilizamos. O servidor aguarda que qualquer aplicação publique (envie) uma configuração com os parâmetros de uma simulação em determinado canal no Redis, configuração essa no formato JSON (JavaScript Object Notation) [3]. O formato JSON foi escolhido por ser uma notação padrão muito utilizada para troca de mensagens entre diferentes aplicações. Quando uma configuração é publicada, o servidor automaticamente recebe a mensagem, cria uma pasta, e coloca a simulação para rodar remotamente, salvando na pasta todos as saídas da simulação e o relatório de execução da

simulação onde ficam registrados eventuais erros. Esta funcionalidade é útil, pois como as simulações costumam durar algumas horas, tudo é feito no servidor, não tendo a necessidade de se deixar o computador ligado e também é possível iniciar simulações com mais rapidez e de qualquer lugar, do celular por exemplo.

O programa possui código aberto, com código e documentação escritos em inglês e seu repositório é mantido no github (<https://github.com/gvpm/CATS>). Sendo o programa aberto e disponível a todos, é possível que novos resultados publicados sejam validados por outros pesquisadores e aqueles que tiverem pesquisas nesse ramo podem utilizar e contribuir para o desenvolvimento do programa.

## **2.2 Quanto ao estudo de Acidentes**

O modelo base dessa nossa pesquisa, e da maioria de pesquisas desse ramo, é o modelo Nasch, criado por Nagel–Schreckenberg [1] no início dos anos 90. Assim, o primeiro modelo de AC, incluindo situações perigosas (DS) no modelo NaSch, foi proposto por Boccara et al. [5]. Huang et al. [6] apresentaram expressões analíticas para acidentes para tráfego com baixas velocidades máximas e Fukui et al [7] para altas velocidades. Modelo de CA de duas pistas também foram considerados para situações simplificadas. Moussa [8] define as situações perigosas devido tanto a duas situações: carros que param de se movimentar e quando há uma desaceleração brusca. Em [9] Moussa analisa a ocorrência de acidentes em uma pista de dois sentidos. Em [10] foi feito um estudo de acidentes devido a carros parados e sua consequência no fluxo de tráfego. Mais recentemente Bentaleb et al. [11] estudam, numericamente, a probabilidade de ocorrência de acidentes de carro no modelo NaSch, considerando veículos rápidos e lentos. Também analisa o efeito da evacuação de veículos danificados da estrada. Os efeitos da velocidade máxima dos carros em situações de perigo com modelos CA foram avaliados em Huang [12]. Pode-se notar [6, 7, 11] que para modelar heterogeneidade nos perfis de motoristas nas vias, considera-se somente a influência da velocidade máxima na existência de possíveis acidentes. Como no presente projeto tem-se tentado modelar os comportamentos dos motoristas, pretendeu-se avaliar se esses diferentes perfis podem influenciar no surgimento de situações perigosas que podem conduzir a acidentes no tráfego e como essa influência ocorre. Assim, para avaliar o surgimento dessas situações perigosas, fizemos modificações no modelo NaSch para incluir diferentes políticas de aceleração. Para isso utilizamos uma discretização mais refinada da pista, fazendo com que um carro ocupasse mais de uma célula, tornando assim possível representar tanto uma aceleração suave como uma aceleração brusca. Além disso, para determinar a política de aceleração do carro a cada instante de tempo, foram utilizadas Funções de Densidade de Probabilidade Beta, que vem sendo utilizada nesta pesquisa para descrever comportamentos através de diferentes políticas de aceleração.

Foram definidas e implementadas diversas métricas para medir a probabilidade de acidente e diversas simulações foram executadas para diferentes perfis de direção, com o intuito de analisar o comportamento deste perfis de condução com relação aos acidentes.

## **4- Resultados e discussão**

Alguns dos resultados gerados por essa pesquisa que avaliam situações perigosas (Dangerous Situations – DS) que podem conduzir a acidentes estão descritos no Position Paper aceito para o 15th International Conference on Scientific Computing, CSC'17, a ser realizado de 17 a 20 de julho do presente ano, 2017, em Las Vegas, USA. O paper intitulado “Traffic Flow Simulation Under the Influence of Different Acceleration Policies” foi aceito para publicação e apresentação

oral, obtendo excelentes avaliações. Assim, para a apresentação dos resultados, anexamos o paper, que estará disponível *on line* a partir de agosto.

Esses resultados comprovam a importância de se incluir a modelagem dos perfis de motoristas ao se avaliar as mais diversas situações do tráfego viário. No caso em questão, consideramos, através de FDP's adequadas, perfis de motoristas (i) que têm uma política de aceleração e frenagem, mais contínua e suave, (ii) que aceleram e freiam bruscamente, (iii) um perfil intermediário. Os resultados mostram que, além da velocidade máxima da via, como analisado por alguns autores e descrito na seção 3, essas diferentes políticas de aceleração podem atuar consideravelmente na dinâmica do fluxo viário e também na possível ocorrência de acidentes nas vias.

## 5- Conclusões e continuidade do trabalho

Como descrito na seção anterior, os resultados que têm sido encontrados no desenvolvimento do presente projeto mostram a importância de se modelar diferentes perfis na direção para tornar possível uma representação mais realista do tráfego viário. Nessa etapa do projeto, estudou-se propostas existentes na literatura para reconhecimento de situações perigosas do tráfego viário e propôs-se uma nova forma de modelar essa questão incluindo, na modelagem, diferentes perfis de aceleração. Resultados mostraram a importância de se considerar esses diferentes comportamentos.

Para obtenção desses resultados, também trabalhamos em um novo software, o CATS. Considera-se que o programa CATS possibilitará que novas funcionalidades, sejam incluídas com maior facilidade e eficiência. Então, com a continuidade da pesquisa pretende-se também dar continuidade ao desenvolvimento do CATS.

Além disso, ao se fazer o estudo de acidentes, verificamos que outras situações devem ser avaliadas e modeladas. Essa questão já vem sendo trabalhada pois percebemos que com a inclusão de diferentes políticas de aceleração devemos modelar fenômenos que não acontecem com um perfil uniforme e discretização menos refinada, como usualmente utilizado. Espera-se dar continuidade a esse estudo, incluindo uma análise de sensibilidade para avaliar quais fatores influenciam mais na ocorrência de situações perigosas no tráfego. Acreditamos que com esses resultados, que são inéditos na área, esse trabalho possa ser submetido a um periódico,

Pretende-se também, na continuidade do trabalho, avaliar estas situações no modelo proposto pelo grupo [2], que é um modelo de antecipação, mais preciso, que poderá representar, de forma mais real, o que acontece na pista. Essa avaliação deve ser feita tanto para pistas simples como para pistas duplas, que incluirá novos desafios a serem modelados.

## 6- Referências bibliográficas

[1] K. Nagel and M. Schreckenberg, A cellular automaton model for freeway traffic, *Journal of physical I France*, Vol. 2, pp. 2221-2229, 1992.

[2] M. Zamith, R. C. P. Leal-Toledo, E. Clua, E. M. Toledo, G. V. P. Magalhães, A new stochastic cellular automata model for traffic flow simulation with drivers behavior prediction, *J. Comp. Science*, 9, 51- 56, 2015.

[3] Redis, <https://redis.io/> em 12 de dezembro de 2016.

[4] JSON - JavaScript Object Notation, <http://json.org/>, em 15 de janeiro de 2017.

[5] N Boccara, H Fuks and Q Zeng, Car accidents and number of stopped cars due to road blockage on a one-lane highway, *Journal Physics*, A 30, 3329, 1997

- [6] D. W. Huang, Exact results for car accidents in a traffic model, *Journal of Phy. A: Math. and General*, V. 31, 29, 1998.
- [7] M. Fukui and Y. Ishibashi, Traffic flow in 1D cellular automaton model including cars moving with high speed, *Journal of the Physical Society of Japan*, vol. 65, pp. 1868-1870, 1996.
- [8] N. Moussa, Car accidents in cellular automata models for one-lane traffic flow, *Phy. Review E* 68, 036127, 2003.
- [9] N. Moussa, Simulation study of traffic accidents in bidirectional traffic models, *Int. J. Mod. Phys. C*, 21, 1501, 2010.
- [10] Y. Xiang-Qing, M. Y. Qiang, Car accidents determined by stopped cars and traffic flow, *J.l of Phy. A: Math.l and General*, 35, 49, 2002.
- [11] K. Bentaleb., N. Lakouari, R. Marzoug, A. Benyoussef, Simulation study of traffic car accidents in single lane highway, *Physica A* 413, 473-480, 2014.
- [12] D. W. Huangm W. C. Tseng, Mean-field theory for car accidents, *Phys. Rev. E* 64, 057106, 2001.

---

*Guilherme Villar Piccinini de Magalhães*  
(bolsista)

---

*Elson Magalhães Toledo – Regina Célia P. Leal Toledo*  
(orientadores)



# PROJETO DE INICIAÇÃO CIENTÍFICA

## Relatório

06/06/2017 à 05/07/2018

### Aplicações de Filtro de Kalman em Robótica

**Aluno:** Hiago Riba Guedes

**Curso:** Engenharia Mecatrônica – Universidade Católica de Petrópolis

**Orientador:** Marcelo D. Fragoso.

**Área:** Engenharia Elétrica

**Sub-área:** Eletrônica Industrial, Sistemas e Controles Eletrônicos

#### Objetivos:

Aplicar técnicas da Teoria de Filtragem de Kalman para o problema de localização de robôs em ambientes internos.

#### Introdução:

O Filtro de Kalman foi desenvolvido na década de 60 e tem se revelado uma ferramenta importante para o tratamento de informações de sistemas sujeitos à incertezas. A teoria de filtros de Kalman tem sido aplicada em diversos problemas em temas diversos tais como sistemas de orientação, design digital, redes neurais, fusão de sensores e fusão de dados, visão computacional, finanças, problemas de rastreamento, em meteorologia (na correção de estimativas de temperatura), dentre outros.

O projeto tem como objetivo utilizar técnicas de filtragem linear no estudo de robôs móveis. Para isso, será montada uma interface gráfica onde será aplicado em um ambiente virtual, o que foi estudado. Essencialmente, a teoria de filtragem será utilizada no problema de tracking e de localização de robôs, onde será feita uma comparação com a aplicação da filtragem e sem a aplicação.

Posteriormente podendo montar um pequeno robô para a representação real dele.

#### **Metodologia:**

Serão utilizadas técnicas de estimação linear (Filtro de Kalman) para resolver o problema de localização dos robôs.

#### **Resultados e Discussão:**

Como o Projeto foi aprovado em 06/06/2017 (27 dias) não temos ainda resultados. Estamos na primeira fase do Projeto referente ao "Estudo sobre a Teoria de Filtros de Kalman", que será utilizada como ferramenta no problema de **localização e estimação de robôs em ambientes internos**.

#### **Referências Bibliográficas:**

[1] Wiener, Norbert (1949) - *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. New York: Wiley.

[2] Kalman, R.E. (1960) A new approach to linear filtering and prediction problems, *Transactions of the ASME-Journal of Basic Engineering*. **82** (1): 35–45.

[3] Davis, M.H.A. (1977) – *Linear Estimation and Stochastic Control*, Chapman and Hall.

[4] Davis, M.H.A. e Vinter, R.B. (1985) - *Stochastic Modelling and Control*, Springer Netherlands

[5] Mataric, M.J. (1969) - *Introdução à Robótica*, Edgard Blucher

Relatório de atividades:

## *Métodos de Discretização para Sistemas Físicos Dispersivos e/ou Dissipativos*

*Aluno: Igor P. dos Santos Pereira*

Orientadora: Sandra M. Cardoso Malta

Co-orientadora: Cristiane Oliveira de Faria

*14 de Julho de 2017*

## 1 Objetivos

Propor esquemas de diferenças finitas para a equação de Korteweg–de Vries e outra equação afim, verificar suas consistências e convergências, além de estudar e analisar a estabilidade numérica de um método explícito para essas EDPs.

## 2 Metodologia

A equação de *Korteweg–de Vries* (KdV) modela fenômenos caracterizados pela propagação de pulsos ondulatórios. Este fenômeno foi inicialmente estudado em longos canais rasos, onde as ondas de água podiam percorrer longas distâncias sem ter sua amplitude alterada. Estas “ondas solitárias”, ou *sólitons*, como ficaram conhecidas, são capazes de interagir entre si de forma não linear e, após a colisão, preservar suas formas originais [1].

A KdV é uma equação diferencial parcial de terceira ordem, não linear, dispersiva, não-dissipativa [2] e homogênea, que pode ser escrita como:

$$\frac{\partial u}{\partial t} + \frac{\partial^3 u}{\partial x^3} + u \frac{\partial u}{\partial x} = 0, \quad (1)$$

sendo  $u$  uma função definida da seguinte maneira

$$u : \begin{array}{l} \Omega \subset \mathbb{R} \times \mathbb{R}^+ \longrightarrow \mathbb{R} \\ (x, t) \longmapsto u(x, t) \end{array} ; u \in \mathcal{C}^3([0, L] \times [0, \infty[).$$

Fisicamente,  $u(x, t)$  é o deslocamento vertical do ponto  $x$  do domínio da onda no instante de tempo  $t$ . O comprimento  $L$  representa o tamanho do canal. Além das características da KdV já citadas, a dependência temporal na primeira parcela da equação (1) a classifica também como um problema de evolução.

A presença de uma não-linearidade nesta equação demanda um tratamento mais específico e detalhado. Por isso, como primeira abordagem é conveniente definir e examinar uma versão modificada da KdV, em que o termo  $uu_x$  é desconsiderado a fim de se entender melhor o comportamento da derivada terceira:

$$u_t + u_{xxx} = 0 \quad \text{em } \Omega. \quad (2)$$

Utilizando um esquema de derivada central para aproximar a derivada de terceira ordem e o método de Euler explícito, é possível escrever a seguinte equação de diferenças finitas:

$$\frac{\mathbf{u}_i^{n+1} - \mathbf{u}_i^n}{\Delta t} + \frac{\mathbf{u}_{i+2}^n - 2\mathbf{u}_{i+1}^n + 2\mathbf{u}_{i-1}^n - \mathbf{u}_{i-2}^n}{2\Delta x^3} = 0, \quad (3)$$

onde  $\mathbf{u}_i^n$  denota a solução numérica num ponto  $(x_i, t_n) \in \Omega$  da malha, construída de forma homogênea a manter  $\Delta t$  e  $\Delta x$  constantes ao longo de toda malha.

Para verificar a consistência do esquema foi calculado o seu erro de truncamento local.

**Definição 2.1** (*Erro de truncamento local*): Sejam  $\mathbf{u}_i^{n+1}$  e  $\tilde{u}(x_i, t_{n+1})$  os valores da solução numérica e exata, respectivamente, no ponto  $(x_i, t_{n+1}) \in \Omega$ , então o erro de truncamento local (ETL) do esquema no ponto  $(x_i, t_n) \in \Omega$  é denotado e definido por:

$$\tau_i^n = \frac{\tilde{u}(x_i, t_{n+1}) - \mathbf{u}_i^{n+1}}{\Delta t}.$$

O esquema (3) será consistente com a equação original (2) se

$$\lim_{(\Delta x, \Delta t) \rightarrow (0,0)} \tau_i^n = 0.$$

Já a convergência do método foi estudada a partir da avaliação do erro global do esquema

**Definição 2.2** (*Erro global*): Sejam  $\mathbf{u}_i^n$  e  $\tilde{u}_i^n$  os valores da solução numérica e exata, respectivamente, no ponto  $(x_i, t_n) \in \Omega$ , então o erro global do esquema no ponto  $(x_i, t_n) \in \Omega$  é denotado e definido por:

$$\varepsilon_i^n = \mathbf{u}_i^n - \tilde{u}_i^n.$$

O intuito dessa análise é descobrir se o erro global do método pode, de alguma forma, ter seu valor absoluto limitado. Nesse caso, dizemos que o erro global do esquema é uma função limitada se  $\exists M > 0$  tal que

$$|\varepsilon_i^n| \leq M; \quad \forall (x_i, t_i) \in \Omega.$$

Uma forma para investigar se o erro global possui tal propriedade em todos os pontos da malha é, em particular, averiguar se este é limitado no ponto em que erro assume o seu maior valor. Caso o erro seja limitado nesse ponto, então, necessariamente, será limitado em todos os demais pontos da malha. Portanto, é oportuno o uso da norma do máximo para o erro global.

**Definição 2.3** (*Norma do máximo*): Seja  $\varepsilon^n = (\varepsilon_0^n, \varepsilon_1^n, \dots, \varepsilon_p^n) \in \mathbb{R}^{p+1}$ , onde  $(\mathbb{R}^{p+1}, \|\cdot\|_\infty)$ , então a norma do máximo de  $\varepsilon^n$  é denotada e definida como:

$$E^n = \max_{0 \leq i \leq p} |\varepsilon_i^n| = \|\varepsilon^n\|_\infty$$

Por (2) se tratar de uma EDP de evolução, é necessário fornecer uma condição inicial para garantir que o problema esteja “bem posto”. Além disso, a presença de uma condição inicial é fundamental para a avaliação do erro global que será proposta.

$$\begin{cases} \frac{\partial u}{\partial t} + \frac{\partial^3 u}{\partial x^3} = 0 & \text{em } \Omega, \quad u \in C^3(\Omega) \\ u(x, 0) = g(x) \end{cases}$$

Finalmente, o esquema explícito para a KdV a ser analisado é bem similar a (3), mantendo a mesma escolha para a derivada temporal e de terceira ordem. Já o operador diferencial contido no termo não-linear será aproximado por uma derivada central:

$$\frac{\mathbf{u}_i^{n+1} - \mathbf{u}_i^n}{\Delta t} + \frac{\mathbf{u}_{i+2}^n - 2\mathbf{u}_{i+1}^n + 2\mathbf{u}_{i-1}^n - \mathbf{u}_{i-2}^n}{2\Delta x^3} + \mathbf{u}_i^n \frac{\mathbf{u}_{i+1}^n - \mathbf{u}_{i-1}^n}{2\Delta x} = 0, \quad (4)$$

O erro de truncamento local desse esquema será calculado e sua consistência demonstrada.

### 3 Resultados e Discussão

Como na definição de *erro de truncamento local* aparecem apenas termos no passo  $n + 1$ , é conveniente isolar  $\mathbf{u}_i^{n+1}$  da equação (3). Uma vez que o método é explícito, isso pode ser feito de forma trivial:

$$\mathbf{u}_i^{n+1} = \mathbf{u}_i^n - \frac{\Delta t}{2\Delta x^3} \left( \mathbf{u}_{i+2}^n - 2\mathbf{u}_{i+1}^n + 2\mathbf{u}_{i-1}^n - \mathbf{u}_{i-2}^n \right), \quad (5)$$

Logo, a expressão para o erro de truncamento local é:

$$\tau_i^n = \frac{1}{\Delta t} \left[ \tilde{u}(x_i, t_{n+1}) - \left( \mathbf{u}_i^n - \frac{\Delta t}{2\Delta x^3} \left( \mathbf{u}_{i+2}^n - 2\mathbf{u}_{i+1}^n + 2\mathbf{u}_{i-1}^n - \mathbf{u}_{i-2}^n \right) \right) \right].$$

Supondo que a solução numérica deva ser, de forma idealizada, igual a solução exata, segue-se que

$$\frac{\tilde{u}_i^{n+1} - \tilde{u}_i^n}{\Delta t} = \tau_i^n - \frac{\tilde{u}_{i+2}^n - 2\tilde{u}_{i+1}^n + 2\tilde{u}_{i-1}^n - \tilde{u}_{i-2}^n}{2\Delta x^3}. \quad (6)$$

Pode-se observar que o erro de truncamento local está medindo justamente o quão a solução exata da equação diferencial deixa de satisfazer o esquema de diferenças finitas [3].

Expandindo cada termo de (6) em série de Taylor em torno do ponto  $(x_i, t_n)$ , percebe-se que as parcelas que contém  $\tilde{u}_{i-1}^n$  e  $\tilde{u}_{i-2}^n$  apresentam conjugação de sinal em suas respectivas séries de potência. Por isso, é útil expressar cada expansão na coordenada espacial como a soma de duas séries distintas: uma compilando as potências pares e outra as ímpares.

$$\begin{aligned} \tilde{u}_i^{n+1} &= \sum_{j=0}^{\infty} \frac{\Delta t^j}{j!} \frac{\partial^j u}{\partial t^j} \Big|_i^n \\ \tilde{u}_{i+2}^n &= \sum_{k=0}^{\infty} \frac{(2\Delta x)^k}{k!} \frac{\partial^k u}{\partial x^k} \Big|_i^n = \underbrace{\sum_{k=0}^{\infty} \frac{(2\Delta x)^{2k}}{(2k)!} \frac{\partial^{2k} u}{\partial x^{2k}} \Big|_i^n}_{\text{potências pares}} + \underbrace{\sum_{k=0}^{\infty} \frac{(2\Delta x)^{2k+1}}{(2k+1)!} \frac{\partial^{2k+1} u}{\partial x^{2k+1}} \Big|_i^n}_{\text{potências ímpares}} \\ \tilde{u}_{i+1}^n &= \sum_{k=0}^{\infty} \frac{\Delta x^k}{k!} \frac{\partial^k u}{\partial x^k} \Big|_i^n = \sum_{k=0}^{\infty} \frac{\Delta x^{2k}}{(2k)!} \frac{\partial^{2k} u}{\partial x^{2k}} \Big|_i^n + \sum_{k=0}^{\infty} \frac{\Delta x^{2k+1}}{(2k+1)!} \frac{\partial^{2k+1} u}{\partial x^{2k+1}} \Big|_i^n \\ \tilde{u}_{i-1}^n &= \sum_{k=0}^{\infty} \frac{(-\Delta x)^k}{k!} \frac{\partial^k u}{\partial x^k} \Big|_i^n = \sum_{k=0}^{\infty} \frac{\Delta x^{2k}}{(2k)!} \frac{\partial^{2k} u}{\partial x^{2k}} \Big|_i^n - \sum_{k=0}^{\infty} \frac{\Delta x^{2k+1}}{(2k+1)!} \frac{\partial^{2k+1} u}{\partial x^{2k+1}} \Big|_i^n \\ \tilde{u}_{i-2}^n &= \sum_{k=0}^{\infty} \frac{(-2\Delta x)^k}{k!} \frac{\partial^k u}{\partial x^k} \Big|_i^n = \underbrace{\sum_{k=0}^{\infty} \frac{(2\Delta x)^{2k}}{(2k)!} \frac{\partial^{2k} u}{\partial x^{2k}} \Big|_i^n}_{\text{potências pares}} - \underbrace{\sum_{k=0}^{\infty} \frac{(2\Delta x)^{2k+1}}{(2k+1)!} \frac{\partial^{2k+1} u}{\partial x^{2k+1}} \Big|_i^n}_{\text{potências ímpares}} \end{aligned}$$

Ao fazer essa distinção, a manipulação algébrica se torna mais clara, onde os termos idênticos, que apenas possuem sinais opostos podem ser cancelados. Caso a equação de diferenças finitas contasse com nenhuma parcela em que a função fosse avaliada em um ponto da forma  $(x_i - \lambda\Delta x, t_n)$ ,  $\lambda > 0$ , esse procedimento não seria necessário, visto que todos os monômios teriam mesmo sinal e, assim, a aritmética entre eles seria imediata.

Substituindo em (6) e realizando as devidas simplificações chegamos que,

$$\tau_i^n = \frac{\partial u}{\partial t} \Big|_i^n + \frac{\partial^3 u}{\partial x^3} \Big|_i^n + \sum_{j=2}^{\infty} \frac{\Delta t^{j-1}}{j!} \frac{\partial^j u}{\partial t^j} \Big|_i^n + \sum_{k=2}^{\infty} (2^\mu - 2) \frac{\Delta x^{\mu-3}}{\mu!} \frac{\partial^\mu u}{\partial x^\mu} \Big|_i^n, \quad (7)$$

onde  $\mu(k) = 2k + 1$ .

Como por hipótese, a função  $u$  é solução da equação (2) em  $\Omega$ , então, é necessário que, em particular, seja solução no ponto  $(x_i, t_n) \in \Omega$ :

$$\frac{\partial u}{\partial t} \Big|_i^n + \frac{\partial^3 u}{\partial x^3} \Big|_i^n = 0$$

Extraindo dos somatórios os termos de ordem mais baixa, obtemos o erro de truncamento local do esquema:

$$\begin{aligned}\tau_i^n &= \frac{\Delta t}{4} \frac{\partial^2 u}{\partial t^2} \Big|_i^n + \frac{\Delta x^2}{4} \frac{\partial^5 u}{\partial x^5} \Big|_i^n + \sum_{j=3}^{\infty} \frac{\Delta t^{j-1}}{j!} \frac{\partial^j u}{\partial t^j} \Big|_i^n + \sum_{k=3}^{\infty} (2^k - 2) \frac{\Delta x^{k-3}}{\mu!} \frac{\partial^k u}{\partial x^k} \Big|_i^n \\ \tau_i^n &= \frac{\Delta t}{4} \frac{\partial^2 u}{\partial t^2} \Big|_i^n + \frac{\Delta x^2}{4} \frac{\partial^5 u}{\partial x^5} \Big|_i^n + O(\Delta t^2 + \Delta x^4)\end{aligned}\quad (8)$$

O erro, portanto, é da ordem de  $O(\Delta t + \Delta x^2)$  e o esquema de diferenças finitas é consistente, já que, como se pode perceber facilmente, no limite, quando  $\Delta x$  e  $\Delta t$  vão a zero, o erro de truncamento local tende a se anular sob certas condições. Mais precisamente, é possível garantir a convergência do ETL se as derivadas  $u_{tt}$  e  $\partial_x^5 u$  forem limitadas em todo domínio  $\Omega$ . Por outro lado, se não for possível requerer essa regularidade para tais funções, pode ser, então, que o ELT apresente um comportamento divergente. No entanto, como será visto adiante, a hipótese de regularidade das derivadas se fará necessária para o prosseguimento da análise.

$$\left| \frac{\partial^2 u}{\partial t^2} \right| \leq A \quad \text{e} \quad \left| \frac{\partial^5 u}{\partial x^5} \right| \leq B, \quad A, B \in \mathbb{R}^+.$$

Para avaliar o erro global do esquema, começamos substituindo (5) e (6) na definição 2.2:

$$\begin{aligned}\varepsilon_i^n &= [\mathbf{u}_i^{n-1} + \eta (\mathbf{u}_{i+2}^{n-1} - 2 \mathbf{u}_{i+1}^{n-1} + 2 \mathbf{u}_{i-1}^{n-1} - \mathbf{u}_{i-2}^{n-1})] - [\tilde{u}_i^{n-1} + \eta (\tilde{u}_{i+2}^{n-1} - 2 \tilde{u}_{i+1}^{n-1} + 2 \tilde{u}_{i-1}^{n-1} - \tilde{u}_{i-2}^{n-1})] + \Delta t \tau_i^{n-1} \\ \varepsilon_i^n &= (\mathbf{u}_i^{n-1} - \tilde{u}_i^{n-1}) + \eta [(\mathbf{u}_{i+2}^{n-1} - \tilde{u}_{i+2}^{n-1}) - 2(\mathbf{u}_{i+1}^{n-1} - \tilde{u}_{i+1}^{n-1}) + 2(\mathbf{u}_{i-1}^{n-1} - \tilde{u}_{i-1}^{n-1}) - (\mathbf{u}_{i-2}^{n-1} - \tilde{u}_{i-2}^{n-1})] - \Delta t \tau_i^{n-1} \\ \text{onde } \eta &= -\frac{\Delta t}{2\Delta x^3}.\end{aligned}$$

As subtrações entre os parênteses definem o erro global calculado em pontos variados da malha:

$$\varepsilon_i^n = \varepsilon_i^{n-1} + \eta (\varepsilon_{i+2}^{n-1} - 2 \varepsilon_{i+1}^{n-1} + 2 \varepsilon_{i-1}^{n-1} - \varepsilon_{i-2}^{n-1}) - \Delta t \tau_i^{n-1} \quad (9)$$

Aplicando a norma do máximo ao erro e utilizando a desigualdade triangular é possível escrever a seguinte inequação:

$$\begin{aligned}E^n &\leq E^{n-1} - 6\eta E^{n-1} + \Delta t \mathfrak{T}^{n-1} \\ E^n &\leq (1 - 6\eta)E^{n-1} + \Delta t \mathfrak{T}^{n-1}, \quad \forall n \leq N\end{aligned}\quad (10)$$

onde,  $\mathfrak{T}^{n-1} = \max_{0 \leq p \leq I} |\tau_p^{n-1}|$ , denota a norma do máximo para o erro de truncamento local do esquema.

Quando  $n = 1$ , a desigualdade (10) assume a forma

$$E^1 \leq (1 - 6\eta)E^0 + \Delta t \mathfrak{T}^0$$

Uma vez que  $E^0$  corresponde ao maior erro global ao longo da malha no tempo  $t_0 = 0$  e, nesse instante, a solução numérica  $\mathbf{u}_i^0$  é dada pela condição inicial do problema,  $g(x_i)$ ; a solução exata é idêntica a solução numérica, o que implica em  $E^0 = 0$ .

$$E^1 \leq \Delta t \mathfrak{T}^0 \quad (11)$$

Utilizando (11) e usando indução finita é possível mostrar que o maior erro global em cada passo segue uma relação de recorrência, que pode ser expressa da seguinte forma:

$$E^{n+1} \leq \Delta t \left( \sum_{k=0}^n (1 - 6\eta)^k \mathfrak{T}^{n-k} \right) \quad (12)$$

Nesse ponto, faz-se o uso da hipótese de limitação das derivadas  $u_{tt}$  e  $\partial_x^5 u$  para afirmar que:

$$|\tau_i^n| \leq A \frac{\Delta t}{4} + B \frac{\Delta x^2}{4}$$

Por conseguinte,

$$\begin{aligned} E^{n+1} &\leq \Delta t \left( \sum_{k=0}^n (1-6\eta)^k \tau^{n-k} \right) \leq \Delta t \left( \sum_{k=0}^n (1-6\eta)^k \left( A \frac{\Delta t}{4} + B \frac{\Delta x^2}{4} \right) \right) \\ E^{n+1} &\leq \Delta t \left( A \frac{\Delta t}{4} + B \frac{\Delta x^2}{4} \right) \left( \sum_{k=0}^n (1-6\eta)^k \right) \end{aligned}$$

O somatório entre parênteses corresponde a soma dos  $n$  primeiros termos de uma progressão geométrica de razão  $1 - 6\eta$  e, por isso, a desigualdade pode ser escrita da seguinte forma:

$$\begin{aligned} E^{n+1} &\leq \left( A \frac{\Delta t^2}{4} + B \frac{\Delta t \Delta x^2}{4} \right) \left( \frac{1 - (1-6\eta)^{n+1}}{6\eta} \right) \\ E^{n+1} &\leq \frac{1}{24\eta} \left( A \Delta t^2 + B \Delta t \Delta x^2 \right) \left( 1 - (1-6\eta)^{n+1} \right) \end{aligned} \quad (13)$$

O cálculo do erro de truncamento local para o esquema (4) para KdV se beneficia do fato de ser uma extensão do esquema (3) para a KdV modificada. Posto que todos os termos em comum entre as duas EDPs são aproximados pelo mesmo tipo de derivada numérica, este ETL difere de (7) apenas pelo termo não-linear:

$$\tau_i^n = \frac{\partial u}{\partial t} \Big|_i^n + \frac{\partial^3 u}{\partial x^3} \Big|_i^n + \sum_{j=2}^{\infty} \frac{\Delta t^{j-1}}{j!} \frac{\partial^j u}{\partial t^j} \Big|_i^n + \sum_{k=2}^{\infty} (2^\mu - 2) \frac{\Delta x^{\mu-3}}{\mu!} \frac{\partial^\mu u}{\partial x^\mu} \Big|_i^n + \tilde{u}_i^n \frac{\tilde{u}_{i+1}^n - \tilde{u}_{i-1}^n}{2\Delta x}$$

Como já foi feito para as primeiras parcelas, é necessário expandir em série de Taylor os termos da não-linearidade em torno do ponto  $(x_i, t_n)$ , obtendo:

$$\tau_i^n = \frac{\partial u}{\partial t} \Big|_i^n + \frac{\partial^3 u}{\partial x^3} \Big|_i^n + \sum_{j=2}^{\infty} \frac{\Delta t^{j-1}}{j!} \frac{\partial^j u}{\partial t^j} \Big|_i^n + \sum_{k=2}^{\infty} (2^\mu - 2) \frac{\Delta x^{\mu-3}}{\mu!} \frac{\partial^\mu u}{\partial x^\mu} \Big|_i^n + \left[ u_i^n \frac{\partial u}{\partial x} \Big|_i^n + \frac{\Delta x^2}{3!} u_i^n \frac{\partial^3 u}{\partial x^3} \Big|_i^n + u_i^n \sum_{k=2}^{\infty} \frac{\Delta x^{\mu-1}}{\mu!} \frac{\partial^\mu u}{\partial x^\mu} \Big|_i^n \right]$$

Como  $u$  é solução da KdV, deve satisfazer

$$\frac{\partial u}{\partial t} \Big|_i^n + \frac{\partial^3 u}{\partial x^3} \Big|_i^n + u_i^n \frac{\partial u}{\partial x} \Big|_i^n = 0$$

Segue que,

$$\tau_i^n = \frac{\Delta t}{4} \frac{\partial^2 u}{\partial t^2} \Big|_i^n + \frac{\Delta x^2}{4} \frac{\partial^5 u}{\partial x^5} \Big|_i^n + \frac{\Delta x^2}{6} u_i^n \frac{\partial^3 u}{\partial x^3} \Big|_i^n + \sum_{j=3}^{\infty} \frac{\Delta t^{j-1}}{j!} \frac{\partial^j u}{\partial t^j} \Big|_i^n + \sum_{k=3}^{\infty} (2^\mu - 2) \frac{\Delta x^{\mu-3}}{\mu!} \frac{\partial^\mu u}{\partial x^\mu} \Big|_i^n + u_i^n \sum_{k=2}^{\infty} \frac{\Delta x^{\mu-1}}{\mu!} \frac{\partial^\mu u}{\partial x^\mu} \Big|_i^n$$

Por fim, o erro de truncamento local do esquema é:

$$\tau_i^n = \frac{\Delta t}{4} \frac{\partial^2 u}{\partial t^2} \Big|_i^n + \frac{\Delta x^2}{4} \left( \frac{\partial^5 u}{\partial x^5} \Big|_i^n + u_i^n \frac{2}{3} \frac{\partial^3 u}{\partial x^3} \Big|_i^n \right) + O(\Delta t^2 + \Delta x^4) \quad (14)$$

Assim como no primeiro esquema trabalhado, no limite, quando  $(\Delta x, \Delta t) \rightarrow (0, 0)$ , o ETL tende a se anular, desde que  $u$ ,  $u_{tt}$ ,  $u_{xxx}$  e  $\partial_x^5 u$  sejam limitadas em  $\Omega$ . Por se tratar de um modelo ondulatório, é razoável esperar que a solução  $u$  seja “bem comportada”, isto é, que além de ser limitada (por uma amplitude máxima), tenha derivadas suficientes e que estas também sejam limitadas.

## 4 Conclusão

Os esquemas de diferenças finitas propostos para a KdV e sua versão linear apresentam consistência, contanto que algumas condições sobre o comportamento da solução sejam impostas. Em relação ao erro global dos esquemas, os resultados preliminares parecem ser inconclusivos, a princípio. Outras abordagens ainda serão experimentadas na tentativa de avaliar melhor a convergência desses métodos.

Por ser um EDP não linear, a teoria clássica de análise de estabilidade linear não pode ser empregada para a KdV. A proposta que pretende-se usar é a dos esquemas conservativos em diferenças finitas, que envolvem algumas sofisticacões as quais podem auxiliar no estudo da estabilidade numérica da KdV.

Assim, na sequência, os esquemas conservativos estão sendo estudados e a implementação numérica dos esquemas propostos está sendo desenvolvida.

## Referências

- [1] C. Constantino, Ondas não-lineares em meio dissipativos descritas pela Equação de Korteweg–de Vries–Burgers. Dissertação de Mestrado em Física, Unicamp, (1979).
- [2] J. Aminuddin, *Numerical Solution of The Korteweg–de Vries Equation*, *International Journal of Basic & Applied Sciences IJBAS-IJENS*, 11(2):62–67, 2011.
- [3] J. A. CUMINATO, Discretização de Equações diferenciais Parciais - Técnicas de Diferenças Finitas. USP. (2002).



## RELATÓRIO DE ATIVIDADES

**Título do Projeto:** Explorando Sistemas de Gerência de *Workflows* Científicos para Experimentos de Bioinformática em Ambientes HPC

**Nome do bolsista:** Joice Cristina Alves Gouveia Pedro

**Nome do orientador:** Kary Ann del Carmen Ocaña Gautherot

**Tipo de bolsa:** PIBIC

**Período do relatório:** 01/01/2017 a 10/07/2017

### OBJETIVOS

A proposta desta iniciação científica, para o presente período de avaliação, se centra na análise de desempenho em ambientes de processamento de alto desempenho (PAD) de ferramentas de bioinformática como o RAxML<sup>1</sup> e o Sistema de Gerência de *Workflows* Científicos (SGWfC) SciCumulus<sup>2</sup>. O ambiente usado para a execução dos experimentos em larga escala foi o supercomputador Santos Dumont<sup>3</sup>. O objetivo principal foi explorar esses ambientes para alocar de forma eficiente experimentos de bioinformática em larga escala, que levem a uma melhor compreensão do comportamento tanto das ferramentas como do ambiente de PAD. Desta maneira, pretende-se diminuir os possíveis gargalhos de custo e tempo computacional de experimentos em larga escala, a fim de dar suporte aos cientistas na obtenção de resultados em tempo hábil.

### INTRODUÇÃO

O uso de processamento de alto desempenho (PAD) [Foster and Kesselman 2004] é cada vez mais requerido nas diversas áreas das Ciências como biologia, física, astronomia, entre outros. Na bioinformática, devido ao crescimento exponencial de dados biológicos usados na investigação médica, biológica e biotecnológica, ela requer

---

<sup>1</sup> Randomized Axelerated Maximum Likelihood

<sup>2</sup> <https://scicumulus2.wordpress.com/>

<sup>3</sup> <http://sdumont.lncc.br/>

de tecnologias eficientes escaláveis. Pesquisas em bioinformática relacionadas a estudos sobre a evolução dos organismos *e.g.*, em estudos de genômica comparativa, filogenia e filogenômica fazem parte das áreas mais usadas, porém são as mais custosas computacionalmente [Delsuc et al. 2005].

O programa RAxML [Stamatakis 2014] é um dos poucos programas que possui um algoritmo escalável e paralelizável, e por isso, foi escolhido nessa primeira fase da presente pesquisa. Logo depois da exploração do RAxML, a seguinte fase a ser explorada (atualmente) visa acoplar eficientemente essa atividade por meio da construção de *workflows* científicos e sua instanciação em Sistemas de Gerência de *Workflows* Científicos (SGWfC) como o SciCumulus [Oliveira et al. 2010].

O RAxML constrói árvores para inferências filogenéticas utilizando dois métodos importantes, são eles a Máxima Verossimilhança (MV) e o *bootstrap*. Ambos métodos são computacionalmente intensivos, o que faz com que a execução de grandes dados biológicos (árvores da vida de organismos) em ambiente serial seja uma tarefa dispendiosa no quesito tempo, ou até inviável. O RAxML disponibiliza algoritmos paralelos como versões PThreads, MPI e uma versão híbrida MPI/PThreads, as quais foram exploradas no presente projeto.

Na primeira fase, objetivamos fazer uma análise de desempenho dessas quatro versões do RAxML com intuito de atribuir o comportamento em ambientes de PAD baseado no desempenho (tempo de execução/*speedup*) de cada uma das versões, demonstrando assim o ganho que as versões paralelas fornecem (versão híbrida foi a versão com melhor desempenho).

Fruto desses resultados levou a submissão e aceitação do seguinte resumo expandido na Escola Regional de Alto Desempenho<sup>4</sup> (ERAD-RJ 2017): “*Computação Paralela para a Inferência Filogenética*”. Joice Pedro (FAETEC/RJ), Marcelo Galheigo (LNCC), Carla Osthoff (LNCC), Roberto Pinto Souto (LNCC) e Kary Ocaña (LNCC). ERAD-RJ 2017. Com essas análises foi possível evidenciar alguns gargalos na aplicação que precisam de pesquisas mais detalhadas, através de ferramentas de perfilamento, que será tema de trabalhos futuros.

## **METODOLOGIA**

---

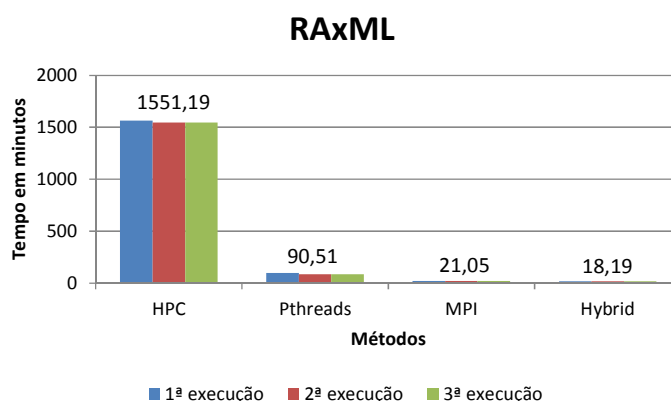
<sup>4</sup> <https://www.dcc.ufrj.br/erad2017/>

A primeira etapa envolveu uma revisão bibliográfica exaustiva sobre a ferramenta RAxML 8.2.9 utilizada para as execuções de inferências filogenéticas; o ambiente de PAD do SINAPAD, em especial o Santos Dumont; e um estudo sobre bibliotecas e *scripts* de execução específica dos *clusters* do SINAPAD.

Na segunda etapa iniciamos as execuções das quatro versões do RAxML (Serial, Pthreads, MPI e Híbrida) no Santos Dumont. Para avaliação de desempenho utilizamos dados biológicos reais, um superalinhamento de aminoácidos (4.941 bases) de 31 genes de 12 genomas de protozoários, o modelo evolutivo WAG e 1.000 de *bootstrap*.

## RESULTADOS E DISCUSSÃO

A Figura 1 apresenta a média de tempo de execução total (TET) das execuções das versões do RAxML. As execuções foram repetidas três vezes para confirmação dos resultados.



**Figura 1. Média de TET do RAxML.**

A Figura 1 demonstra que o ganho no tempo de execução da versão Pthread em relação à versão serial é bastante expressivo, chegando a ser 17 vezes mais rápido que a versão serial. O ganho continua com o uso da versão MPI onde a execução ficou 73 vezes mais rápida, ficando apenas aproximadamente três minutos acima da versão Híbrida que apresentou um ganho de 85 vezes, que diminuiu de 1.551,19 para 18,19 minutos (98,83%).

Pelo melhor desempenho a versão híbrida foi eleita para o teste de escalabilidade.

A Figura 2 apresenta o gráfico de escalabilidade para o RAxML versão híbrida. Os testes foram repetidos três vezes e mostraram similaridades nos resultados das suas execuções. A Figura 3 apresenta a média do TET dessas repetições com o valor de execução para cada núcleo computacional.

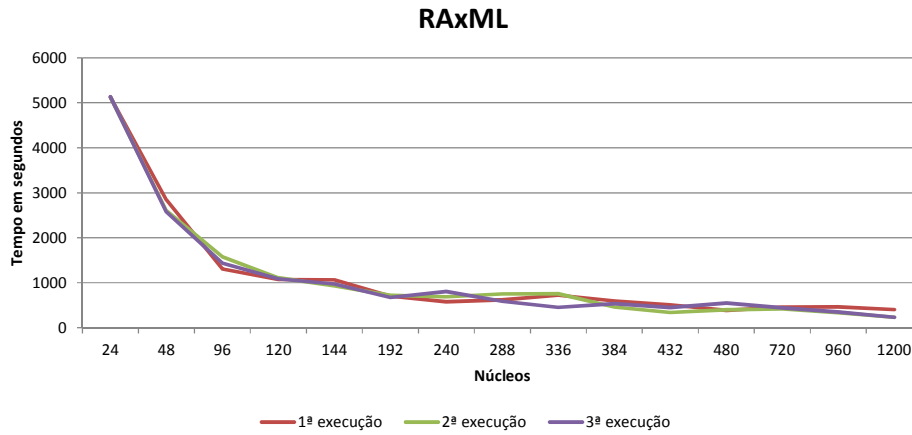


Figura 2. Três execuções do RAxML versão Híbrida.

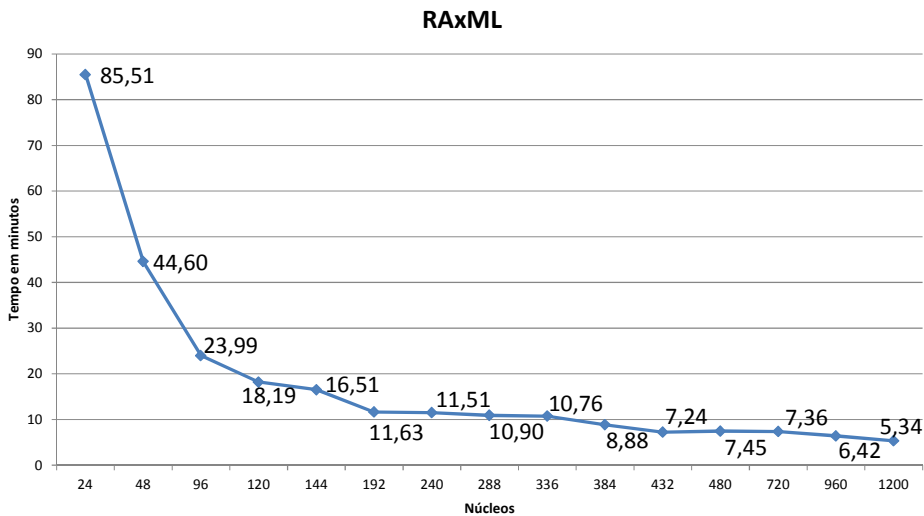


Figura 3. Média do TET da versão Híbrida.

A Figura 4 apresenta o gráfico de *speedup* do RAxML. Pode ser observado que em 120 núcleos há um ganho de 85 vezes. Depois disso, o RAxML versão híbrida apresentou uma redução na escalabilidade à medida que o número de núcleos aumenta, apresentando um ganho de 318 vezes em 1200 núcleos.

Essa redução pode estar relacionada à alocação de memória, afinidade de processos/núcleos, escrita em disco ou até mesmo características na arquitetura do *cluster*. Acreditamos que essa redução na escalabilidade seja menor no ExaML (a versão do RAxML para ambientes de PAD), uma vez que o algoritmo trata questões como a alocação de memória de forma mais refinada.

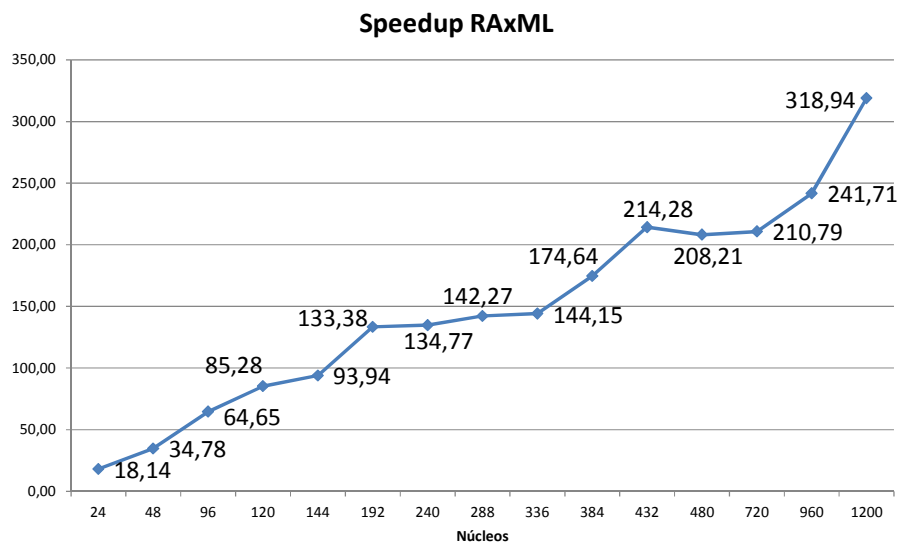


Figura 4. Speedup da versão Híbrida.

## CONCLUSÕES

Nos testes de desempenho, o RAxML versão híbrida apresentou uma melhora em até 98,83% comparada à execução serial. Depois disso, ele apresentou redução na escalabilidade lineal. A melhora apresentada da versão híbrida em relação à versão serial corrobora a importância do uso de ferramentas escaláveis em ambientes paralelos.

Análises mais detalhadas no Santos Dumont estão previstas, como o uso de ferramentas de tuning (*perfilador*), para explorar melhor execuções intensivas tanto do RAxML como do seu derivado, o ExaML<sup>5</sup>, específico para supercomputadores.

Alguns treinamentos foram necessários com o intuito de melhor entender os resultados e aprimorar os conhecimentos como Introdução a Programação MPI, Introdução a Programação CUDA, Introdução a Programação com *Swift: Workflows*

<sup>5</sup> Exascale Maximum Likelihood

Científicos Paralelos e Introdução a Programação Paralela de Aceleradores em OpenCL. Além dos cursos acima mencionados, participaremos da Escola Supercomputador SDUMONT para aprimorar os conhecimentos dos usuários no desenvolvimento de algoritmos paralelos.

## **REFERÊNCIAS BIBLIOGRÁFICAS**

Delsuc, F., Brinkmann, H. and Philippe, H. (may 2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews. Genetics*, v. 6, n. 5, p. 361–375.

Foster, I. and Kesselman, C. (2004). *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann.

Oliveira, D., Ogasawara, E., Baião, F. and Mattoso, M. (2010). SciCumulus: A Lightweight Cloud Middleware to Explore Many Task Computing Paradigm in Scientific Workflows. In *International Conference on Cloud Computing*. . <http://dx.doi.org/10.1109/CLOUD.2010.64>, [accessed on Nov 2].

Stamatakis, A. (1 may 2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, v. 30, n. 9, p. 1312–1313.



## Coordenação de Ciência da Computação (CCC)

Relatório

Programa Institucional de Bolsas de Iniciação Científica (IC)

# Sensoriamento Autônomo com Inteligência Baseada em Nuvens Computacionais

Jonathas Rodrigues Almeida

Orientador - Bruno Schulze

Período: Dezembro 2016 a Junho 2017

## 1 Dados Gerais

- **Instituição:** Laboratório Nacional de Computação Científica
- **Bolsista:** Jonathas Rodrigues Almeida
- **Coordenador/Orientador:** Bruno Schulze (CCC-LNCC)
- **Título:** Sensoriamento Autônomo com Inteligência Baseada em Nuvens Computacionais
- **Modalidade:** Bolsa de Iniciação Científica
- **Vigência:** 01/01/2017 a 31/07/2017
- **Projeto Relacionado:** Computação Científica Distribuída (ComCiDis)

## 2 Objetivos

Os seguintes objetivos foram estipulados para o plano de trabalho com o intuito de contribuir para o cenário de integração de *Internet of Things* (IoT) à Nuvens Computacionais, foram eles:

1. Montagem e teste do veículo;
2. Testes de comunicação dos sensores do veículo com a nuvem;
3. Análise e testes de algoritmos de sensoriamento e avaliações de rotas;
4. Teste do veículo autônomo e relatório;

## 3 Introdução

A proposta apresentada teve por finalidade dar continuidade ao uso de arquiteturas de processadores abertos em apoio à pesquisa científica. O uso de arquitetura ARM para a construção de sensores e dispositivos móveis não tripulados tem se tornado cada vez mais presente no mundo. Para a pesquisa científica, o uso destes dispositivos, datam desde os lançamentos das sondas espaciais para explorar espaço profundo, e das pesquisas submarinas nas profundezas abissais e hoje se encontram presentes para aplicação em locais extremamente nocivos ao seres humanos, como vulcões, plantas industriais com resíduos tóxicos, em desarme de bombas ou minas e sensores militares ou vetores de armamento.

Com a possibilidade de hoje poder montar veículos semelhantes utilizando processadores com capacidade de processamento de instruções mais complexas e sobretudo circuitos inteiros na forma de componentes de prateleira, fáceis de serem adquiridos, torna-se possível prospectar a possibilidade de conjugar a capacidade de processamento distribuído existente em nuvens computacionais, a mobilidade desses dispositivos, a comunicação remota e capacidade de injetar novas instruções ampliando a gama de atividade dos dispositivos em função do processamento de dados remoto. Isso em síntese possibilitaria dispositivos mais inteligentes e flexíveis que possuiriam a capacidade de processamento de uma nuvem computacional e que se adaptariam a situação em função do processamento e análise das informações





coletadas e enviadas para a nuvem. Essa abordagem se mostra bem atrativa no momento que a inteligência fica remota, em um ambiente com muito maior capacidade de processamento e dispositivos autônomos mais simples, poderiam adaptar-se em função da situação, resolvendo problemas específicos.

Tendo em vista que o grupo de Computação Científica Distribuída adquiriu um conjunto de placas para a construção de um sensor e um *cluster* baseado em arquitetura ARM, de forma a avaliar o uso destas arquiteturas na execução de aplicações científicas e seu respectivo consumo de energia em comparação as arquiteturas atuais. A continuidade desse trabalho era adquirir mais uma placa e os respectivos acessórios necessários a construção de um veículo autônomo, com a capacidade de executar o sensoriamento em uma área específica e em função dos dados obtidos, processados remotamente em nuvem, esse veículo autônomo, em tempo real, teria o conjunto de instruções alterado para realizar uma avaliação mais detalhada da discrepâncias observadas nessa área.

No plano de trabalho apresentado foi definido em duas etapas, sendo a primeira dedicada a aquisição e montagem do veículo autônomo e a segunda dedicada elaboração dos algoritmos para o processamento dos dados e a comunicação do veículo em Nuvem.

## 4 Material e Métodos ou Metodologia

Os resultados obtidos serão apresentados de acordo as metas, relacionando as atividades realizadas durante este período no cronograma.

### Meta 1 - Montagem e teste do veículo

O objetivo desta meta foi definir quais os materiais necessários para a construção do veículo autônomo e realizar a aquisição destes componentes. Através das avaliações foi definido o uso dos seguintes equipamentos de acordo com suas funções:

#### Arduíno R3

O Arduíno é uma plataforma de computação *open-source* baseado em uma placa com um microcontrolador Atmega328, com entradas e saídas digitais e analógicas, um ambiente de desenvolvimento baseado em Linguagem C através de uma comunicação USB. Esta plataforma é muito utilizada para criação de veículos autônomos por possuir uma infinidade de *Shields* (placas de hardware que podem ser plugadas no Arduino) com diversas aplicação como controle de motores, sensoriamento e comunicação Wi-Fi, facilidade de montagem tanto mecânica quanto eletrônica além de possuir um ambiente particularmente simples de programação.

#### Kit Tanque Tamiya

O kit tanque Tamiya é dividido em três conjuntos, o primeiro (Tamiya 70097) consiste em dois motores DC independentes com suas caixas de redução ajustável em duas relações de transmissão, 58:1 e 204:1, o segundo (Tamiya 70100) possui as duas partes da esteira juntamente com as rodas a se utilizar e os respectivos eixos e por último (Tamiya 70098) é o prato de montagem rígido com diversas furações que permite a acoplagem não só dos dois primeiros itens como também de outros componentes na construção do veículo autônomo. Este kit, por ser em baseado em forma de tanque, possui uma melhor aderência no deslocamento em diversos tipos de superfície, possui também a possibilidade, através das relações de transmissão, de termos um maior controle na velocidade do veículo além da facilidade na montagem e empilhamento dos componentes.

### **Driver Motor L298N**

Este Driver é baseado no chip L298N, construído para controlar cargas indutivas como relés, motores DC e motores de passo. Com o uso deste Driver é possível controlar independentemente a velocidade e rotação dos dois motores DC do veículo autônomo.

### **Sensor de Refletância QRE1113**

Como o veículo autônomo irá ter um deslocamento em um determinado ambiente para realizar algumas funções como sensoriamento, seria necessário possuímos uma espécie de rotas onde ele iria percorrer parada devidos fins. Para que o veículo saiba se deslocar em uma determinada trajetória utilizaremos quatro sensores analógicos de reflectância, que consiste em um LED infravermelho e um foto-transistor sensível à luz infravermelho, que permite de uma maneira fácil detectar uma faixa na superfície através da variação de luz infravermelho refletida.

### **Sensor Ultrassônico - HC-SR04**

Como o nome já diz, o sensor ultrassônico possui um emissor e um receptor sônico e com ele é possível calcular a distância de objetos através do tempo entre a emissão e a recepção de pulsos sônicos. A utilização do sensor ultrassônico foi por uma questão de segurança, tanto para quem se encontra por perto quanto para a proteção física do veículo, pois com ele podemos definir com será o comportamento do veículo quando houver algum obstáculo a sua frente.

### **Sensor de Temperatura e Umidade DHT22**

O DHT22 é um sensor de temperatura e umidade com comunicação digital de um fio, o sensor é calibrado e não requer componentes extras para funcionar, sendo somente necessário sua conexão ao Arduíno e sua programação. Com este sensor será necessário realizar testes de leitura e processamento de dados.

### **Módulo WiFi - ESP8266**

O módulo WiFi ESP8266 é um SOC com protocolo TCP/IP integrado que consegue dar a qualquer microcontrolador acesso a sua rede WiFi. O ESP8266 é capaz tanto de hospedar uma aplicação quanto descarregar todas as funções de redes WiFi a partir de outro processador de aplicação.

### **Componentes Diversos**

Além dos componentes citados, foi necessário a utilização de mais alguns componentes como uma bateria de Zipo de 11.1V 2200mAh para a alimentação dos componentes eletrônicos, uma placa de MDF para a fixação dos componentes, um botão ON/OFF para ligar e desligar o veículo, fios para a ligação dos sensores com a placa Arduíno, entre outros necessários para a montagem mecânica e eletrônica.

A próxima etapa realizada foi a elaboração dos circuitos, a montagem dos componentes mecânicos e eletrônicos e a realização de testes básicos a fim de definir o comportamento de cada um dos componentes para podermos atribuir suas funções de acordo com o objetivo desejado. Após a realização de testes de funcionamento envolvendo cada componente separadamente, definiu-se qual a pinagem do Arduíno seria utilizado por cada um deles e após isso realizar o esquema de ambos componentes em funcionamento.

O esquema a seguir, Figura 1, apresenta como o circuito deverá ser montado, já com os sensores de refletância, temperatura, ultrassônico, Módulo WiFi e o Driver de motor conectado a placa Arduíno. A partir deste esquema será possível a montagem de toda a parte eletrônica do veículo autônomo, onde o Arduíno irá coletar todos os dados dos sensores, através das portas digitais e analógicas, e enviar estes dados pela porta serial ou módulo Wifi para um determinado fim.

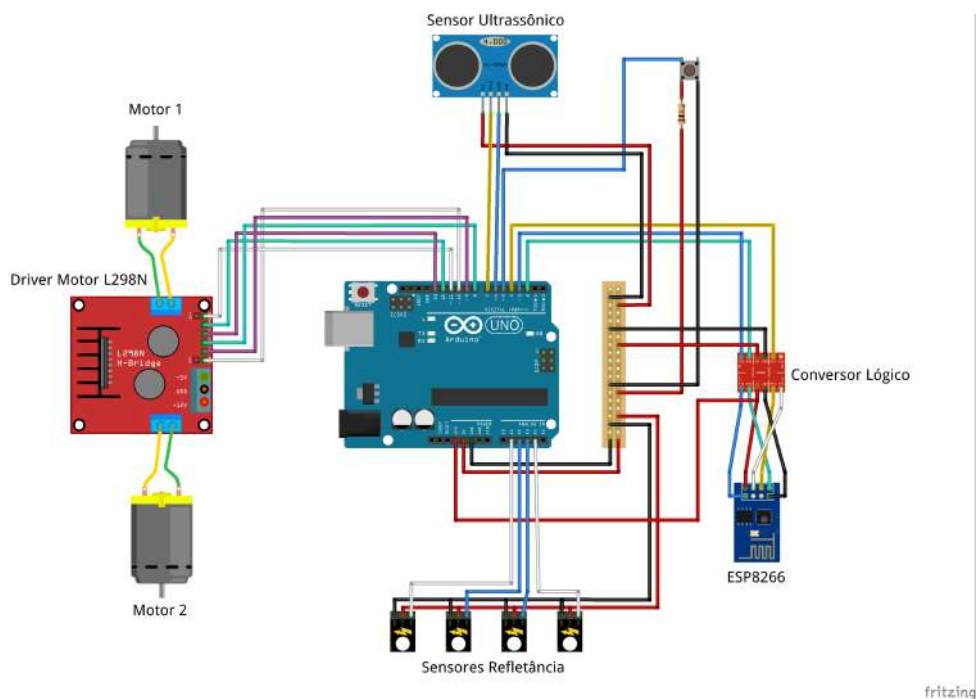


Figure 1: Esquema lógico para a montagem do circuito com Arduino

Em seguida foi feita a montagem dos componentes mecânicos do kit Tamiya composto por dois motores com suas respectivas caixas de redução (Tamiya 70097), as rodas com seus eixos e as esteiras (Tamiya 70100) e o prato de montagem (Tamiya 70098). Na figura 2 podemos ver como fica o chassi do veículo após a junção de ambos componentes.



Figure 2: Montagem do Kit Tamiya

A partir desse momento foi acoplado ao chassi do veículo uma placa de MDF para facilitar a fixação dos componentes eletrônicos ao chassi, nessa placa parafuso-se o Arduíno, o Driver de Motor e o sensor ultrassônico. Na parte inferior do chassi um suporte foi adicionado para a fixação dos quatro sensores de refletância e a bateria juntamente com o botão ON/OFF ficou localizada no interior do chassi como visto na Figura 3.

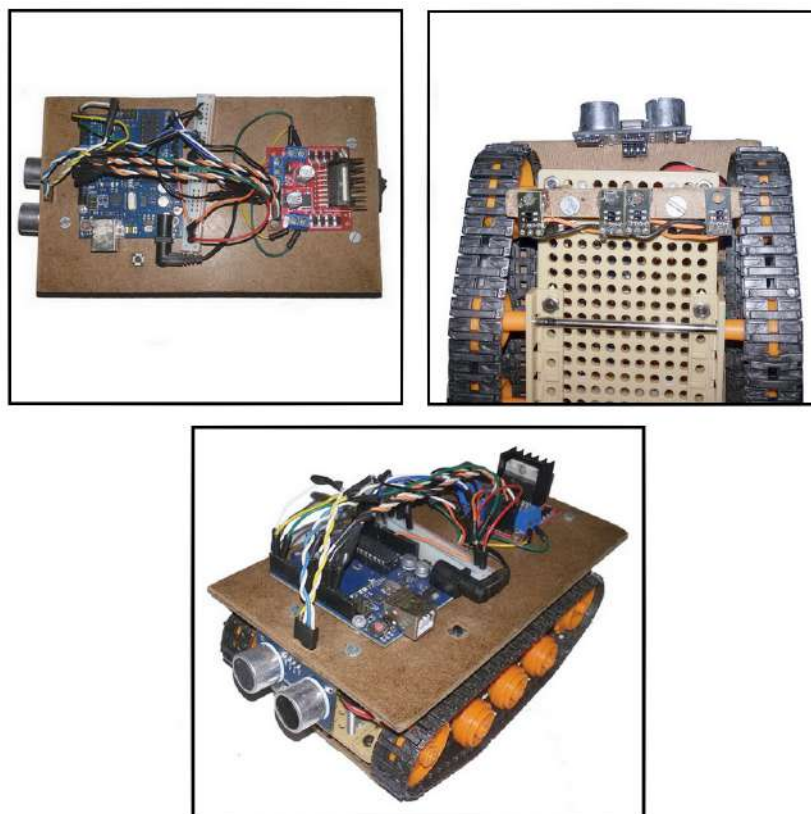


Figure 3: Montagem final do veículo autônomo

## Meta 2 - Testes de comunicação dos sensores do veículo com a nuvem

Objetivo dessa etapa é realizar o desenvolvimento do código para o estabelecimento da comunicação entre o veículo autônomo e a nuvem, a inicialização dos *apliances* e carregamento dos programas de controle do veículo e sensores.

Com base nos objetivos, foi desenvolvido um código para com que o veículo se conecte a uma rede *wireless* através do ESP8266, possibilitando assim a comunicação da mesma com a nuvem. Esse código ainda necessita de algumas adaptações e melhoramento para facilitar a interface de comunicação, afim de facilitar a interpretação por parte do Arduino dos dados a serem enviados pela nuvem.

A respeito do carregamento dos programas pelo veículo e tendo como objetivo o uso da nuvem para processamento de dados, foi definido que o veículo não ficará responsável por nenhuma função de reconhecimento de rotas e de seleção de funções. Sendo assim, a nuvem será responsável por ter o conhecimento das rotas e informar em que direção o veículo deve seguir, receber as "missões" a serem realizadas, processar qual a melhor trajetória e forma de realizá-las, além de realizar o tratamento dos da-

dos recebidos dos sensores implantados no veículo, enquanto isso o veículo terá em seu código apenas o necessário para, através dos comandos recebidos pela nuvem, realizar o seu deslocamento através das rotas, coletar dados sensoriais e realizar a comunicação.

### **Meta 3 - Análise e testes de algoritmos de sensoriamento e avaliações de rotas**

Nesta etapa, no que se refere aos sensores, serão realizados os testes dos algoritmos de sensoramento e envio dos dados coletados. Em relação ao veículo serão realizados experimentos com diferentes algoritmos para avaliação de rotas e tratamento de colisão.

Em um primeiro momento foi desenvolvido um algoritmo afim de analisar os dados das leituras do sensor de temperatura e umidade, evitando possíveis erros e leituras imprecisas. Além disso, sabendo que os sensores utilizados no veículo para a interpretação de sua posição na rota são suscetíveis a erros e com o objetivo de obter uma melhor performance, evitando erros e fugas das rotas foi desenvolvido e implementado um código conhecido como PID. Esse fica responsável por realizar o tratamento dos dados recebido pelo Arduino dos sensores de refletância centrais (responsáveis por manter o veículo na linha) e com isso obter os valores ideais para o melhor controle dos motores.

PID (Proporcional-Integral-Derivativo) é o algoritmo de controle muito utilizado nos sistemas de controle industrial. A popularidade de controladores PID pode ser atribuída em parte ao seu desempenho robusto em uma ampla gama de condições de funcionamento e em parte à sua simplicidade funcional, que permite operá-los de uma forma simples e direta. O algoritmo PID é composto por três coeficientes: proporcional, integral e derivativo, que são variados para obter a resposta ideal.

Para as primeiras avaliações e testes do veículo foi desenvolvido um mapa de forma matricial de ordem 5, isto é, um mapa formado por cinco linhas na vertical e cinco na horizontal onde cada interceptação das linhas representa um ponto no mapa, com isso temos um mapa com 25 pontos distintos para a movimentação do veículo. No veículo utilizamos quatro sensores de refletância, os dois centrais ficaram responsáveis por manter o veículo na linha e os outros dois externos para identificar os cruzamentos no mapa, com isso se torna possível informar a nuvem cada vez que se passa por cruzamento, possibilitando o melhor controle da nuvem sob as ações do veículo.

A partir desse mapa, foi desenvolvido o primeiro algoritmo de avaliação de rotas, como no código do veículo não haverá nada a respeito do mapa, esse algoritmo funciona da seguinte forma: É injetado no programa em qual posição do mapa o veículo se encontra e qual a direção de sua frente, em seguida o programa recebe as informações do local em que o veículo deve ir para coletar os dados do sensor DHT22, a partir daí o programa deve se comunicar com o veículo e informar cada movimento que ele deve realizar até chegar ao local para realizar a coleta, isto é, através das coordenadas do veículo e do destino o programa deve informar se o veículo deve seguir em frente, virar a direita ou esquerda ou dar meio volta à cada vez que o veículo informar que esta em um cruzamento, além de informar o momento de coletar e enviar os dados dos sensores. Com isso o programa fica responsável não só de conhecer o mapa e indicar a ação que o veículo deve tomar, mas também de acompanhar passo a passo o posicionamento do veículo no mapa.

Além disso, para o tratamento de colisões foi implementado um algoritmo no veículo com o objetivo de evitar as colisões do veículo. O algoritmo através das interpretações dos dados recebidos pelo sensor ultrassônico faz com que o veículo pare e informe a nuvem caso algum objeto entre em sua frente. Enquanto o objeto permanecer em sua frente nada acontece, caso o objeto libere o caminho o veículo sai do estado de parado e continua a realizar a ação ordenada.

## Meta 4 - Teste do veículo autônomo e relatório

Nesta etapa foram realizados os testes para verificação da integração entre veículo, sensores e nuvem computacional. O veículo autônomo é capaz de andar pelo ambiente, coletando informações através de sensores conectados à um Arduino preso em sua carcaça. Esse robô pode seguir rotas definidas através de linhas no chão ou através de comandos enviados para seu controlador.

Para criar instâncias sob demanda, o veículo publicará uma mensagem na rede que será enviada ao serviço Charon, *middleware* para dispositivos IoT e Nuvem. Assim, os dados coletados dos sensores e os comandos são enviados por uma instância criada por esse serviço.

Os teste entre o veículo com os sensores enviando dados para *appliances* dedicados na Nuvem foram realizados e o ambiente é capaz de receber essas dados através de um *gateway* IoT. O comunicação da Nuvem controlando o veículo ainda encontra-se em desenvolvimento, pois os equipamento necessário para esta comunicação foi adquirido recentemente. Porém, alguns teste já foram realizados, dentre eles, a capacidade de atuar sobre o circuito através de um servidor Web dentro do próprio Arduino, recebendo dados via método GET e POST.

### 4.1 Resultados

Com o objetivo de utilizarmos utilizarmos a capacidade de processamento de uma nuvem computacional a fim de realizar um sensoriamento autônomo inteligente, o desenvolvimento deste trabalho possibilitou, através dos estudos do uso de arquiteturas de processadores abertos, realizar a construção de um veículo autônomo com capacidade de se locomover através de uma rota pré-determinada, realizar coleta de dados de sensores, realizar o tratamento afim de evitar colisões, além de interagir com nuvem computacional, afim de obter instruções das ações a serem tomadas.

Além disso, foi desenvolvido um algoritmo de avaliação de rotas, que através dos dados inseridos do posicionamento do veículo no mapa e das "missões" a serem realizadas pelo veículo, é capaz de avaliar a melhor rota e informar ao veículo quais os movimentos devem ser realizados passo-a-passo para o seu deslocamento até o destino da realização do sensoriamento.

Sobretudo, com este trabalho foi possível explorar tecnologias de Nuvem e sistemas distribuídos, desenvolvimento de aplicações e comunicação de serviços, sendo fundamental a troca de conhecimento entre os integrantes do grupo de ComCiDis. Além disso, está bolsa permitiu por em prática conhecimentos de eletrônica aprendidos no curso de graduação, o que proporcionou no crescimento técnico/científico como um todo.

# Relatório Final Bolsa I.C 2017

Nome do bolsista: João Guilherme Nobre Rittmeyer

Nome do Orientador: Dr. Fabio Porto

Tipo de bolsa: PIBIC

Período da Bolsa: 01/08/2016 à 31/07/2017

## Título: Ambiente Integrado para execução de *workflows* científicos

O processamento de grandes volumes de dados por aplicações científicas é um problema cada vez mais recorrente, pela crescente disponibilidade de dados que é produzida em escalas cada vez maiores.

Neste contexto a escalabilidade e o desempenho das aplicações científicas tornam-se necessárias, pois o volume de dados processado aumenta de forma exponencial o tempo de execução.

Uma das áreas que exemplificam este problema é a astronomia, pois, a evolução da tecnologia utilizada nos telescópios durante os anos, resulta no aumento massivo das informações captadas ao mapear o céu.

### Objetivos

1. Estudo dos framework Apache Spark[1,10], Apache Hadoop[2]

Nesta tarefa busca-se entender a lógica desses *frameworks* em sua teoria e analisa o seu funcionamento através de exemplos mostrados nas bibliografias e em aplicações desenvolvidas por outros autores.

2. Estudo da estrutura de dado *PhTree*[3]

O objetivo desta tarefa é entender a fundo os conceitos, limitações e a forma de implementação da estrutura de dados *PhTree* em uma aplicação no campo da astronomia.

A *PhTree* é uma das muitas implementações da estrutura de dados chamadas de árvore, porém, esta por sua vez é multidimensional e permite processar de forma eficiente até 62 dimensões.

3. Estudo e desenvolvimento da estrutura de dado *QuadTree*[4,9,12]

De forma, semelhante ao objetivo da segunda tarefa, esta busca entender e implementar a estrutura de dados chamada *QuadTree*, também considerada uma

árvore, porém, assim como a *PhTree*, é multidimensional com seu limite de dimensões igual a dois.

A *QuadTree* é usada para representar pontos e regiões. A sua forma de representação é bidimensional, cujo princípio fundamental é similar ao conceito da subdivisão binária.

4. Desenvolvimento da aplicação “*Isotropic Constellations Queries*” com o uso da *QuadTree* na Linguagem de Programação Python

Esta tarefa busca desenvolver uma aplicação na linguagem de programação *Python* integrada aos *frameworks* *Apache Spark* e *Apache Hadoop*. A aplicação tem como objetivo encontrar padrões geométricos com ou sem o uso da escala ao procurar possíveis soluções nos catálogos da astronomia com o uso da *QuadTree*.

Um padrão geométrico[8] representa um conjunto de pontos que expressam uma consulta geométrica no qual o seu relacionamento é dado pelas distâncias entre cada ponto, formando uma figura.

5. Desenvolvimento da aplicação “*Indexing Tree*” com o uso da *PhTree* na Linguagem de Programação Java

Esta tarefa busca desenvolver uma aplicação na linguagem de programação *Java* sem o uso dos *frameworks* *Apache Spark* e *Apache Hadoop*. A aplicação tem o objetivo de realizar o pré-processamento dos catálogos da astronomia, buscando inicialmente os vizinhos dado uma distância.

6. Publicação do artigo “*Pre-Processing and Indexing Techniques In Constellation Query*” na décima nona conferência internacional de “*Big Data Analytics and Knowledge Discovery*” (*DaWaK 2017*)

Este artigo, descreve os testes realizados e publicados nas implementações que utilizam a *QuadTree* e *PhTree* ao comparar o desempenho do pré-processamento das duas aplicações, mostrando assim a melhor estrutura de dados para realizar as buscas espaciais com os dados da astronomia.

7. Desenvolvimento da aplicação “*Non-Isotropic Constellations Queries*” com o uso da *QuadTree* na Linguagem de Programação Python

Esta tarefa implementa uma aplicação na linguagem de programação *Python* integrada aos *frameworks* *Apache Spark* e *Apache Hadoop*. A aplicação tal qual a “*Isotropic Constellations Queries*” encontra soluções com base nos padrões geométricos, porém, estas possuem uma distância a mais ou a menos, sendo que está diferença é a taxa de erro em relação ao padrão geométrico.



## Introdução

As aplicações científicas são desenvolvidas nas áreas de banco de dados[5], bioquímica[6], astronomia[7] e entre outras. Estas simulam por exemplo eventos com base em modelos matemáticos e análises de série temporal

Neste trabalho, estamos interessados no problema da busca por formas geométricas no céu. Em particular, na astronomia, um exemplo de padrão geométrico é o *Einstein Cross*, um fenômeno da astronomia previsto na teoria da relatividade de Albert Einstein, que ocorre pelo efeito das lentes gravitacionais[11]. Sua representação geométrica é semelhante a uma cruz, como mostra a figura 1.

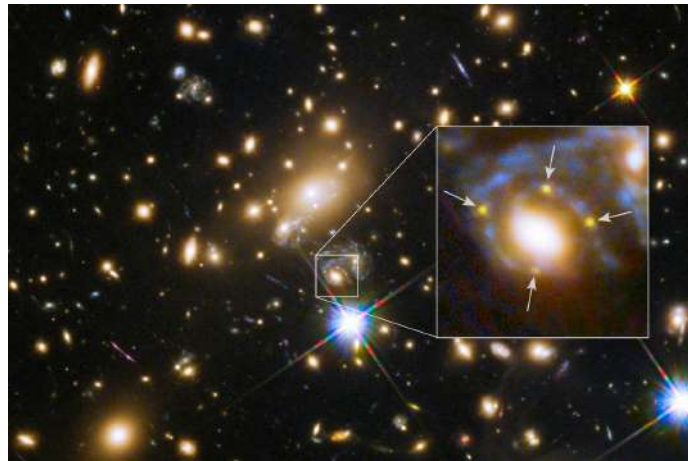


Figura 1: *Einstein Cross*

Fonte: S. Rodney / JHU / NASA / ESA / FrontierSN

A principal motivação deste trabalho é estudar e avaliar estratégias de execução paralela na identificação dos padrões geométricos ocultos nos grandes catálogos da astronomia. Utilizaremos como padrão exemplo a *Einstein Cross*.

Nossa estratégia reduz o número de comparações e otimiza o tempo de processamento ao selecionar apenas os conjuntos de estrelas que ajudam na composição de uma possível solução.

## Material e Métodos ou Metodologia

A realização deste trabalho requereu o estudo e a configuração dos *frameworks* Apache *Spark* e Apache *Hadoop*. A segunda etapa da pesquisa considerou o estudo e a implementação da estrutura de dados *Quadtree*, enquanto na *PhTree* somente o estudo foi realizado, pois, o código já existia.

Na terceira etapa foram desenvolvidos e testadas as aplicações “*Isotropic Constellations Queries*” e “*Indexing Tree*”. Na quarta etapa, após obter e comparar os resultados dos testes realizados na terceira etapa, estes foram utilizados na publicação do artigo “*Pre-Processing and Indexing Techniques In Constellation Queries*”.

Na quinta e última etapa foi desenvolvido “*Non-Isotropic Constellations Queries*” com o objetivo criar uma nova abordagem no processamento dos padrões geométricos. Esta nova abordagem flexibiliza a solução aceitando candidatos com uma pequena diferença a mais/menos em relação ao padrão geométrico.

## Resultados e Discussão

Durante a pesquisa os resultados obtidos na primeira e segunda etapa foram satisfatórios, pois em sua maior parte o esperado era o entendimento dos *frameworks* e das estruturas de dados utilizados. Na terceira etapa os resultados obtidos após a implantação foram superiores ao esperado como pode ser visto na figura 2. Este gráfico mostra a comparação do desempenho das aplicações “*Isotropic Constellations Queries*” e “*Indexing Tree*” ao variar o tempo de execução (*Exec Time*) em segundos e o raio (*Radius*). Como pode ser visto, o tempo da *QuadTree* com a altura definida igual a 7 (QT-7), 8 (QT-8) e 9 (QT-9) mostra-se constante enquanto o tempo da *PhTree*, aumenta à medida que o raio aumenta.

Embora a *QuadTree* apresente um resultado superior com essas alturas, a árvore pode esconder possíveis soluções, tornando as buscas imprecisas e assim a *PhTree* torna-se a melhor solução.

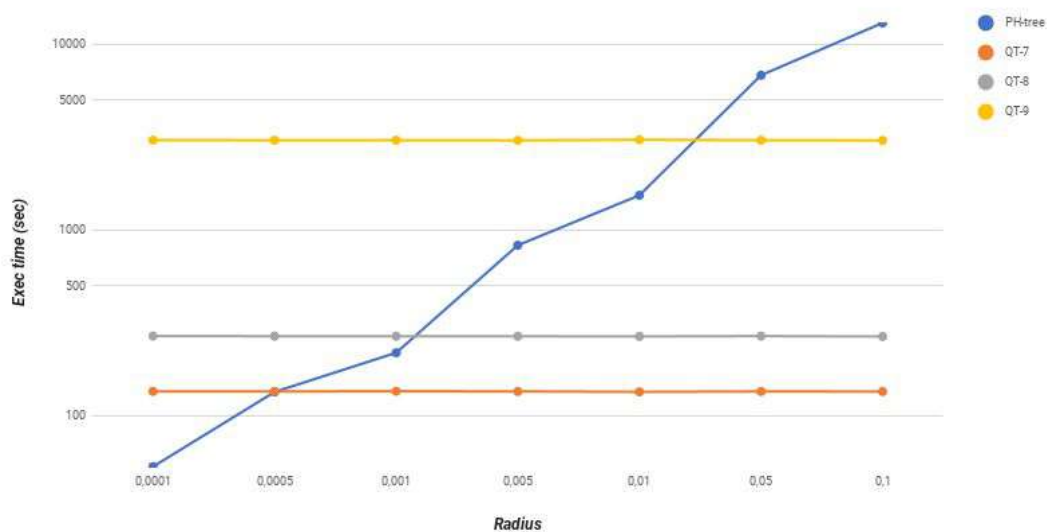


Figura 2: Tempo de busca X raio

Além desse teste, foi avaliado o consumo de memória das duas árvores durante a criação da estrutura de dado, não foi considerado o custo da memória na busca dos vizinhos.

A *QuadTree* com a altura definida em 9 níveis, consumiu 3.02 GB enquanto a *PhTree* utilizou 4.53 GB mostrando o uso 1.5 vezes menor pela *QuadTree*. Em um outro experimento, referente a *QuadTree* mostrado na figura 3, relata um teste no qual mede o uso da memória e o tempo com vários arquivos de tamanhos cada vez maiores. O

comportamento da memória ao lidar com esses arquivos, mostra que o consumo dobra a medida que o tamanho do arquivo aumenta quando a altura da *QuadTree* não é definida.

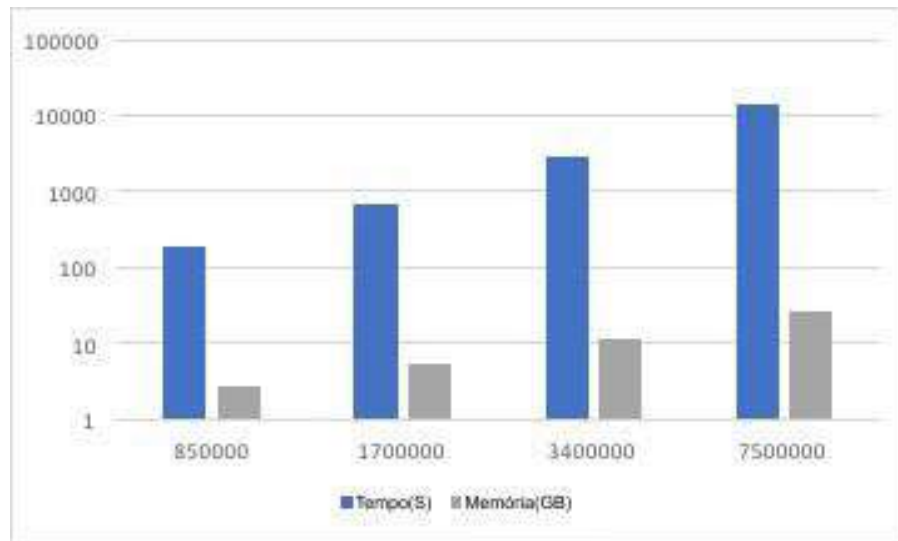


Figura 3: Tempo e memória

Com relação à última etapa desta pesquisa, estamos na fase de teste da implementação.

## Conclusões

Este trabalho permitiu entender de forma clara as técnicas de pré-processamento e processamento paralelo quando busca-se eficiência ao trabalhar com grandes volumes de dados, além disso, possibilitou o estudo dos algoritmos de composição voltados à astronomia e da própria astronomia.

De modo geral, as aplicações desenvolvidas segundo a teoria dos algoritmos de composição e na busca do padrão geométrico são semelhantes, embora a implementação do pré-processamento apresente diferenças técnicas.

Nos testes realizados, verificou-se que entre os trabalhos apresentados a *PhTree* mostra a melhor solução tanto no tempo de execução quanto na memória utilizada ao comparar com a *QuadTree*.

Logo a análise das soluções para encontrar o padrão geométrico em função da complexidade, mostra a necessidade de uma ferramenta eficiente e otimizada no contexto da astronomia que envolve profundamente o conceito de *BigData*.

Finalmente, gostaria de salientar que os resultados obtidos durante este período de bolsa PIBIC superaram as expectativas iniciais. Além do aprendizado técnico, conseguimos 1 publicação internacional. Adicionalmente, participei do Congresso da Sociedade Brasileira de Computação, em São Paulo, onde fui assistente do curso do Prof Fabio Porto, e assisti à palestras.

## Referências bibliográficas

- [1] Karau, Holden, et al. *Learning spark: lightning-fast big data analysis*. " O'Reilly Media, Inc.", 2015.
- [2] White, Tom. *Hadoop: The definitive guide*. " O'Reilly Media, Inc.", 2012.
- [3] Zäschke, Tilmann. "The PH-Tree Revisited." (2015).
- [4] Porto, F., Khatibi, A., Nobre, J. R., Ogasawara, E., Valduriez, P., and Shasha, D. "Constellation Queries over Big Data." *arXiv preprint arXiv:1703.02638* (2017).
- [5] Lustosa, Hermano, et al. "Database system support of simulation data." *Proceedings of the VLDB Endowment* 9.13 (2016): 1329-1340.
- [6] Costa, Raquel L., Carolina M. Voloch, and Carlos G. Schrago. "Comparative evolutionary epidemiology of dengue virus serotypes." *Infection, Genetics and Evolution* 12.2 (2012): 309-314.
- [7] de Oliveira, Douglas Ericson M., et al. "Avaliação da Localidade de Dados Intermediários na Execução Paralela de Workflows BigData." (2015).
- [8] Bishop, Christopher M. *Pattern recognition and machine learning*. springer, 2006.
- [9] Porto, Fabio, et al. "Constellation Queries over Big Data." *arXiv preprint arXiv:1703.02638* (2017).
- [10] Zaharia, Matei, et al. "Apache Spark: A unified engine for big data processing." *Communications of the ACM* 59.11 (2016): 56-65.
- [11] D. Overbye, "Astronomers observe supernova and find they are watching reruns," New York Times, USA, 2015.
- [12] Khatibi, A., Porto, F., Nobre, J. R., Ogasawara, E., and Valduriez, P. "Pre-processing and indexing techniques in constellation queries". (*dawak*, 2017)
- .

# Relatório Iniciação Científica – LNCC

João Victor Marinho Bechara

## 1 Dados gerais

- Título do projeto: Análise Comparativa Baseada em Redes Complexas dos Efeitos da Crise Econômica na Malha Aérea Brasileira
- Nome do bolsista: João Victor Marinho Bechara
- Nome do orientador: Artur Ziviani (LNCC)
- Tipo de bolsa e período do relatório: Iniciação Científica (07/2016-06/2017)

## 2 Objetivos

Análises recentes da estrutura da malha aérea brasileira [1], bem como da malha aérea global [2, 3] foram elaboradas utilizando grafos agregados da malha aérea, contendo apenas informações sobre aeroportos (localidades) e rotas (ligações entre aeroportos). Esse tipo de análise desconsidera a informação temporal de escalonamento dos voos, fazendo com que caminhos compostos por várias rotas não sejam confiáveis, já que não existe informação que garanta que o encadeamento de voos sugerido por esses caminhos vistos na rede agregada seja mesmo realizável. Ou seja, nesses modelos com rede agregada não é possível garantir a sequência temporal das rotas, uma vez que não há informação temporal no grafo de rotas.

A análise proposta neste trabalho é elaborada de tal forma que todas as informações temporais de agendamento dos voos sejam preservadas no modelo de malha aérea considerado. Assim é sempre possível garantir que essas informações de agendamento dos voos sejam automaticamente utilizadas na construção de caminhos no modelo. Desta forma, através desta análise, o principal objetivo do trabalho é verificar as diferentes estratégias das empresas aéreas na estruturação de suas malhas aéreas, incluindo a comparação de suas adaptações face à recente crise econômica.

## 3 Introdução

Os sistemas de transporte estão muito presentes no dia-a-dia das pessoas, sendo componente fundamental da sociedade moderna. Por isso, torna interessante analisá-los para entender como funciona o deslocamento da população e como as empresas que providenciam esse serviço se preparam para suprir a demanda que vem crescendo bem como se modificando ao longo dos anos. Nesse contexto, a malha aérea doméstica brasileira pode ser entendida naturalmente

como uma rede não trivial, tendo em vista que existem diversas companhias que atuam em diferentes regiões do país, oferecendo serviços para diferentes tipos de pessoas com um número diferentes de aviões.

A grave crise econômica enfrentada pelo país impacta diretamente diversos setores da economia, levando estes a se ajustarem face à nova realidade. Nesse contexto o setor de aviação civil não é exceção e houve por parte das companhias aéreas um ajuste na sua disponibilidade de voos, tendo por consequência uma recente modificação na malha aérea doméstica brasileira. Este trabalho analisa a malha aérea doméstica brasileira, também comparando sua estrutura com dados anteriores da mesma malha. Essas bases de dados foram coletadas do sítio eletrônico da Agência Nacional de Aviação Civil (ANAC) nos dias 3 de junho de 2015 e 13 de maio 2016. Essa modelagem é realizada através de um MAG [4], já que desta forma, pode-se analisar com naturalidade diferentes aspectos (aerportos, companhia aérea, instante de tempo e versão da base de dados).

## 4 Material e Métodos ou Metodologia

Neste estudo de caso, para identificação do núcleo central da rede foi utilizado o algoritmo K-Core [5]. O núcleo correspondente ao K-Core é obtido através de um processo de decomposição progressiva da rede, no qual remove-se progressivamente todos os vértices com conectividade inferior ao valor crescente de  $k$  e suas respectivas conexões. Por exemplo, o 2-Core de uma rede é formado pelo núcleo resultante da remoção de todos os vértices de grau 1, ou seja, o núcleo contendo somente vértices com no mínimo grau 2. O objetivo final do algoritmo é de que o último valor de  $k$  nos forneça apenas o núcleo central da rede, sem que haja a possibilidade de remoção de outros vértices.

A adoção do K-Core disponibiliza maior ganho nas análises, tendo em vista que retorna rede mais básica de núcleo da rede. Por exemplo, no caso de uma análise sob a perspectiva de voos, o núcleo é formado pelos aeroportos com maior incidência de voos. Porém, é possível aplicar qualquer outra métrica na base de dados que seja conveniente para a análise.

Com essa ferramenta de análise é possível encontrar as principais propriedades da malha aérea brasileira das duas bases de dados. Os núcleos dos grafos resultantes de cada empresa aérea informa o quanto cada aeroporto é importante para cada companhia. Outras informações relevantes que são obtidas com o K-Core são: localização da concentração das atividades de cada empresa aérea, comparação entre as estratégias das empresas, identificação dos hubs das empresas e assim por diante. Esse método pode ser aplicado no grafos resultantes das operações que serão descritas nas próximas seções.

## 5 Resultados e Discussão

A grave crise econômica enfrentada pelo país impacta diretamente diversos setores da economia, levando estes a se ajustarem face à nova realidade. Nesse contexto, o setor de aviação civil não é exceção e houve por parte das companhias aéreas um ajuste na sua disponibilidade de voos, tendo por consequência uma recente modificação na malha aérea doméstica brasileira.

Durante a análise comparativa das malhas aéreas em 2015 e 2016, foram encontradas diversas mudanças estruturais. Em primeiro, o número total de aeroportos foi reduzido de 110 para 109, mas a interseção da primeira base de dados (2015) e da segunda (2016) são apenas 101. Assim, 9 aeroportos não funcionam mais e 8 novos aeroportos começaram a operar. Outra mudança estrutural refere-se às companhias aéreas. A base de dados de 2015 possui 7 companhias aéreas, já a base de dados de 2016 possui 8 companhias aéreas. Porém, a interseção entre o conjunto das companhias das bases de dados resultam em 6 companhias aéreas. A companhia aérea Sete parou de operar da primeira (2015) para a segunda base de dados (2016). Enquanto isso, duas companhias aéreas começaram a operar voos comerciais nesse intervalo de tempo, as companhias TRIP e a FlyWays.

O número de componentes fortemente conexas na primeira base de dados foi de 7, uma por empresa aérea. Agora, este número aumentou para 10. Fato curioso é que a companhia aérea Trip possui 3 componentes fortemente conexas, mostrando como a crise econômica afetou a demanda dos voos e assim, as companhias aéreas tentaram se adaptar.

Uma das principais formas que as empresas aéreas encontraram para otimizar a utilização dos aviões, foi através do *codeshare*, ou seja, as companhias estão fazendo mais parcerias entre si para compartilhar mesmos voos. Entre as bases de dados analisadas, o número de *codeshare* passou de 923 para 1234 voos compartilhados por semana, um aumento de cerca de 33% no número de voos compartilhados por semana entre junho de 2015 e maio de 2016.

A Tabela 1 mostra a diferença do número de voos entre junho de 2015 e maio de 2016. É possível notar que as três maiores companhias aéreas do país (Gol, Azul e TAM) e uma das empresas pequenas (Passaredo) tiveram reduções significativas em seus voos, especialmente a Gol com 37% de redução no período. Já as empresas médias e pequenas (Avianca e Map) tiveram aumento no seu número de voos (13% e 46% de aumento respectivamente). O crescimento em média das empresas menores em tempos de crise pode ser visto como estratégia de preenchimento das lacunas dos voos deixadas pelas empresas maiores.

Table 1: Número de voos.

Empresa aérea	Voos antes da crise	Voos depois da crise	Diferença (%)
Gol	6644	4188	-37
Azul	5839	5075	-13
TAM	4814	4158	-14
Avianca	1386	1561	13
Passaredo	661	612	-7
Map	98	143	46

Resultados preliminares deste trabalho foram publicados em [6] junto aos resultados alcançados pelo trabalho do também bolsista de IC Bernardo Botelho Antunes da Costa.

## 6 Conclusões

Neste trabalho foi possível analisar as principais características da malha aérea brasileira nas duas bases de dados (junho de 2015 e maio de 2016). Uma análise comparativa traz resultados preliminares sobre como a crise econômica impactou na malha aérea brasileira e nas estratégias adotadas pelas companhias aéreas para se adaptar a essa crise.

Como resultado deste artigo, foi possível verificar os métodos adotados pelas companhias em resposta ao contexto de crise econômica. Verificou-se que, em média, as principais companhias aéreas tiveram um decréscimo considerável em sua quantidade de voos disponibilizados. Fato esse que levou a quantidade de voos cair cerca de 22% de uma base de dados para a outra. Enquanto isso, as pequenas e médias empresas viram como uma oportunidade de crescimento e por isso expandiram sua quantidade de voos e o número de aeroportos que operam. Foi visto que um dos principais métodos utilizados pelas grandes empresas foi a utilização do *codeshare*, havendo um aumento de 33% da quantidade de voos compartilhados de uma base de dados para a outra.

Com base nos resultados preliminares publicados em [6], um artigo mais completo encontra-se em fase final de redação visando sua submissão a um periódico indexado.

## References

- [1] G. S. Couto, A. P. C. d. Silva, L. B. Ruiz, and F. Benevenuto, “Structural Properties of the Brazilian Air Transportation Network,” *Anais da Academia Brasileira de Ciências (AABC)*, vol. 87, pp. 1653–1674, Setembro 2015.
- [2] T. Verma, N. a. M. Araújo, and H. J. Herrmann, “Revealing the structure of the world airline network,” *Scientific Reports*, vol. 4, pp. 1–6, 2014.
- [3] P. Wei, L. Chen, and D. Sun, “Algebraic connectivity maximization of an air transportation network: The flight routes’ addition/deletion problem,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 61, pp. 13–27, 2014.
- [4] K. Wehmuth, É. Fleury, and A. Ziviani, “On MultiAspect graphs,” *Theoretical Computer Science (TCS)*, vol. 651, pp. 50–61, Outubro 2016.
- [5] S. B. Seidman, “Network structure and minimum degree,” *Social Networks*, vol. 5, no. 3, pp. 269 – 287, 1983.
- [6] B. B. A. Costa, J. V. M. Bechara, K. Wehmuth, and A. Ziviani, “Modelagem e análise da malha aérea doméstica brasileira com grafos multiaspectos (MAGs),” in *Anais do X Encontro Acadêmico de Modelagem Computacional – EAMC 2017*, (Petrópolis, RJ), Fevereiro 2017.



## **Laboratório Nacional de Computação Científica**

Relatório de atividades PIBIC      Período: 01/03/2017 á 31/07/2017

Processo CNPq 800333/2016-0

## Dados Gerais

Este documento é uma descrição das pesquisas feitas pelo bolsista **Lucas Borsatto Simão** (CPF 139.440.047-01) para o Programa de Bolsas de Iniciação Científica (PIBIC) do Laboratório Nacional de Computação Científica (LNCC). O trabalho possui como título: **Caracterização de Texturas Utilizando Redes Neurais e Espaços Fractais com Aplicações a Imagens Radiológicas** e os estudos foram supervisionados pelo professor Gilson Antônio Giraldi, pesquisador da instituição.

## Introdução

Dentro da área de aprendizado de máquina, a sub área de modelos geradores tem ganhado maior notoriedade nos últimos anos. Isso se deve ao fato de que, por conta dos avanços nos algoritmos, estes modelos tem se tornado mais precisos. Dentre os modelos geradores mais usados estão as *Generative Adversarial Nets* (GAN)[1], que se utiliza do conceito de aproximação entre a distribuição gerada e a distribuição que se deseja obter.

Desde sua criação, várias variações das GANs já foram propostas, e sua importância recai sobre diversas áreas e aplicações, principalmente quando se lida com imagens. Como exemplos desses processos, pode-se citar a criação de imagens artificiais, retirada de ruído, aumento de resolução de imagens e preenchimento automático de falhas [2, 3, 4].

Também entre estes processos esta o de tradução image-to-image, que consiste na transição de uma imagem de um domínio para outro. Esta operação de transformação num espaço é possível para qualquer domínio que seja requerido. Dentro deste contexto, este trabalho tem a proposta de utilização das DualGANs [5], principalmente com texturas, para remapear espaços de dimensão  $N$  para que sejam representados em dimensões menores, processo conhecido como redução de dimensionalidade. Através desta redução podem-se obter, por exemplo, as componentes de maior importância ou uma representação mais significativa do próprio problema abordado. Para isto, é considerada a utilização da função LOGMAP.

## Generative Adversarial Nets

Modelos que contam com a aproximação direta da máxima verossimilhança, tais como os Variational Autoencoders [6], normalmente apresentam problemas com geração de imagens borradas por conta da sua proposta de aproximar a distribuição real corrigindo suas respostas

através da própria imagem de entrada.

Diferentemente destes modelos, as GANs contam com uma função de aproximação entre as próprias distribuições. Assim, a estrutura do modelo se divide em duas redes, a geradora e a discriminadora, conforme mostra a imagem da figura 1.

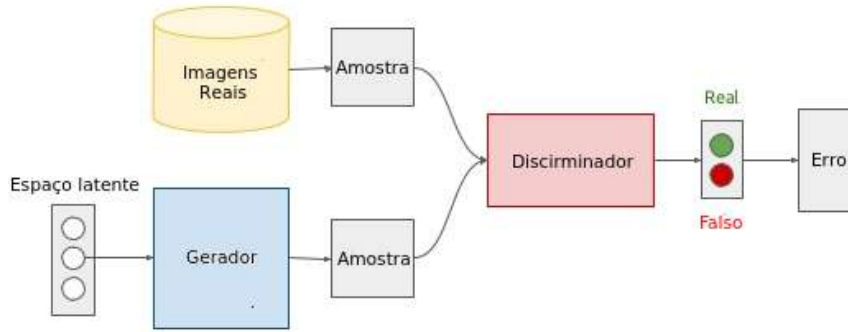


Figura 1: Representação do modelo GAN.

A rede geradora recebe como entrada um vetor de valores aleatórios, gerados através de uma distribuição gaussiana, por exemplo. Esta rede é responsável pela geração das imagens, que irão ser comparadas com as imagens do conjunto original. Esta comparação é feita pela rede discriminadora. Esta compara ambas as imagens, e como resposta define se a imagem é real ou falsa.

Então, o processo consiste no gerador tentando criar imagens cada vez mais parecidas com as reais, enquanto o discriminador é aprimorado para discernir estas das reais. Para isso, ele se utiliza da função de log verossimilhança, numa operação conhecida como minimax, mostrada através da equação:

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)} [\log(D(x))] + E_{x \sim P_z(z)} [\log(1 - D(G(z)))], \quad (1)$$

onde  $G$  representa o modelo gerador e  $D$  o modelo discriminador.  $P_{data}(x)$  é a distribuição de dados reais e  $P_z(z)$  é a distribuição associada ao ruído aleatório  $z$ .

Desta forma, pode se assemelhar o processo com um jogo de dois agentes, onde o gerador tenta minimizar as chances do discriminador classificar corretamente, enquanto o discriminador tenta maximizar esta mesma função.

Por consequência, é mostrado em [1] que o modelo pode ser representado aproximando a função de divergência de Jensen-Shannon, cujo valor mínimo é  $-\log(4)$ , e a função de custo aproximada é calculada através da equação:

# **BaMBa *Analytics*: Uma Arquitetura baseada em Banco de Dados e Visualização no apoio à Análise da Biodiversidade Marinha Brasileira**

**Bolsista: Matheus Machado da Rosa Albuquerque<sup>1</sup>**  
**Orientador: Luiz Manoel Rocha Gadelha Júnior<sup>2</sup>**  
**Coorientadora: Kary Ann del Carmen Ocaña Gautherot<sup>2</sup>**  
**Colaborador: Pedro Milet Meirelles<sup>3</sup>**

<sup>1</sup>Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (Cefet/RJ)  
Petrópolis – RJ – Brasil

<sup>2</sup>Laboratório Nacional de Computação Científica (LNCC)  
Petrópolis – RJ – Brasil

<sup>3</sup>Universidade Federal do Rio de Janeiro (UFRJ)  
Rio de Janeiro – RJ – Brasil

{matheusm,lgadelha,karyann}@lncc.br, pedrommeirelles@gmail.com

## **PIBIC - 2017**

### **1. OBJETIVOS**

O objetivo do presente estudo é desenvolver uma estrutura *online* baseada em bancos de dados e visualização no apoio à análise da Biodiversidade Marinha Brasileira. O projeto aborda a problemática da exploração e gerenciamento de dados coletados pela Rede Nacional de Pesquisa em Biotecnologia Marinha (BiotecMar<sup>1</sup>).

Devido não somente à dificuldade de se obter dados consistentes e completos como também a crescente preocupação em relação a degradação da biodiversidade, constatamos a importância do desenvolvimento de uma ferramenta que auxilie na análise e manipulação desses dados.

A aplicação está sendo desenvolvida na linguagem estatística R [Team 2000], que torna a implementação da interface robusta e permite a inclusão de *scripts* específicos em R para a exploração dos dados. A integração da aplicação com o banco de dados permite gerar estatísticas e consultas de forma efetiva e facilitada. Estimamos que a realização das análises ocorra em um tempo de execução otimizado, uma vez que a aplicação será instalada nos servidores e clusters de computadores no LNCC.

### **2. INTRODUÇÃO**

O presente trabalho visa o desenvolvimento de um ambiente gráfico para análise e manipulação de dados via *web* da Biodiversidade Marinha Brasileira, através do projeto BiotecMar.

O Projeto Biotecmar tem como principal objetivo desenvolver pesquisas inovadoras nas áreas de biodiversidade, prospecção, genômica, pós-genômica (ômicas) e trans-

---

<sup>1</sup><http://www.biotecmar.com.br>

ferência para o setor produtivo. Em âmbito geral, a Rede Nacional de Pesquisa em Biotecnologia Marinha está estruturada em quatro grandes eixos temáticos e conta com grupos consolidados e articulados, visando o desenvolvimento de temáticas de fronteira da biotecnologia marinha.

O projeto Bamba (Brazilian Marine Biodiversity Database) [Meirelles et al. 2015] aborda a problemática da exploração e gerenciamento de dados coletados pelo BiotecMar. Podem ser integrados ao banco de dados conjuntos de dados certificados obtidos a partir de estudos holísticos integrados, que compreendem diversos tipos de parâmetros, permitindo acesso a uma grande variedade de recursos marinhos.

Constatamos a dificuldade dos pesquisadores obterem dados consistentes, completos e compatíveis com os *softwares* mais usados nas respectivas análises. Ferramentas que apoiem a análise de dados, envolvem etapas desde a concepção de bancos de dados que armazenem tais dados, até a implementação de rotinas que garantam a qualidade desses dados, como sistemas de gerência ou sistemas baseados na *web*.

Nessa proposta visa-se o desenvolvimento de BaMBa *Analytics*, uma arquitetura baseada em interface *web* integrada a banco de dados, que permite ao usuário consultar e minerar dados e gerar estatísticas com eficiência e facilidade.

## 2.1. TRABALHOS RELACIONADOS

Uma das maiores barreiras para mensurar e monitorar de forma efetiva as alterações na biodiversidade, deve-se a falta de ferramentas que auxiliem nesse controle. É fundamental para a conservação e gestão da biodiversidade um monitoramento com maior precisão e integrabilidade com os sistemas e dados já existentes.

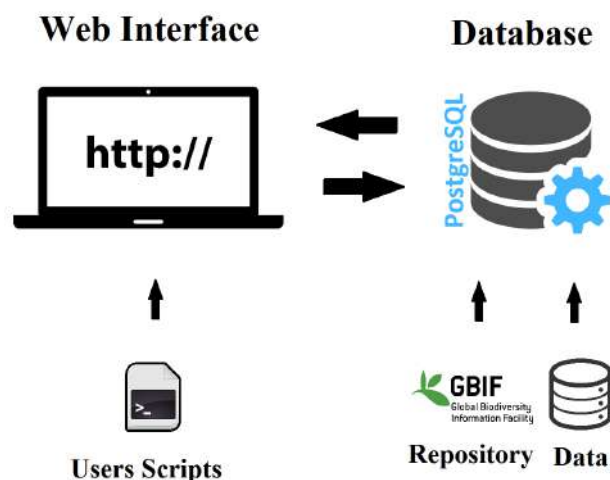
”As aves são janelas incomparáveis em processos bióticos em todos os níveis e são indicadores de bem-estar ecológico comprovada”[Wood et al. 2011]. A compreensão da distribuição e dinâmica das espécies são aspectos fundamentais para a conservação e gestão da ecologia.

O projeto eBird [Wood et al. 2011] tem como objetivo recolher registros de observação de aves. A partir desses dados é possível desenvolver modelos geográficos de distribuição até realizar inferências sobre associações de espécies e habitats. Entretanto, os modelos gerados pelo projeto podem ser de difícil entendimento e análise, em visão disso foi desenvolvido o BirdVis, um sistema de visualização interativo que permite a análise dos dados, podendo identificar correlações e padrões relevantes.

É notória a crescente demanda da comunidade científica em buscar soluções para averiguar e deter o constante avanço da degradação da biodiversidade. ”Reduzir a taxa de perda de biodiversidade e prevenção de perigosas alterações na biodiversidade são desafios internacionais...”[Pereira et al. 2013]. Essa preocupação pode ser notada em *Essential Biodiversity Variables* [Pereira et al. 2013]. O artigo aborda a problemática de controle real da biodiversidade de forma efetiva salientando a busca de uma base de programas de monitoramento da biodiversidade em todo o mundo.

## 3. METODOLOGIA

O BaMBa *Analytics* está formado por duas camadas principais: a interface *web* e o banco de dados, como mostrado na Figura 1.



**Figura 1. Esquema conceitual do BaMBa Analytics**

A interface está sendo desenvolvida na linguagem estatística R. A escolha é justificada pois o R torna a implementação da interface robusta e de integração facilitada com o banco de dados devido às suas capacidades relacionais de importação e exportação de dados. A linguagem R possibilita a criação de estatísticas específicas através de gráficos graças a sua vasta coleção de bibliotecas e funções voltadas para este tipo de análise.

O R já possui um framework<sup>2</sup> voltado para a criação e desenvolvimento de aplicações *web*, o Shiny<sup>3</sup>, gerando uma melhor integração com todo código. Conseqüentemente, a interface irá permitir o *upload* de *scripts* específicos em R para manipulação dos dados. As tabelas de dados metagenômicos geradas pelo projeto BiotecMar seguem o padrão de saída gerado pela aplicação MG-RAST [Meyer et al. 2008]. Foi desenvolvida uma rotina que converte os dados em formato tabular gerado pelo MG-RAST para o padrão Darwin Core [Wieczorek et al. 2012], o que permite a publicação automática dos mesmos.

O banco de dados foi implementado em PostgreSQL e faz utilização de alguns frameworks que são responsáveis por assumir o controle de diversos procedimentos. Através desta aplicação *web*, uma quantidade massiva de dados contidos em planilhas, são importados e armazenados em um banco de dados relacional.

O BaMBa *Analytics* está integrado ao repositório de dados GBIF [Edwards et al. 2000], através de uma API via serviço RESTful que importa informações de diversas categorias de dados e metadados que seguem padrões como o Darwin Core.

#### 4. RESULTADOS E DISCUSSÃO

O BaMBa *Analytics*, assim como as suas camadas de interface *web* e banco de dados, foram instalados para fins de teste nos servidores de clusters de computadores no LNCC.

<sup>2</sup>Um framework (ou biblioteca) é uma abstração que une códigos comuns entre vários projetos de *software* provendo uma funcionalidade genérica.

<sup>3</sup><http://shiny.rstudio.com/>

O uso da linguagem R auxilia a criação de um ambiente automático que favorece ao processamento de dados via *scripts*. Também será possível que o pesquisador gere análises usando os modelos de estatística pré-definidos, não necessitando de conhecimento sobre a linguagem R para explorar os dados de forma mais profunda.

O gerenciamento e armazenamento de dados por meio do banco de dados relacional possibilita efetuar buscas pontuais e gerar visualizações geográficas da Biodiversidade Marinha Brasileira. A Figura 2 mostra o resultado da integração das camadas do BaMBA *Analytics* em interação constante com o pesquisador, que realiza a análise e validação dos dados de maneira integrativa. A Figura 3 mostra a rotina de conversão que foi desenvolvida em funcionamento, onde o arquivo convertido é disponibilizado para *download*.

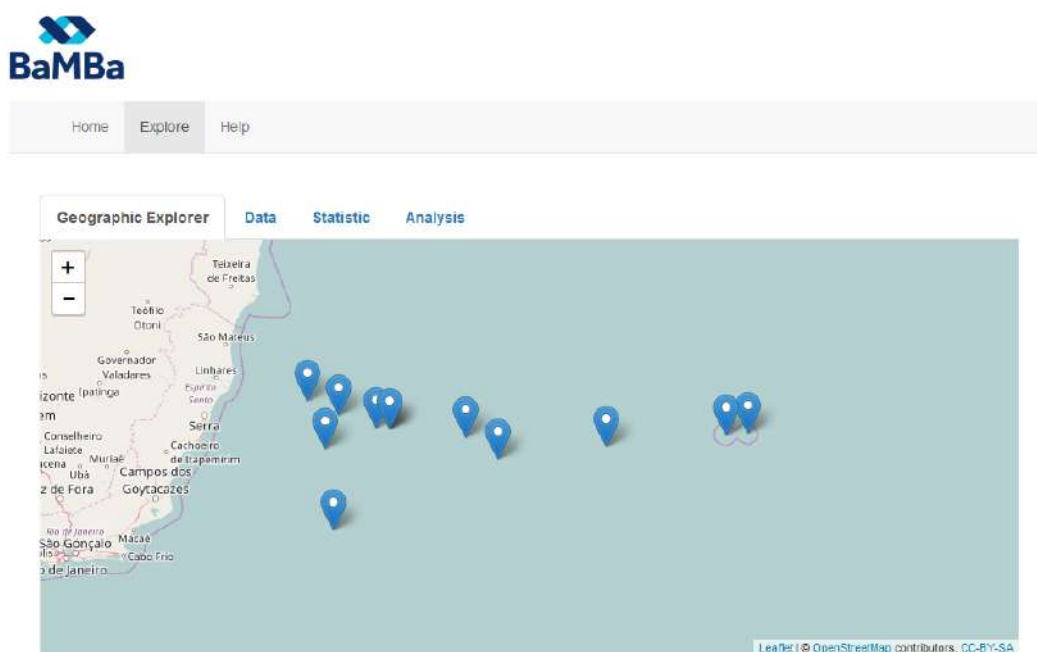


Figura 2. Resultado da interface *web* do BaMBA *Analytics*

**Uploading Files**

**Choose Taxonomic TSV File**

**Choose Functional TSV File**

**Choose Metadata TSV File**

OcurrenceID	BasisOfRecord	EventDate	EventID	Kingdom	Phylum	Class
1	HumanObservation	2007	4480742.30	Archaea	Crenarchaeota	Thermo
2	HumanObservation	2007	4487911.30	Archaea	Crenarchaeota	Thermo
3	HumanObservation	2007	4487911.30	Archaea	Crenarchaeota	Thermo
4	HumanObservation	2007	4487909.30	Archaea	Crenarchaeota	Thermo
5	HumanObservation	2007	4487911.30	Archaea	Crenarchaeota	Thermo
6	HumanObservation	2007	4487911.30	Archaea	Crenarchaeota	Thermo
7	HumanObservation	2007	4487909.30	Archaea	Crenarchaeota	Thermo
8	HumanObservation	2007	4480742.30	Archaea	Crenarchaeota	Thermo
9	HumanObservation	2007	4480742.30	Archaea	Crenarchaeota	Thermo
10	HumanObservation	2007	4487911.30	Archaea	Crenarchaeota	Thermo
11	HumanObservation	2007	4487910.30	Archaea	Crenarchaeota	Thermo
12	HumanObservation	2007	4480743.30	Archaea	Crenarchaeota	Thermo

Figura 3. Resultado da interface do *web* conversor

## 5. TRABALHOS FUTUROS

Realizar análises dos perfis funcionais de comunidades microbianas tem sido uma das tarefas mais importantes no campo da metagenômica. O perfilamento derivado dessas análises tem contribuições fundamentais para a pesquisa biológica, pois dessa forma é possível prever informações como a abundância dos genes funcionais dos organismos presentes na amostra.

Com o avanço dos métodos empenhados na coleta e análise de dados metagenômicos o volume de dados tal como o comprimento das leituras produzidas por plataformas de sequenciamento tem aumentado de forma gradativa a não escalar bem com o funcionamento das ferramentas atuais de análise metagenômica, exigindo cada vez mais eficiência nas execuções.

Foi possível identificar o destaque de desempenho da ferramenta de análise SUPER-FOCUS [Silva et al. 2015] dentre as demais. O SUPER-FOCUS (SUBsystems Profile by databases Reduction using FOCUS), tem seu funcionamento baseado no conceito de homologia ágil, utilizando uma base de dados reduzida para relatar os subsistemas presentes na amostra original e perfilar suas abundâncias. A escolha dessa ferramenta não deve-se apenas ao seu largo poder de processamento mas também a seu alto potencial de otimização.

Com isso, foi previsto a possibilidade de implementar o paralelismo de algumas etapas do *workflow* do SUPER-FOCUS. Estimamos que com esse método a ferramenta consiga reduzir significativamente seu tempo de execução permitindo o pesquisador optar por alinhadores mais precisos sem perder em desempenho. Para essa tarefa, foi escolhida a plataforma Swift [Zhao et al. 2007]. Assim, é estimado que a aplicação ganhe em desempenho de forma a estar preparada para atender a todo tipo de demanda de análise computacional tal como para o futuro.

## 6. CONCLUSÕES

O uso da linguagem R no BaMBA *Analytics* favorece um ambiente transparente e amigável, onde o pesquisador poderá fazer o *upload* de seus próprios dados e *scripts* e assim, analisá-los. A partir da importação e armazenamento dos dados será possível efetuar buscas pontuais por região, demonstrar estatísticas e obter uma visualização geográfica através de mapas.

O BaMBA *Analytics* será instalado em servidores alocados no LNCC com o intuito de atender os requisitos da comunidade científica em relação ao tempo de execução das tarefas efetuadas, oferecendo um ambiente paralelo e distribuído para a execução de análises de maneira eficiente e segura. Por conseguinte, é estimado que a arquitetura torne a execução dessas pesquisas e consultas de dados biológicos simples e possível em um tempo adequado, como requerido pelos cientistas.

Como passos futuros serão realizados testes de desempenho para analisar os pontos de otimização da arquitetura, a fim de reduzir o tempo e gasto computacional.



## Referências

- Edwards, J. L., Lane, M. A., and Nielsen, E. S. (2000). Interoperability of biodiversity databases: biodiversity information on every desktop. *Science*, 289(5488):2312–2314.
- Meirelles, P. M., Gadelha, L. M., Francini-Filho, R. B., de Moura, R. L., Amado-Filho, G. M., Bastos, A. C., Paranhos, R. P. d. R., Rezende, C. E., Swings, J., Siegle, E., et al. (2015). Bamba: towards the integrated management of brazilian marine environmental data. *Database*, 2015.
- Meyer, F., Paarmann, D., D’Souza, M., Olson, R., Glass, E. M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., et al. (2008). The metagenomics rast server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics*, 9(1):386.
- Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R., Scholes, R. J., Bruford, M. W., Brummitt, N., Butchart, S., Cardoso, A., et al. (2013). Essential biodiversity variables. *Science*, 339(6117):277–278.
- Silva, G. G. Z., Green, K. T., Dutilh, B. E., and Edwards, R. A. (2015). Super-focus: a tool for agile functional analysis of shotgun metagenomic data. *Bioinformatics*, 32(3):354–361.
- Team, R. C. (2000). R language definition. *Vienna, Austria: R foundation for statistical computing*.
- Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., and Vieglais, D. (2012). Darwin core: an evolving community-developed biodiversity data standard. *PloS one*, 7(1):e29715.
- Wood, C., Sullivan, B., Iliff, M., Fink, D., and Kelling, S. (2011). ebird: engaging birders in science and conservation. *PLoS biology*, 9(12):e1001220.
- Zhao, Y., Hategan, M., Clifford, B., Foster, I., Von Laszewski, G., Nefedova, V., Raicu, I., Stef-Praun, T., and Wilde, M. (2007). Swift: Fast, reliable, loosely coupled parallel computation. In *Services, 2007 IEEE Congress on*, pages 199–206. IEEE.

# Simulação Computacional de Algoritmos Quânticos

Aluno: Matheus Roberto Araújo Borsato/Faeterj

Orientador: Renato Portugal/LNCC

07 de julho de 2017

## 1 Introdução

A Computação Quântica teve início na década de 80 com os trabalhos de Benioff e Feynman [1, 2]. Feynman argumentou que um computador clássico pode simular com eficiência diversos problemas da Física Clássica, mas o mesmo não ocorre com os problemas da Física Quântica, uma vez que a dimensão do espaço de Hilbert cresce exponencialmente em função do número de partículas acrescentadas ao sistema. Feynman perguntou se um dispositivo que usasse as leis da Mecânica Quântica para realizar cálculos não poderia simular eficientemente a própria Mecânica Quântica. Os argumentos de Feynman estimularam David Deutsch a generalizar o modelo mais fundamental da Computação Clássica, a saber, a máquina de Turing, para o seu equivalente quântico num trabalho histórico de 1985 [3]. Posteriormente generalizou também o modelo de circuitos baseado em portas lógicas clássicas. Operadores unitários tomaram o lugar das usuais portas lógicas AND, OR e NOT [3]. O modelo de circuitos é o modelo adequado para o desenvolvimento de novos algoritmos. Baseado nele, Shor elaborou um algoritmo exponencialmente mais rápido para fatoração de números inteiros e cálculo de logaritmo discreto [5]. Esses algoritmos permitem a quebra dos principais códigos de criptografia usados atualmente, como RSA, Diffie-Hellman e ElGamal, caso um computador quântico de tamanho razoável esteja disponível. O artigo de Shor foi um grande estímulo para o crescimento dessa área. Atualmente, é possível simular classicamente computadores quânticos com 42 qubits usando um eficiente programa multi-core em um supercomputador<sup>1</sup>. Esse simulador foi usado para fatorar 15707 em 113x139 usando

o algoritmo de Shor. Por outro lado, na construção do hardware quântico verdadeiro, devemos levar em conta que pelo menos dois terços da memória principal serão usados para códigos corretores de erros. Assim, as novidades em termos de aplicações só virão quando computadores quânticos com mais de 126 qubits estiverem disponíveis. O estado da arte do hardware quântico, com comprovação da comunidade científica, é um computador de 12 qubits<sup>2</sup>. Por outro lado, a empresa Canadense D-Wave já produziu um computador quântico de 128 qubits usando a técnica “computação quântica adiabática” e vendeu uma unidade para a empresa de equipamentos bélicos Lockheed Martin<sup>3</sup>. Recentemente a Nasa adquiriu uma nova versão com mais memória.

## 2 Objetivo

O objetivo geral deste plano de trabalho é complementar a minha formação na área de Computação Quântica com foco em algoritmos quânticos e na sua simulação usando computação de alto desempenho. Os objetivos específicos visam ao desenvolvimento dos fundamentos matemáticos dessa teoria, o estudo das principais portas lógicas quânticas e introdução aos algoritmos de busca usando passeios quânticos com simulações em Python.

## 3 Resultados

Até este momento tive um bom desenvolvimento quanto a linguagem de programação Python, tendo começado estudando um pouco sobre Variáveis, expressões e sentenças.

Python é uma linguagem de programação de alto nível, interpretada, de script, imperativa, orientada a objetos, funcional, de tipagem dinâmica e forte. Foi lançada por Guido van Rossum em 1991. Apesar de várias partes da linguagem possuírem padrões e especificações formais, a linguagem como um todo não é formalmente especificada.

A linguagem foi projetada com a filosofia de enfatizar a importância do esforço do programador sobre o esforço computacional. Prioriza a legibilidade do código sobre a velocidade ou expressividade. Combina uma sintaxe concisa e clara com os recursos poderosos de sua biblioteca padrão e por módulos e frameworks desenvolvidos por terceiros.

Devido às suas características, ela é principalmente utilizada para processamento de textos, dados científicos e criação de CGIs para páginas dinâmicas para a web.

Variáveis são espaços da memória RAM do computador no qual são armazenadas informações passadas pelo próprio programador, de acordo com o desejado.

Uma sentença é uma unidade de código que o interpretador Python pode executar.

Uma expressão é uma combinação de valores, variáveis e operadores.

E começando por esse ponto foi também visto: Execução Condicional, Funções, Interação, Strings, Arquivos e Listas.

Nesse tempo saindo um pouco da teoria e também aplicando esse conhecimento em Python em alguns exercícios pude aprender um pouco mais sobre a linguagem, alguns conceitos foram entendidos de uma forma mais clara ao aplicar e utilizar nos problemas sugeridos nos exercícios.

Com base de estudos na apostila **Python para Informática – da Universidade Federal de Mato Grosso do Sul – UFMS[10]** e também em vídeo aulas no site do **Curso em Vídeo[12]**, o conhecimento até aqui foi bem amplo, podendo conhecer aplicações muito interessantes e problemas da Computação Quântica, pois Python é uma linguagem de programação que não leva muito trabalho de se entender, pois a mesma trabalha com palavras bem próximas a nosso vocabulário, e também pois a forma de processamento da linguagem é mais funcional e rápida que as outras para os problemas de Computação Quântica.

Como dito no início deste tópico um dos outros objetivos era a aplicação de Grafos.

Ao iniciar o estudo de Grafos foi visto o conteúdo pela apostila **Grafos – Uma Introdução de Samuel Jurkiewicz[9]** tendo como conteúdo: O que é um Grafo, Grau de um Vértice, Alguns Problemas com as Definições, Isomorfismo e Tipos Especiais de Grafos.

Tendo base de estudos também o site Wikipedia, pude me aprofundar um pouco mais na parte de teoria dos Grafos, entendo um pouco mais da história e como foi a aplicação mais conhecida que tem o nome de: O Problema das Sete Pontes de Königsberg.

A teoria dos grafos é um ramo da matemática que estuda as relações entre os objetos de um determinado conjunto. Para tal são empregadas estruturas chamadas de grafos,  $G(V,E)$ , onde  $V$  é um conjunto não vazio de objetos denominados vértices e  $E$  é um subconjunto de pares não ordenados de  $V$ , chamados arestas.

Dependendo da aplicação, arestas podem ou não ter direção, pode ser permitido ou não arestas ligarem um vértice a ele próprio e vértices e/ou arestas podem ter um peso (numérico) associado. Se as arestas têm uma direção associada temos um dígrafo. Um grafo com um único vértice e sem arestas é conhecido como grafo trivial.

O artigo de Leonhard Euler, publicado em 1736, sobre o problema das sete pontes de Königsberg, é considerado o primeiro resultado da teoria dos grafos. É também considerado um dos primeiros resultados topológicos na geometria; isto é, não dependente de quaisquer medidas. Isso ilustra a profunda conexão entre a teoria dos grafos e topologia.

Ao iniciar o estudo de Grafos foi visto o conteúdo pela apostila Grafos – Uma Introdução de Samuel Jurkiewicz tendo como conteúdo: O que é um Grafo, Grau de um Vértice, Alguns Problemas com as Definições, Isomorfismo e Tipos Especiais de Grafos.

Ainda falta conhecimento a ser aprendido, porém já ocorreu um grande avanço, e até o final espero adquirir o máximo de conhecimento que eu puder.

Figura 1:

Grafo com quatro vértices e 6 arestas. É um grafo completo, conexo e planar.

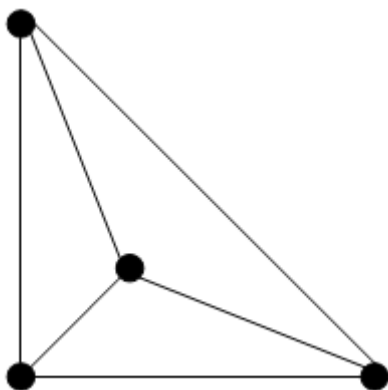


Figura 2:

O problema das pontes de Königsberg



## Bibliografia:

[https://pt.wikipedia.org/wiki/Teoria\\_dos\\_grafos](https://pt.wikipedia.org/wiki/Teoria_dos_grafos)

<https://pt.wikipedia.org/wiki/Python>

## Referências:

- [1] P. Benioff. The Computer as a Physical System: A Microscopic Quantum Mechanical Hamiltonian Model of Computers as Represented by Turing Machines. *Journal of Statistical Physics*, 22, 1980.
- [2] R. Feynman. Simulating Physics with Computers. *International Journal of Theoretical Physics*, 21:467–488, 1982.
- [3] D. Deutsch. Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer. *Proceedings of the Royal Society of London, A* 400:97–117, 1985.
- [4] D. Deutsch. Quantum Computational Networks. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 425(1868):73–90, 1989.
- [5] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, 1997.
- [6] Portugal, R. [et al.], *Uma Introdução à Computação Quântica SBMAC*, São Carlos, 2010.
- [7] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, UK, 2000.
- [8] R. Portugal. *Quantum Walks and Search Algorithms*. Springer, NY, 2013.
- [9] *Grafos – Uma Introdução* de Samuel Jurkiewicz
- [10] *Python para Informática – da Universidade Federal de Mato Grosso do Sul – UFMS*
- [11] <https://www.youtube.com/channel/UCrWvhVmt0Qac3HgsjQK62FQ>

# Relatório de Atividades PIBIC 2017

Matheus Tonelli de Souza<sup>1,2</sup>,  
Kary Ocaña<sup>1</sup>, Luiz M. R. Gadelha Jr<sup>1</sup>

<sup>1</sup>Laboratório Nacional de Computação Científica (LNCC)  
Petrópolis – RJ – Brazil

<sup>2</sup>Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (Cefet/RJ)  
Petrópolis – RJ – Brazil

{matheust, karyann, lgadelha}@lncc.br

## 1. Dados Gerais

Este documento é uma descrição das pesquisas feitas pelo bolsista Matheus Tonelli de Souza para o Programa de Bolsas de Iniciação Científica (PIBIC) do Laboratório Nacional de Computação Científica (LNCC) com vigência de **01/08/2016 a 31/07/2017**. O trabalho possui como título: **Aplicações de técnicas de aprendizagem de máquinas à análise de dados de domínio e de tempo de execução de workflows de bioinformática** os estudos foram supervisionados por Luiz Gadelha e Kary Ocaña.

## 2. Introdução

Em experimentos científicos de alto desempenho, faz-se o uso de grandes simulações com um fluxo de diferentes programas sendo executados. Essas tarefas podem ser apoiadas por tecnologias como *workflows*, e seus respectivos sistemas de gerenciamento de *workflows* científicos (SGWfC). A linguagem Swift é usualmente utilizada para modelar *workflows* de alto desempenho, pois esta linguagem possui um alto nível de paralelismo nativo, otimizando os recursos do ambiente computacional utilizado. Contudo, esses experimentos científicos estão em constante evolução e o volume de dados consumidos e produzidos tem sido cada vez maior. A execução de qualquer experimento produz diretamente ou indiretamente dados de proveniência. Essas informações estão disponíveis, nas suas fases de composição, execução e análise. Esses dados gerados podem ser registrados ou não, pois nem todas as aplicações apoiam o registro de proveniência. Como esses dados tendem a ser volumosos, lidar com sua extração, manipulação, visualização e tratamento não tem sido uma tarefa trivial. O gerenciamento e manipulação da proveniência de um experimento é importante, por exemplo, para a reprodutibilidade de *workflows* e para a análise de dados que levem a uma descoberta científica importante. O presente trabalho visa apoiar experimentos científicos na fase da análise, por meio do desenvolvimento de um perfilador acoplado a técnicas de aprendizagem de máquinas que apoie o cientista na análise e visualização dos dados relativo ao funcionamento do seu experimento.

## 3. Análise de Dados de Proveniência

Um experimento científico deve possuir um alto grau de reprodutibilidade, pois tendem a ser reexecutados diversas vezes. Os dados gerados pela execução do experimento, são tão valiosos quanto os resultados finais. Em contrapartida, esses dados acabam sendo ignorados por alguns cientistas, por desconhecerem métodos capazes de avaliar tais dados

de maneira eficiente. A proveniência permite que o usuário compreenda os resultados gerados pelo experimento, assim como, também fornece a possibilidade de identificar um problema no experimento e solucioná-lo. Esta proveniência pode ser extraída de duas maneiras, prospectiva e retrospectiva [Freire et al. 2008]. O método prospectivo registra a especificação da tarefa computacional e seus respectivos processos até o término da atividade. A forma retrospectiva funciona registrando detalhadamente os passos de uma tarefa computacional e seu ambiente computacional. Segundo [Mattoso et al. 2008] os experimentos científicos devem gerar resultados de alta qualidade baseando-se na reutilização dos dados explorados a partir dos dados de proveniência. Nestes critérios de experimentos científicos, os processos são registrados com interesse de serem interpretados e gerar resultados que serão de benefício futuro no experimento. A utilização de scripts que realizam aprendizagem de máquina e mineração de dados, tem sido uma excelente companheira a análise desses registros de experimento, dado que produzem métricas a respeito do seu funcionamento entre outros parâmetros específicos definidos pelo cientista, enriquecendo as informações já coletadas. Permitindo rastrear as atividades e arquivos acessados durante sua execução, por exemplo. Os dados de proveniência servem tanto para análise dos experimento, quanto para validação de resultados encontrados, evitando gerar estudos falhos e construir bases de conhecimento sem credibilidade.

### **3.1. Aprendizagem de Máquina**

Técnicas de aprendizagem de máquinas vem sendo empregadas no reconhecimento de padrões estruturados para extração de conhecimento, gerando métodos de classificações eficientes. É possível aplicar técnicas de aprendizagem de máquina diretamente nos dados de proveniência gerados por *workflows* científicos, que levem a um melhor entendimento sobre o comportamento de um experimento e incluso a sua otimização. Essas técnicas podem gerar resultados quanto a sua classificação segundo alguns parâmetros, assim como previsões quanto ao seu funcionamento.

### **3.2. Perfilador de Workflow**

A fim de suprir essa necessidade de análise dos dados de proveniência dos *workflows* científicos, foram surgindo alguns perfiladores. Sendo estes, ferramentas que consomem os dados de proveniência de um *workflow* e a partir de suas análises conseguem traçar um perfil referente ao seu funcionamento. Porém, na maioria das vezes, acabam sendo modelados para efetuar a análise de um *workflow* específico, logo restringe sua análise ao mesmo. Essas ferramentas, necessitam dos dados da execução do *workflow*, existem diferentes maneiras de extrair esses dados, por meio de *scripts* que agem durante ou após sua execução, por exemplo.

## **4. HPSW-Prof**

O HPSW-Prof [Mondelli 2016] é um perfilador baseado na arquitetura de experimentos gerenciados pelo SGWfC Swift [Gadelha et al. 2012] no apoio à manipulação e tratamento de dados de proveniência. Na sua arquitetura, além das camadas de bancos de dados, de proveniência e interface *Web*, neste trabalho foram acopladas eficientemente ferramentas de aprendizado de máquinas para o tratamento estatístico das informações de proveniência relacionadas à execução do *workflow* científico.



Figura 1. Exemplo de funcionamento do HPSW-Prof

O HPSW-Prof é uma arquitetura independente que trabalha após a execução do *workflow*, logo não interfere no desempenho do experimento. Ele faz o uso da proveniência gerada pela execução do *workflow* modelado em Swift, onde são explorados os dados que contém características computacionais e processos que ocorrem durante a aplicação. Em posse desses dados de proveniência, é possível implementar funções preditivas, que podem prever o tempo de execução de um novo experimento baseado no tamanho dos arquivos de entrada. O método de avaliação utilizado é fundamentado na comparação de qualidade dos resultados gerados com os conjuntos de teste e treinamento usados para realizar a regressão estatística. Por exemplo, alguns *workflows* podem ser executados milhares de vezes, e o cientista precisa avaliar as informações de uma maneira comparativa. Sendo assim, o HPSW-Prof apresenta uma ferramenta de análise em larga escala, pois permite navegar por esse grande conjunto de *workflows* e a partir de datas ou algum parâmetro específico, carregar as informações a serem analisadas.

#### 4.1. Implementação

Na implementação do HPSW-Prof, os bancos de dados (*logs* em SQLITE) do Swift contendo as informações de proveniência de execuções de *workflows* foram centralizados em um PostgreSQL e sobre ele foram acopladas funções da linguagem R. Essas funções extraem informações relevantes a análises de perfilamento e geram uma série de gráficos estatísticos. Após, as funções foram integradas ao *framework* Shiny com o objetivo de disponibilizar os resultados através de uma interface Web. Para obter algumas respostas quando a predição do experimento, foi utilizada a biblioteca que conta com scripts de aprendizagem máquina do programa Weka [Hornik et al. 2009].

#### 4.2. Resultados Obtidos

O HPSW-Prof é uma arquitetura que pode ser facilmente acoplada ao *workflow* modelado na linguagem Swift. Esta linguagem de alto nível possui um grande grau de parale-



```

J48 pruned tree
-----

sum_read <= 2517363567487: [ 20.6, 6201.1) (170.0)
sum_read > 2517363567487
| total_apps <= 4560
| | avg_real_secs <= 597.199162: [ 6201.1,12381.5) (3.0)
| | avg_real_secs > 597.199162: [12381.5,18562.0) (6.0)
| total_apps > 4560: [24742.4,30922.9] (9.0/1.0)

Number of Leaves : 4
Size of the tree : 7

```

**Figura 2. Funcionamento do script J48 no HPSW**

lismo nativo, sendo assim, amplamente utilizada por muitos cientistas. A execução desses *workflows*, tem sua proveniência registrada em arquivos log, e importado para algum ambiente de banco de dados. Posteriormente, é realizado o perfilamento estatístico em cima desses dados, que também servem de base para fornecer predições de futuras execuções [Hartmann 2016].

A necessidade da visualização desses dados em uma ferramenta como o perfilador HPSW-Prof, se dá pela facilidade da compreensão ao visualizar os dados. Na *Tabela 1* por exemplo, podemos identificar os dados de proveniência produzidos pela execução de um *workflow*. Ao tentar associar e comparar os parâmetros gerados por cada aplicação, encontra-se alguma dificuldade. Assim, confirmamos a importância de ferramentas que auxiliem uma melhor visualização desses dados de proveniência. Deve ser levado em conta que alguns experimentos executam diversas atividades, logo a quantidade de resultados a serem analisados torna-se muito grande, tomando cada vez mais tempo para compreensão dos resultados.

App Name	AVG Duration	AVG Percent cpu	AVG Fs Writes	AVG Fs Reads	AVG Max Rss
FragHits	2.65	97.00	420.00	77.99	10468.67
Combine Frags	0.15	100.00	424.00	0.80	1370.14
Csv Generator	0.09	0.00	32.00	0.00	1407.81
Filter Hits	0.13	50.00	688.00	222.11	516.09
Hits	0.88	98.50	23904.00	213.21	391193.43
Reverse Complement	0.17	98.00	1256.00	0.00	12441.30
Sort Hits	0.44	148.50	23904.00	468.42	391355.91
Sort Words	0.52	186.00	30064.00	0.00	235100.05
w2hd	0.88	99.00	50104.00	2438.80	525.45
Words	0.32	97.00	30056.00	0.00	475.90

**Tabela 1. Exemplo dos dados gerados após a execução de um *workflow*.**

Com os resultados expostos na interface gráfica, como podemos observar na *Figura 2*, torna-se compreensível e agradável a visualização do funcionamento e duração de cada atividade, mostrando o próprio comportamento do *workflow* ou algum potencial problema de paralelização ou erro no código. Logo, o cientista será capaz de entender de maneira clara as atividades que compõem o seu experimento com os resultados gerados.

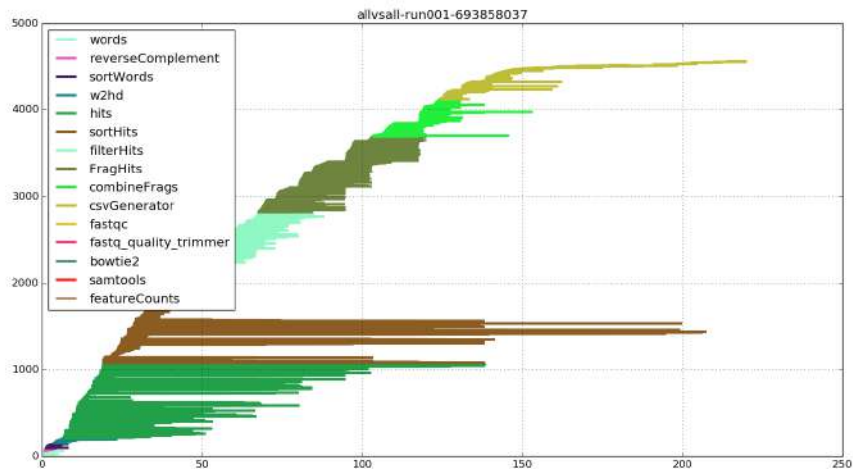


Figura 3. Exemplo de gráfico gerado com HPSW-Prof

## 5. Conclusão

Neste breve trabalho, apresentamos algumas definições de análise proveniência, como a prospectiva, retrospectiva e suas respectivas diferenças. Além das características existentes dentro dos dados de proveniência e sua influência nos resultados obtidos, assim como as diferentes técnicas que podem ser implementadas para predição desses dados.

O HPSW-Prof é uma ferramenta que foi desenvolvida no intuito de apoiar o perfilamento estatístico de experimentos de alto desempenho. Proporcionando a visualização relacionada à execução desses experimentos em interface web, viabilizando o cientista identificar alguma falha de paralelização e uma melhor compreensão do *workflow*. Como o perfilador inclui técnicas de aprendizagem de máquina, torna-se possível a predição de tempo consumido na execução do *workflow*, a partir de alguns parâmetros. Deve ser levado em conta que alguns experimentos podem perdurar de por um longo período de execução e obter um estimativa da duração é de grande valia para o cientista. Uma ferramenta como o perfilador é de grande importância para análise dos dados de proveniência, já que esses dados podem ser utilizados no rastreamento das ações de um experimento, possibilitando a identificação de erros, funcionamento dos componentes computacionais e reprodutibilidade dos experimentos para verificar os resultados gerados.

## Referências

- Freire, J., Koop, D., Santos, E., and Silva, C. T. (2008). Provenance for computational tasks: A survey. *Computing in Science & Engineering*, 10(3).
- Gadelha, L. M., Wilde, M., Mattoso, M., and Foster, I. (2012). Mtcprow: a practical provenance query framework for many-task scientific computing. *Distributed and Parallel Databases*, pages 1–20.
- Hartmann, H. (2016). Statistics for engineers. *Communications of the ACM*, 59(7):58–66.
- Hornik, K., Buchta, C., and Zeileis, A. (2009). Open-source machine learning: R meets weka. *Computational Statistics*, 24(2):225–232.

- Mattoso, M., Werner, C., Travassos, G., Braganholo, V., and Murta, L. (2008). Gerenciando experimentos científicos em larga escala. *SBC*, page 121.
- Mondelli, M. L. ; Tonelli, M. S. . O. K. . G. J. L. M. R. . V. A. T. R. (2016). HPSW-Prof: A provenance-based framework for profiling high performance scientific workflows. *Simpósio Brasileiro de Banco de Dados*, pages 117–122.

**Título:** Controle ótimo de vigas de Euler - Bernoulli

**Bolsista:** Monique Ribeiro da Costa.

**Orientador:** Jaime Edilberto Muñoz Rivera.

Bolsista de Iniciação Científica- CNPq.

### **Objetivo:**

Formular o problema de controle ótimo de uma viga no modelo de Euler Bernoulli, onde minimiza-se um funcional quadrático que otimiza a deformação de acordo com uma função objetivo.

### **Introdução:**

Muitas estruturas são construídas usando vigas, as quais defletem ou distorcem por decorrência do próprio peso ou de uma força externa, podendo ocasionar ruptura dependendo da intensidade da força exercida ou pela constituição de seu material. Simularemos esta viga de comprimento  $L$  usando o modelo de Euler Bernoulli. Este modelo é baseado no seguinte princípio:

***“A posição de equilíbrio dos corpos elásticos é aquela que minimiza sua energia potencial elástica.”***

Usando este princípio obtemos que a posição de equilíbrio da viga definida pelo gráfico da função  $u$  que é solução da seguinte equação:

$$\alpha u + \beta u_{xxxx} = f \quad (1)$$

Verificando a condição de contorno:

$$u(0) = u_x(0) = 0, u(L) = u_x(L) = 0 \quad (2)$$

Sendo  $f$  a força distribuída que está atuando sobre a viga. Assim, a função  $u$  será aquela cujo gráfico define a posição de equilíbrio da viga. Então para cada  $f$  temos uma posição de equilíbrio  $u$ .

O problema de controle de vigas consiste em controlar a solução  $u$  da equação descrita em (1). Deve-se escolher a força  $f$  de tal forma que a posição de equilíbrio esteja mais parecida com a função objetivo  $Z_d$ . Isto é, como deve ser  $f$  de tal forma que  $u(x) - z_d(x)$  seja o menor possível. Aqui aparece um problema de minimização sobre o conjunto de todas as possíveis forças distribuídas que podem agir sobre a viga.

## **Métodos:**

Utilizaremos os métodos dos espaços de Hilbert e de minimização sobre estes espaços. Estudaremos o problema de controle ótimo da equação (1). O problema consistem em determinar a força que deve ser concentrada acima da viga de tal forma que a viga esteja próxima da nossa função objetivo. O funcional custo será dado por um funcional quadrático e as forças atuantes sobre a viga serão os controles.

O problema tende a encontrar o controle ótimo que minimiza o funcional custo, e a caracterizá-lo de tal forma que seja possível calcular numericamente, para isto precisa-se encontrar o sistema de otimalidade que consiste na equação de estado e a equação do sistema adjunto.

A solução numérica do problema será desenvolvida usando o método de diferenças finitas baseado nas diferenças centrais, método a qual aproxima-se a derivada de uma função pela sua diferença.

## **Resultados e Discussão:**

Encontrando o sistema de otimalidade que consiste na equação de estado e a equação do sistema adjunto, calcularemos sua solução numericamente, utilizando o método de diferenças finitas, resolveremos assim o sistema discreto computacionalmente, utilizando o método de eliminação de Gauss, em um programa feito em MATLAB<sup>®</sup>, obtendo desta forma os valores numéricos da deformação da viga em cada ponto da discretização dela. Utilizando a plataforma gráfica do MATLAB<sup>®</sup> para simular as deformações.

## **Conclusões:**

- 1) Temos deduzido o modelo da viga de Euler Bernoulli usando argumentos físicos da teoria de elasticidade.
- 2) Calculamos soluções aproximadas do modelo de Euler Bernoulli usando diferenças finitas.
- 3) Definimos o problema de controle ótimo para vigas unidimensionais onde tentamos controlar através de um funcional quadrado a deformação de viga.
- 4) Concluimos que para funcionais devidamente "coercivos" o problema de controle ótimo possui existência e unicidade.

## Referências bibliográficas:

RIVERA, JAIME . Título:Cálculo Difrencial II e equações diferenciais.

RIVERA, Jaime Edilberto Muñoz. Métodos Computacionales Aplicados a las Ecuaciones Difrenciale.

Editora Thomsom.

ZILL, DENNIS G. Equações Diferenciais com Aplicações em Modelagem. XXX: Thomson, 2003.

SOUZA, Salete. Disciplina: Resistência dos materiais. Escola de engenharia metalúrgica de Volta Redonda. Universidade Federal Fluminense sd. Disponível em <<http://www.professores.uff.br/salete/res1/aula10.pdf>>

## **Relatório de Atividades**

### **Programa Institucional de Bolsas de Iniciação Científica - PIBIC/LNCC**

#### 1) Dados gerais

Introdução à Computação Quântica

Paulo Ricardo Borré Reis

Renato Portugal

PIBIC – março de 2017 a julho de 2017

#### 2) Objetivos

Este trabalho de iniciação científica teve como objetivo estudar aspectos fundamentais da computação quântica e seus aspectos algorítmicos, reforçando a formação nas áreas de Álgebra Linear e programação em Python, que são necessárias para o entendimento e simulação de algoritmos quânticos. As principais metas são:

- Domínio dos conceitos de Álgebra Linear
- Capacidade de programação em Python
- Entendimento e interpretação dos postulados e princípios da mecânica quântica
- Entendimento e simulação de alguns algoritmos quânticos importantes.

#### 3) Introdução

A Computação Quântica é uma área de pesquisa nova, que está em ascensão, e que utiliza elementos de várias outras áreas do conhecimento, como Matemática, Física e Ciência da Computação. Um computador quântico é um dispositivo de computação que faz uso direto dos fenômenos da Mecânica Quântica, tais como sobreposição (ou superposição) de estados e emaranhamento, para realizar operações sobre dados.

O início da computação quântica deu-se nos anos 80 com os trabalhos de Benioff [1] e Feynman [2]. Feynman foi o primeiro a propor a utilização de fenômenos quânticos para executar rotinas computacionais. Ele mostrou que um computador tradicional levaria um tempo extremamente longo para simular um simples experimento de física quântica. Isso deve ao fato de que a dimensão do espaço de Hilbert associado ao sistema cresce

exponencialmente em função do número de partículas acrescentadas ao mesmo. Por outro lado, sistemas quânticos podem executar enormes quantidades de cálculos num curto espaço de tempo. É possível utilizar essa capacidade para se calcular algo útil?

Os argumentos de Feynman estimularam David Deutsch a generalizar o modelo mais fundamental da computação clássica, a saber, a máquina de Turing, para o seu equivalente quântico num trabalho histórico de 1985 [3]. Ele mostrou que da mesma forma que uma Máquina de Turing pode simular outra máquina de Turing eficientemente, um computador quântico universal é capaz de simular o funcionamento de outro computador quântico com complexidade, no máximo, polinomial. Isso fez crescer a esperança de que um dispositivo simples seja capaz de executar muitos algoritmos quânticos diferentes. Posteriormente, o autor também generalizou o modelo de circuitos baseado em portas lógicas. Nesse novo contexto, operadores unitários tomaram o lugar das usuais portas lógicas AND, OR e NOT [4].

#### 4) Metodologia

A metodologia utilizada nesta pesquisa consistiu-se, basicamente, em estudos dirigidos e reuniões com os orientadores. O estudo dirigido em Álgebra Linear ocorreu por meio de consultas ao acervo do LNCC, bem como por livros presentes no Laboratório do Grupo de Computação Quântica. Em especial, foi utilizado como base para os estudos o livro *Linear Algebra*, de Serge Lang [8].

Para o estudo dirigido em linguagem Python foi consultado um manual introdutório disponível online, enquanto que foram realizadas aplicações práticas unindo a Álgebra Linear com programação na linguagem Python. Foram assim escritos algoritmos para a obtenção de Determinantes, autovalores, autovetores, etc.

Foram ainda realizados estudos dirigidos não previstos no cronograma inicial, estes em Teoria de Grafos. O estudo se deu por meio de consulta ao acervo do LNCC, bem como ao livro-texto *Grafos – Uma Introdução*, de Samuel Jurkiewics [9]. Como forma de averiguação dos estudos, foram elaborados alguns algoritmos que para trabalhar determinadas propriedades dos Grafos. Foi dada atenção especial a um algoritmo de busca, que consistia em procurar em um grafo qualquer  $G$  por um grafo diamante como subgrafo de  $G$ . Esse algoritmo retorna *True* caso  $G$  não contenha nenhum grafo diamante e *False* caso contrário.



## 5) Resultados e Discussão

Durante a realização da pesquisa de iniciação científica foi possível um grande aprofundamento do conhecimento em Álgebra Linear e da linguagem Python, alcançando-se estes objetivos de trabalho.

Com relação à Álgebra Linear foram aprofundados conhecimentos em espaços vetoriais, transformações lineares, matrizes e determinantes, todos estes conhecimentos sendo aplicados em algoritmos na linguagem Python.

Já com relação à linguagem em si, foram aprofundadas questões de sintaxe e semântica da linguagem, bem como em estruturas de repetição, estruturas de decisão, tratamento de erros, funções, listas e a lógica de programação.

Em razão do período curto de tempo para realização do projeto inicial (visto que a pesquisa se iniciou no mês de março), não foi possível alcançar os objetivos pretendidos que se relacionavam à mecânica quântica e aos algoritmos quânticos. Contudo, foram realizados estudos em Teoria de Grafos, que inicialmente não estavam previstos no cronograma da pesquisa.

Com relação à Teoria de Grafos foram aprofundados os conceitos de propriedades gerais dos grafos, como isomorfismo, caminhos e circuitos, subgrafos, etc. Também foram elaborados algoritmos em linguagem Python para simular os conceitos aprendidos.

## 6) Conclusões

No decorrer do trabalho de pesquisa foi possível perceber a importância do estudo e aprofundamento em Álgebra Linear e Teoria de Grafos para a compreensão da Computação Quântica e Algoritmos Quânticos.

Também foi possível perceber que o maior progresso nos estudos se deu apenas com execução prática de atividades propostas pelo orientador. Em especial, os conhecimentos da linguagem Python só foram aprendidos de forma significativa na execução das atividades de criação de algoritmos para simular os conceitos de Álgebra Linear e Teoria de Grafos, fato este que levou a uma melhor compreensão destas demais áreas.

## 7) Referências bibliográficas

- [1] P. Benioff. The Computer as a Physical System: A Microscopic Quantum Mechanical Hamiltonian Model of Computers as Represented by Turing Machines. *Journal of Statistical Physics*, 22, 1980.
- [2] R. Feynman. Simulating Physics with Computers. *International Journal of Theoretical Physics*, 21:467–488, 1982.
- [3] D. Deutsch. Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer. *Proceedings of the Royal Society of London, A* 400:97–117, 1985.
- [4] D. Deutsch. Quantum Computational Networks. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 425(1868):73–90, 1989.
- [5] N. Koblitz. *Algebraic aspects of cryptography*. Berlin ; New York : Springer-Verlag, Rio de Janeiro, 1998.
- [6] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2003.
- [7] P. W. Shor. *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*. *SIAM J. Comput.*, 26(5):1484–1509, 1997.
- [8] S. Lang. *Linear Algebra*. Third Edition. Springer-Verlag. New York. 1987
- [9] S. Jurkiewicz. *Grafos – Uma Introdução*. 2009. Disponível em <http://www.obmep.org.br/docs/apostila5.pdf>. Acesso em 04/07/2017

# Relatório Iniciação Científica – LNCC

Rayssa Barreto Miranda da Rosa

## 1 Dados gerais

- Título do projeto: Análise de dados de atendimento da área de saúde com base em redes complexas.
- Nome do bolsista: Rayssa Barreto Miranda da Rosa.
- Nome do orientador: Artur Ziviani (LNCC).
- Tipo de bolsa e período do relatório: Iniciação Científica (03/2017-06/2017).

## 2 Objetivos

Na área de gestão do sistema de saúde, há uma tendência recente de análise de dados de atendimento com base na teoria de redes complexas [1, 2, 3]. O objetivo de trabalhos nessa linha é o estudo e a melhor compreensão da organização e operação de serviços de saúde, ao mesmo tempo buscando uma melhor coordenação de cuidados e uso de recursos de forma centrada no paciente.

Neste trabalho é proposta a análise e visualização de fluxos de pacientes através de seus atendimentos de saúde usando técnicas de ciência de redes e tendo por base dados de atendimento básico de saúde em uma grande metrópole brasileira [4].

Dado o curto período de bolsa até o presente (4 meses), foi analisado até o momento a bibliografia relacionada e atualizada com os artigos relevantes dos anos de 2016 e 2017. Essas atividades estão alinhadas e em consonância com o cronograma proposto.

## 3 Introdução

A interconectividade dos diversos componentes que englobam as redes de atenção à saúde (unidades de saúde, tipos de serviço prestados, profissionais e pacientes) tipicamente gera um sistema muito complexo, tendo em vista a diversidade e densidade destas interações, dificultando a análise do comportamento desta rede. Nesse contexto, observa-se uma tendência na utilização de técnicas de modelagem e análise de redes complexas na área da saúde, não apenas para modelos de propagação de doenças, mas também as relacionando à organização dos serviços de saúde, visando a melhor coordenação de cuidados e uso de recursos de forma centrada no paciente [1, 2, 3].

## 4 Material e Métodos ou Metodologia

A primeira fase do plano de trabalho da bolsa de IC correspondia a um estudo bibliográfico da área. Para fazer esse estudo bibliográfico foram analisados os artigos sobre redes complexas, em especial as dinâmicas, em que é possível perceber diferentes fluxos de pacientes na rede.

## 5 Resultados e Discussão

Durante a análise das redes complexas, foram selecionados 11 artigos publicados entre 2016 e 2017 que têm relação com redes complexas dinâmicas na área da saúde. Com base nesses artigos, foi realizada a atualização da bibliografia usada no projeto.

## 6 Conclusões

Nesta primeira etapa do plano de trabalho previsto para a bolsa de IC foi possível analisar a bibliografia recente no que se refere às redes complexas no atendimento na área da saúde. Para isso, foi feito também um estudo das técnicas de caracterização das redes dinâmicas.

Concluída a primeira etapa do plano de trabalho, o objetivo das atividades propostas visam a realização de análises básicas próprias e suporte a pós-graduandos em procedimentos de modelagem e análise mais complexos em dados de atendimento de saúde com base em técnicas de ciência de redes, incluindo o estudo de técnicas e realização de métodos para visualização desses dados e análises.

## Referências

- [1] C. E. Pollack, G. E. Weissman, K. W. Lemke, P. S. Hussey, and J. P. Weiner, “Patient sharing among physicians and costs of care: A network analytic approach to care coordination using claims data,” *Journal of General Internal Medicine*, vol. 28, no. 3, pp. 459–465, 2013.
- [2] M. Ito, L. G. Moyano, A. P. Appel, and V. F. Santana, “Análise do relacionamento na comunidade de médicos de seguradoras de saúde,” in *Anais do XV Congresso Brasileiro de Informática em Saúde (CBIS)*, 2016.
- [3] M. L. Barnett, B. E. Landon, A. J. O’Malley, N. L. Keating, and N. A. Christakis, “Mapping physician networks with self-reported and administrative data,” *Health Services Research*, vol. 46, no. 5, pp. 1592–1609, 2011.
- [4] L. O. Jasmim, A. Ziviani, M. Ito, and P. Bandiera-Paiva, “Caracterização de atendimentos em uma rede de atenção à saúde,” in *Anais do Workshop de Informática Médica (WIM)*, Julho 2017.

## Relatório Bolsa Iniciação Científica - IC - PIBIC

Nome do bolsista: Rodrigo Botelho de Souza

Nome do Orientador: Dr. Fábio Porto

Processo: 122071/2016-6

Modalidade - Categoria: Iniciação Científica - IC

Período da Bolsa: De 01/08/2016 a 31/07/2017

### Paralelização de análises sobre dados de simulação utilizando o paradigma MapReduce

Vivemos um período de grande aumento na capacidade de armazenamento de dados eletronicamente, estamos na era dos “Zettabytes”. Contudo, esse aumento da capacidade de armazenamento não foi acompanhado pelo aumento da capacidade de leitura desses dados, acarretando um fator altamente desigual, pois em grandes volumes de dados perdemos muito tempo no processo de leitura, isso se utilizando de técnicas de processamento concentrado. Para contornarmos este problema utilizaremos técnicas e ferramentas de processamento distribuído.

Diferentes projetos de astronomia utilizam-se de distintos equipamentos telescópios, tais equipamentos possuem configurações com relevantes diferenças como poder de visão, calibragem, tecnologia entre outras, esses telescópios mapeiam determinadas áreas do espaço obtendo seus corpos celestes em prol de seus projetos, com toda essa gama de projetos e seus catálogos relativos aos corpos celestes faremos a junção dessas diferentes áreas, contudo encontramos o problema de objetos celestes correspondente em áreas de fronteira com outros projetos.

O estudo consiste nas técnicas para particionar, processar e, finalmente, integrar os dados, tornando seu resultado final íntegro. Ocorre que estratégias de particionamento de dados tipicamente agem de duas formas: (i) divisão física, em blocos de dados de um certo tamanho, (ii) particionamento de valores em uma chave de identificação. Em vários cenários, no entanto, o critério de particionamento se apresenta de forma mais complexa. O particionamento por si, pode ser visto como um processo em duas etapas: a definição do critério de particionamento, e identificação dos dados com as respectivas partições; e a alocação das partições sobre o sistema distribuído. Desta forma, este trabalho apresenta um dataflow para particionamento e alocação de dados em um sistema distribuído.

O desenvolvimento se compreendeu na preparação de um ambiente distribuído de alto desempenho com a utilização do banco de dados em memória Apache Ignite, sendo composto do sistema operacional Centos OS 6, também possuímos a framework Spark atuando como ponto de chamada e processamento em nosso trabalho. Na etapas que compreendem nosso dataflows temos a entrada dos dados via HDFS(Hadoop distributed filesystem), posteriormente carregamos esses dados ao

sistema de gerenciamento de bancos dados Ignite para então realizarmos o processamento via Spark.

Concluimos neste trabalho que as técnicas de particionamento de dados usufruindo das frameworks apache spark e apache ignite, se mostraram com alto ganho de desempenho, gerando no processamento de nossos dataflows maior velocidade nas suas análises. Fizemos uso de uma ferramenta desenvolvida em nosso laboratório FRANCE(fragmentador de catálogos espaciais) um algoritmo iterativo para particionar dados em espaços unidimensionais.

Objetivo alcançados:

Construção do ambiente de alto desempenho

Desenvolvimento da Aplicação para particionamento de dados distribuídos

Desenvolvimento de aplicações em MapReduce(Hadoop)

Estudo e Desenvolvimento de técnicas de particionamento de dados

Desenvolvimento de técnicas para armazenamento de banco de dados em memória (Apache Ignite)

### **Atividades**

- Desenvolvimento de aplicações para banco de dados em memória(apache Ignite) com processamento distribuído via apache Spark;
- Desenvolvimento de aplicação framework Hadoop(Map-Reduce, Hdfs);
- Desenvolvimento ambientes de alto desempenho distribuídos GNU/Linux;
- Estudo de novas framework para Big data;
- Estudo do banco de dados Postgresql;
- Desenvolvimento e implementação do ambiente de alto desempenho, virtualização;

### **Cursos**

- LNCC - Curso de verão Sistemas de controle de versão – SVN e Git
- LNCC - Curso de verão Python: Visão geral da linguagem com aplicações numéricas
- LNCC - Curso de verão Programação em Shell Script usando o Bash
- LNCC - Curso de verão Análise de Dados
- LNCC - Curso de verão Gerência de Grandes Volumes de Dados
- LNCC - Introdução à Programação com Swift: Workflows Científicos Paralelos

### **Palestras**

- LNCC - the practical challenges of managing big data in the cloud

- LNCC - Big data to tackle Urban Mobility Challenges

# INVESTIGAÇÃO DE POLIMORFISMOS DE NUCLEOTÍDIO ÚNICO INDUZIDOS PELA AÇÃO DA POLIMIXINA B E EFEITOS ABIÓTICOS EM *KLEBSIELLA PNEUMONIAE*

## Laboratório Nacional de Computação Científica.

Bolsista: Thiago Cardoso Pereira Carneiro

Orientadora: Dra. Marisa Fabiana Nicolás

Tipo de Bolsa:

Período:

### Resumo

SNPs ou Polimorfismos de nucleotídeo único são mutações que podem ser caracterizadas como a variação da sequência observada em um único nucleotídeo no genoma e isso pode incidir de forma direta e relevante na formação das proteínas. Mutações causadas em bactérias por conta do uso inadequado de antibióticos ao longo dos anos causou um aumento no número de bactérias resistentes a antibióticos, principalmente a Betalactâmicos. A *Klebsiella pneumoniae* se caracteriza como sendo uma bactéria gram-negativa e possui como uma de suas principais características a resistência a antibióticos como a polimixina B (colistina B), última saída para tratamentos quando se trata de bactérias resistentes as betalactâmicos. Embora haja estatísticas insuficientes, evidências acumuladas ainda sugerem que a resistência as cefalosporinas de espectro estendido em *E. coli* e, bem como em *K. pneumoniae* tornaram-se um problema mundial [1]. Tendo em vista a importância do tema, o presente estudo tem como proposta investigar os SNPs induzidos pelo efeito da polimixina B conjuntamente com efeitos abióticos, realizando também a classificação dos SNPs em sinônimos e não sinônimos, buscando também genes ligados a resistência da bactéria. Os dados de transcriptoma analisados de *Klebsiella pneumoniae subsp. pneumoniae Kp13* foram gerados através de sequenciamento de nova geração (NGS) utilizando a técnica de Rna-seq nas condições induzida (polimixina B) e concentrações elevadas de Fe e Ca, baixo Mg e pH (5,8). A investigação foi feita através de ferramentas de bioinformática.

**Palavras chave:** *Klebsiella Pneumoniae*, SNP's, Rna-Seq, Polimixina B.

### OBJETIVOS GERAIS



Investigar os polimorfismos de nucleotídeo único (SNPs) que possam estar ocorrendo em razão da ação da polimixina e fatores abióticos (Ph, Fe, Mg e Ca) na bactéria *Klebsiella pneumoniae* subsp.kp13 utilizando dados gerados através da técnica de RNA-seq em conjunto com ferramentas de bioinformática.

#### **Objetivos específicos:**

- Realizar o mapeamento dos transcritos.
- Quantificar os SNPs encontrados.
- Classificar os SNPs em sinônimos e não sinônimos.
- Encontrar SNPs não sinônimos ligados a resistência.

## **INTRODUÇÃO**

Com a utilização incorreta e/ou indiscriminada de antibióticos ao longo dos anos o número de bactérias resistentes a eles tem se tornado cada vez maior e a problematização em cima da falta de controle das infecções se agrava cada vez mais, nos últimos cinco anos, pode-se observar um aumento dramático nos casos de bactérias resistentes [2]. Podemos citar os casos de bactérias resistentes a polimixina, utilizada como último recurso contra infecções causadas por bactérias multiresistente, como é o caso da gram-negativa *Klebsiella pneumoniae*, que possui mecanismos efetivos próprios contra as polimixinas em conjunto com sua facilidade de transferir material genético. *Klebsiella pneumoniae* são produtoras da chamada KPC (*Klebsiella pneumoniae* carbapenemase), enzima produzida por bactérias gram-negativas que conferem resistência a um amplo espectro de  $\beta$ -lactâmicos e aos antimicrobianos carbapenêmicos. Essa bactéria possui ainda a complexa combinação da impermeabilidade da membrana com betalactamases cromossômicas (AmpC) ou de amplo espectro [3], alvo deste estudo, foi causadora de um surto de infecção hospitalar em 2009 no sul do país.

Os chamados polimorfismos de nucleotídeo único (SNPs) podem ser causadores de alterações relevantes na formação das proteínas; podem ser usados para caracterizar tipos de estirpes de bactérias e indicadores sitio específico para estudo sobre mutações e já foram detectados desde os primeiros sequenciamentos de fragmentos específicos [4]. O estudo sobre mutações é de grande importância para elucidar mecanismos de resistência das bactérias, já se sabe que essas mutações ocorrem sobre stress pelo uso de antibiótico. Em um estudo recente *Pseudomonas aeruginosa* foi revelado que uma cultura em contato com ferro livre e agentes anti-microbianos elevam a taxa de mutação de seus genes [5]. Se torna evidente a importância de estudos sobre mutações em genes contidos em bactérias resistentes. A abordagem apresentada neste estudo dos SNPs se tornou possível graças a uma revolução ocorrida na década de 80 e que uniu pelo menos dois grandes campos de estudo: a biologia e a computação, possibilitando o surgimento da chamada bioinformática que inicialmente, foi tida como um paradoxo entre o mundo real e o imaginário, já que

parecia remoto que biólogos pudessem empregar ferramentas computacionais para desenvolver estratégias a serem empregadas em trabalhos científicos [6]. Desde então as técnicas de sequenciamento, armazenamento e manipulação dos dados foram refinadas e ao longo dos anos resultou em maior precisão, aumento do volume de dados gerados e possibilitou o aumento do alcances dos campos de estudo dentro da genômica. Baseado nos chamados sequenciamentos de nova geração (NGS) a técnica para análise de transcriptoma RNA-seq surge em 2009 [7] permitindo que faça a análise da expressão gênica, este foi utilizado como base para criação dos dados analisados neste trabalho.

## **METODOLOGIA**

A metodologia utilizada foi realizada utilizando ferramentas de bioinformática a partir dos dados do transcriptoma da bactéria *K. Pneumoniae*. O dados do transcriptoma foram gerados por sequenciamento de nova geração (NGS) através de plataforma illumina Hiseq-2500. As sequências foram trimadas (eliminando erros e sequências de baixa qualidade) através do programa Skewer [8]. O alinhamento dos trimados foi realizado através do programa bowtie2 [9]. Os arquivos de saída com extensão .sam foram convertidos através do programa samtools [10] em formato bam afim de se otimizar a análise, diminuindo o tempo e custo computacional. A chamada dos SNPs foi feita a partir dos arquivos obtidos na etapa anterior pelo programa Genome Analysis Toolkit [11]. Para finalização dos SNPs obtidos, os arquivos foram submetidos ao programa SNPeff em conjunto com um script em python para possibilitar que os arquivos de saída pudessem obter um novo campo contendo a classificação da sua incidência em uma região intergênica ou codificante, caso caindo em região condificante, fosse classificada pelo tipo de mutação (sinônima ou não sinônima).

Foram criados diagramas de venn referentes a quantidade de SNPs encontrados, assim como a incidência de sinônimos e não sinônimos. A anotação de cada gene e proteína foi feito através de coordenadas e locus tag fornecidas através dos resultados obtidos pelo programa SnpEff e foram verificadas através do arquivo gff (anotado) referência e através do site NCBI. Ao final foram obtidas tabelas para cada condição com todos os Snp já contendo sua posição, seu efeito, gene e proteína expressa.

## **RESULTADOS**

O resultado do sequenciamento gerou um total de 23.353.870 de reads, após a utilização do programa skewer esse número reduziu para 23.169.374. O mapeamento das 5 bibliotecas alvo (CLSI, Ca, pH, Mg e Fe) contra o cromossomo gerou uma média de 95,10%.

A busca por SNPs a partir do mapeamento do cromossomo utilizando a plataforma GATK revelou um total de 1654 SNPs no cromossomo distribuídos entre as 5 condições, com o filtro de qualidade do programa, 599 deste total foi desconsiderado. Foram encontrados 27 SNPs em comum entre todas as condições com Mg como a condição que mais obteve SNPs.

Entre os plasmídeos da bactéria, foram obtidos um total de 468 SNPs já com o filtro do programa.

A classificação do programa SNPeff revelou a incidência de 377 SNPs sinônimos e 537 não sinônimos ocorrendo no cromossomo e 189 sinônimos e 159 não sinônimos ocorrendo nos plasmídeos.

Como objetivo específico do trabalho, foram encontrados SNPs com efeito não sinônimo ocorrendo em importantes genes ligados a mecanismos de resistência e dentre eles, o gene *bla* SHV, mutações nesse gene conferem um espectro de hidrólise estendido, codificando novas enzimas denominadas  $\beta$ -lactamases de espectro estendido (ESBLs) (JACOBY e BUSH, 200). O gene *bla* SHV foi encontrado em todas as condições. Foram encontrados polimorfismos nos genes *acrA* e *acrB*, onde a ocorrência de mutações está relacionada ao aumento da atividade de bomba e conseqüentemente aumento do efluxo dos antimicrobianos [12] e o gene *marA*, também relacionado ao aumento de atividade da bomba e aumento da resistência aos antimicrobianos [13]. Na condição Mg, foi encontrado polimorfismo no gene *phoP*, envolvido com o sistema de dois componentes e que pode ativar a resistência à polimixina B sob baixo teor de magnésio como mostra um estudo em *S. enterica* [14].

## CONCLUSÃO

O trabalho cumpriu todos os objetivos, sendo encontrados Snps em todas as condições através da investigação. O mapeamento foi realizado com sucesso e sem erros proporcionando o sucesso das próximas etapas e conseqüentemente a conclusão dos demais objetivos.

A investigação revelou a ocorrência da quantidade total de 1055 Snps indicando a indução de Snps em todas as 5 condições de stress, mostrando que Mg foi a condição que gerou a maior ocorrência de polimorfismos.

Os resultados da classificação e quantificação dos efeitos sinônimo e não sinônimo revelaram uma maior incidência de Snps não sinônimos.

Por fim, a busca por Snps em genes ligados a resistência apontou que polimorfismos com efeito não sinônimos em genes ligados a bombas de efluxo multidrogas, sistemas de 2 componentes e  $\beta$ -Lactâmicos. Concluindo a investigação, que demonstra que a ação da

polimixina e dos fatores abióticos induzem não só a uma expressiva quantidade de Snps na bactéria, como também afetam genes que aumentam sua resistência.

## REFERÊNCIAS

1-Giske, Christian G. et al. **Clinical and Economic Impact of Common Multidrug-Resistant Gram-Negative Bacilli**. Antimicrobial Agents and Chemotherapy, 2008.

2-Kaczor AA, Polski A, Sobótka-Polska K, Pachuta-Stec A, Makarska-Bialokoz M, Pitucha M. **Novel antibacterial compounds and their drug targets - successes and challenges**. Curr Med Chem, 2016.

3-Ramos. P.i.P; Picão, C.R; Almeida, P.G.L ;Lima, B.C.N; Girardello, R;Vivian, P.C.A ; Xavier, E.D ; Barcelos,G.F; Pelisson,M; Vespero,C.E; Médigue, C; Vasconcelo ,R.T.A ; Gales, A.C ; Nicolás, F.M. **Comparative analysis of the complete genome of KPC-2-producing Klebsiella pneumoniae Kp13 reveals remarkable genome plasticity and a wide repertoire of virulence and resistance mechanisms**. BMC Genomics, 2014.

4-Orita, M, Iwahana H, Kanazawa, H, et al. **Detection of polymorphisms of human DNA by gel electrophoresis as single strand conformation polymorphisms**. Proceedings of the National Academy of Sciences USA, 1989.

5-Rodríguez-Rojas, Alexandro et al. **Cationic Peptides Facilitate Iron-Induced Mutagenesis in Bacteria**. Ed. Ivan Matic. PLoS Genetics, 2015.

6-De Araújo ND, de Farias RP, Pereira PB, de Figueirêdo FM, de Moraes AMB, Saldanha LC, Gabriel JE. **The effects and applications of bioinformatics on the biomedical area**. Estud Biol, 2008.

7-Zhong Wang, Mark Gerstein, and Michael Snyder. **RNA-Seq: a revolutionary tool for transcriptomics**. Nature Reviews Genetics, 2009.

8-Jiang, H; L, R; Weil-ding. S; Zhu, S. **Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads**. BMC Bioinformatics, 2014.

9-Ben Langmead and Steven L Salzberg. **Fast gapped-read alignment with Bowtie 2**. Nature Methods, 2012.

10-Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. **The Sequence Alignment/Map format and SAMtools**. Bioinformatics, 2009.

11-Geraldine A. Van der Auwera, Mauricio O. Carneiro, Chris Hartl, Ryan Poplin, Guillermo del Angel, Ami Levy-Moonshine, Tadeusz Jordan, Khalid Shakir, David Roazen, Joel Thibault, Eric Banks, Kiran V. Garimella 1 , David Altshuler, Stacey Gabriel, and Mark A. DePristo. **From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline**, Curr Protoc Bioinformatics, 2014.

12-Padilla E, Llobet E, Doménech-Sánchez A, Martínez-Martínez L, Bengoechea JA, Albertí S. **Klebsiella pneumoniae AcrAB efflux pump contributes to antimicrobial resistance and virulence**. Antimicrob Agents Chemother. 2010.

13-Ferrari RG, Galiana A, Cremades R, Rodriguez JC, Magnani M, Tognim MC, et al. **Expression of the marA, soxS, acrB and ramA genes related to the AcrAB/ToIC efflux pump in Salmonella enterica strains with and without quinolone resistance-determining regions gyrA gene mutations**. The Brazilian journal of infectious diseases : an official publication of the Brazilian Society of Infectious Diseases, 2013.

14-Perez JC, Groisman, EA: **Acid pH activation of the PmrA/PmrB two-component regulatory system of Salmonella enterica**. Molecular microbiology, 2007.

## 1. Dados gerais

### **Aplicação das ferramentas Intel Parallel Studio para modernização de código para métodos numéricos de diferenças finitas para solução de equações diferenciais parciais em arquitetura Intel Haswell/Broadwell.**

Bolsista: Gabriel Pinheiro da Costa

Orientador: Carla Osthoff Barros

Co-orientador: Frederico Luís Cabral

Bolsa de Iniciação Tecnológica pelo CNPq

Esse relatório cobre as atividades realizadas entre os dias 14 de abril de 2017 até 03 de julho de 2017

## 2. Objetivos

O objetivo é o desenvolvimento do conhecimento na área de processamento de alto desempenho para avaliar e otimizar o desempenho de uma aplicação de computação científica em uma máquina de arquitetura paralela haswell/broadwell, com o uso das ferramentas de análise de código.

## 3. Introdução

Tendo como referência o plano de trabalho enviado ao CNPq, a proposta desse projeto de iniciação científica pode ser enunciada como

*"o estudo das oportunidades de modernização (vetorização e paralelização) de códigos para solução de equações diferenciais parciais parabólicas com o uso das ferramentas que compõem o Intel Parallel Studio: (1) Vtune Amplifier; (2) Intel Advisor; (3) Intel Trace Analyser and Collector."*

Entretanto, alguns conhecimentos prévios são necessários antes que temas como vetorização, paralelização e método HOPMOC sejam trabalhados, e ferramentas como o Vtune Amplifier, Intel Advisor e Intel Trace Analyser and Collector, utilizadas.

As atividades até então desenvolvidas buscam o esclarecimento acerca de temas fundamentais para a futura compreensão satisfatória desses tópicos que norteiam a

proposta de trabalho(vetorização, paralelização, HOPMOC e ferramentas Intel). Foram, portanto, estudados e trabalhados temas base, intimamente ligados à arquitetura e organização de computadores, certamente indispensáveis para o desenvolvimento futuro.

Serão abordados na seção 5, referente a Resultados e Discussão, os estudos e as principais informações extraídas sobre temas como pipeline, superescalar, cálculos referentes ao custo computacional de instruções, bem como um olhar mais atento para dois produtos Intel específicos: o Pentium 4 e o KNC(Knights Corner).

O trabalho realizado em cima desses conceitos base, seguido pelo estudo de caso dos dois processadores, fornece um arcabouço teórico suficiente para que se tenha início um estudo mais específico, logo em seguida, envolvendo os conceitos de vetorização e paralelização, e também exercícios práticos com o HOPMOC.

#### **4. Material e Métodos ou Metodologia**

Para essa parte inicial e ainda prematura do desenvolvimento do projeto de pesquisa, o material e a metodologia ficaram restritos ao estudo de livros e artigos. Foram selecionados capítulos específicos dentro de livros didáticos comumente utilizados em disciplinas relacionadas à arquitetura e a organização de computadores, bem como artigos que se relacionam parcial ou totalmente com o conteúdo que foi estudado nessa etapa ainda primitiva do projeto.

Foram feitos, também, resumos informais acerca do que foi lido, com o objetivo de facilitar a compreensão do texto original, e também, através dessa síntese escrita, tornar mais eficiente releituras futuras.

Os livros e artigos utilizados para esses estudos iniciais foram escolhidos com base na sua correlação direta ou suficientemente parcial com o tema, e, também, no caso dos livros, por sua linguagem didática que fornece um entendimento mais genérico do tema, tendo-se a especialização necessária através do artigo.

Todos os livros e artigos usados até então estão listados na bibliografia desse relatório.

#### **5. Resultados e Discussão**

O passo inicial dos estudos foi entender os conceitos e posteriormente o funcionamento de uma pipeline e em seguida de uma arquitetura superescalar.

Referente à pipeline, alguns conceitos destacam-se por possuir correlação direta com a proposta do trabalho, como por exemplo as noções de speedup com base na conclusão de instruções por ciclo de relógio(sempre dependendo da quantidade de estágios da pipeline) e a questão dos pipeline hazards, que impedem uma execução contínua da

pipeline, reduzindo o speedup em relação a um processador sem qualquer tipo de estratégia de paralelismo. Dentro os principais pipeline hazards estudados, os que receberam mais atenção foram os data hazards(quando há conflito no acesso de um operando), e os control hazards(causados por previsões erradas acerca de branches).

Como exemplo prático, foi utilizada a pipeline do Intel 80486, que possui um tamanho relativamente conveniente para fins didáticos. com apenas 5 estágios : fetch, decode 1, decode 2, execute e write back.

Quanto ao tema de arquiteturas superescalares, foi dada atenção especial à questão do paralelismo a nível de instrução e as condições que impedem esse paralelismo, como dependência de dados(Read after Write, Write after Read e Write after Write), dependência de condicionais/saltos(não há como prever se o programa irá executar o salto ou não) e o conflito de recursos. Políticas de emissão e conclusão de instruções como in-order issue/in-order completion, in-order issue/out-of-order completion, e out-of-order issue/out-of-order completion, também foram objeto de atenção especial.

Após a leitura desses temas iniciais, foi feito um breve estudo sobre a avaliação de desempenho de uma CPU genérica, com o intuito de aprender a fazer uma estimativa matemática do custo computacional de um conjunto de instruções, com o uso de grandezas simples, como o número de instruções, o tempo de ciclo e a quantidade de ciclos gasto por um tipo de instrução específica. Ainda nesse estudo, foi possível aprender a calcular taxas como MIPS(Million Instructions Per Second), e MFLOPS(Million Floating-point Operations Per Second). Esse estudo foi concluído com a dedução dos cálculos de desempenho que envolvessem um sistema com uma hierarquia de memória, e como o tempo gasto para processar um fluxo de instruções pode variar de acordo com a presença de múltiplas caches e diferentes frequências de cache hit e cache miss.

Antes de especificar um estudo sobre o Pentium 4 e o KNC, foi feita uma leitura na organização e funcionamento mais geral da arquitetura x86 da Intel. Esse estudo foi dirigido aos pontos da arquitetura x86 que mais se relacionavam com os conceitos de superescalar e pipeline. Analisar a classificação de registradores do tipo x86, bem como suas respectivas funções foi o passo inicial para começar a entender o básico dessa arquitetura. De forma objetiva, é possível separar os registradores x86 nas seguintes classes: general(de propósito geral), segment(que fazem referência a algum segmento de memória), flags(guardam condition codes e mode bits) e instruction pointer(armazena o endereço da instrução atual), além dos registradores específicos da FPU, mais precisamente, numeric(armazenam um número em ponto flutuante), control(controlam operações em ponto flutuante), status(refletem o status atual da FPU), e tag word(contém bits que informam a natureza do valor armazenado em dado registrador do tipo numeric). Em seguida foi feito um aprofundamento na constituição e funcionamento de duas classes de registradores um pouco mais complexos: os do tipo flag e do tipo control. A última parte desse olhar inicial para a arquitetura x86, foi focada em entender o processo básico para se lidar com uma interrupção ou exceção nesse tipo de processador, bem como a natureza desses processos e suas classificações.

O primeiro exemplo prático tomado foi o Pentium 4. Novamente o foco do estudo estava relacionado com o ciclo de instruções, mais precisamente, a parcela do ciclo de instruções que se dá quase inteiramente no processador. As leituras tiveram como



objetivo entender com mais detalhes o esquema de funcionamento básico de um Pentium 4:

- 1) O processador faz fetch de instruções na ordem do programa
- 2) Cada instrução é então traduzida em uma ou mais instruções do tipo RISC, conhecidas como micro-instruções.
- 3) O processador executa essas micro-instruções numa arquitetura superescalar.
- 4) O processador então envia os resultados das micro-instruções para os registradores envolvidos na instrução que compunha o fluxo original do programa.

Um ponto muito interessante e que vale a menção, é que a pipeline do Pentium 4 possui um tamanho consideravelmente grande: 20 estágios. Esse número pode ser ainda maior, devido ao fato de algumas micro-instruções precisarem de múltiplos estágios de execução. A magnitude desse valor contribui para maiores penalizações em caso de previsões erradas acerca de branches.

Foram estudadas as partes mais relevantes desse processo geral descrito acima em 4 etapas, são eles:

**Geração de micro-instruções:** estágio em que uma unidade de fetch/decode busca instruções presente no fluxo original do programa e as traduz em uma, duas, três ou quatro micro-instruções (nesse caso, instruções do tipo RISC, de 118bits), armazenadas numa cache L1 denominada trace cache.

**Ponteiro da próxima instrução da trace cache:** lida com a seleção da próxima instrução da trace cache, e possui um sistema de branch prediction particular, que envolve a utilização de consultas de histórico num buffer especial, denominado Branch Target Buffer, afim de fazer previsões baseadas em experiências anteriores com aquela mesma entrada.

**Fetch na trace cache:** a trace cache recebe as micro-instruções já decodificadas e as armazena de forma sequencial. Nesse estágio, essas micro-instruções são recuperadas de forma sequencial.

**Drive :** Esse é o quinto estágio da pipeline do Pentium 4, que entrega as instruções já decodificadas presentes na trace cache, para o módulo de renomeação/alocação.

**Alocação:** Esse estágio aloca recursos necessários para a execução. Tarefas de destaque incluem a criação de entradas no reorder buffer (ROB) e na filas de micro-instruções para o scheduler.

**Renomeação de registradores:** estágio que remapea os 8 registradores de uso geral e os 8 registradores de ponto flutuante em um outro conjunto de 128 registradores físicos.

**Enfileiramento de micro-instruções:** as micro-operações são então enviadas a uma das duas filas de espera possíveis, para aguardar espaço livre nos schedulers. Uma das filas é destinada a operações que envolvem acesso à memória, e a outra a operações que não acessam a memória.

Escalonamento(scheduling) e Despacho(dispatching): aqui, cada um dos schedulers busca na fila por micro-instruções que estejam preparadas para execução imediata(tenham seus operandos prontos e com recursos requeridos disponíveis), e então as despacha para a unidade de execução.

Já o Knights Corner, é uma implementação específica da arquitetura MIC(Many Integrated Core), que começou com o Larrabee em 2007, inicialmente projetada para ser um produto com características de GPU/HPC, mas logo em seguida reduzida somente a HPC, e apelidado de Knights Ferry (KNF). Nesse contexto, o KNC surge como um sucessor lógico do KNF.

O KNC é um objeto de estudo interessante, uma vez que seus cores são baseados na especificação P54C do Pentium original, obviamente com adesão de funcionalidades e estruturas, como o suporte 64 bits, melhorias no sistema de cache, um bus bidirecional especial, e outras, afim de garantir um desempenho condizente com o contexto temporal em que foi lançado.

## 6. Conclusão

Os estudos realizados até então se relacionam de forma íntima com os objetivos da proposta de trabalho, e formam a base de conhecimentos necessários para um desenvolvimento futuro livre de lacunas teóricas e más interpretações conceituais. Ignorar esses estudos básicos e iniciais, visando uma iniciação mais rápida em temas mais avançados, poderia tanto atrapalhar a compreensão futura, bem como empobrecer análises de resultados e conclusões posteriores.

Todavia, ainda que introdutória, a pesquisa realizada nesses poucos meses já viabiliza uma continuação no campo teórico, com a análise e comparação de outras arquiteturas Intel, até modelos mais novos, e o começo dos trabalhos práticos, utilizando o método HOPMOC seguido de comparações de custo computacional e speedup.

## 7. Bibliografia

STALLINGS, William. **Arquitetura e organização de computadores**. 8ª ed. São Paulo: Pearson Prentice Hall, 2010.

RAUBER, Thomas; RÜNGER, Gudula. **Parallel programming: For multicore and cluster systems**. Springer Science & Business Media, 2010.

TANENBAUM, Andrew S.; ZUCCHI, Wagner Luiz. **Organização estruturada de computadores**. 5ª ed. São Paulo: Pearson Prentice Hall, 2007.

TEIXEIRA, Thiago Daniel Quimas Simões. **Estratégias de otimização para um Método Numérico para resolução de problemas de Convecção-Difusão**. 2016. 72 f. Trabalho de Conclusão de Curso(Graduação em Sistema de Informação) - Universidade Estácio de Sá, Petrópolis, 2016.

CHRYSOS, George. Intel® Xeon Phi™ coprocessor-the architecture. **Intel Whitepaper**, v. 176, 2014.

# Relatório PIBITI

**Candidato:** Maria Eduarda Araújo

**Orientadora:** Sandra Mara Cardoso Malta

**Título:** Introdução aos Métodos Numéricos para Solução de Equações Diferenciais Parciais

**Período:** 05/06/17-10/07/17

## Objetivos

Este projeto de Iniciação Científica, com duração de 12 meses, tem como objetivo principal introduzir conhecimentos básicos de matemática (noções de cálculo, equações diferenciais ordinárias e parciais, álgebra linear e métodos numéricos) e de computação (linguagens computacionais tais como C++, Fortran e etc.) à candidata (cursando o 2º período de Engenharia Civil) visando completar a sua formação e, desse modo, tornando-a apta na análise e solução de modelos matemáticos e numéricos relacionados, por exemplo, ao escoamento em meios não saturados (contaminação de solos). Ao final do projeto a candidata apresentará um texto, resumindo os principais resultados da sua pesquisa (usando o LaTeX), ilustrados por gráficos (uso do Matlab ou Octave e etc.) e obtidos a partir de simulações computacionais (programação). Neste primeiro mês de trabalho a aluna está dedicada ao estudo de métodos de resolução de sistemas lineares (decomposição LU, métodos diretos e etc) e assistindo, como ouvinte, a disciplina de “Introdução a Computação”, da Pós-Graduação do LNCC.

## Introdução

O movimento da água em solos não saturados (zona vadosa) pode ser descrito pela equação de Richards, obtida pela combinação da equação de conservação de massa e da lei de Darcy-Buckingham. Usualmente essa equação é representada sob três formas básicas:  $h$  - baseada na carga de pressão,  $\theta$  - baseada na umidade volumétrica e mista - baseada na combinação das duas variáveis. Essas equações são matematicamente equivalentes; porém, em geral, suas aproximações numéricas podem levar a resultados bastante diferentes.

Portanto, neste projeto, visando o estudo de métodos numéricos para a solução de equações diferenciais parciais, analisaremos inicialmente modelos numéricos para a aproximação da equação de Richards, possivelmente nas três formas acima mencionadas, em uma dimensão, empregado-se o método de diferenças finitas e aplicaremos técnicas para aproximar as não-linearidades nos pontos intermediários da malha e naqueles situados nas interfaces das camadas. A partir de dados disponíveis na literatura serão avaliadas a precisão e a eficiência dessas metodologias nas aproximações da condutividade e da capacidade hidráulica. Finalmente, considerando-se o desempenho dos métodos, serão analisadas as melhores opções para a solução de cada problema, indicando-se assim alternativas na implementação de modelos numéricos para o escoamento da água na zona vadosa

Para a execução do trabalho proposto será empregada a metodologia apresentada abaixo.

## Metodologia

- a) Inicialmente a aluna estudará conceitos matemáticos básicos importantes para o entendimento de equações diferenciais ordinárias e do método de volumes finitos (e/ou diferenças finitas), via resultados de álgebra linear, cálculo diferencial e integral e cálculo numérico;
- b) Paralelamente ao item a), a aluna irá assistir, como ouvinte, o curso “Introdução à Computação”, ministrado pelo Prof. Eduardo Garcia, na PG-LNCC;
- c) Serão também introduzidos conceitos de física, visando a modelagem do problema de difusão de contaminantes na zona não-saturada, considerando-se meios homogêneos e heterogêneos (equação de Richards, entre outras), em um e duas dimensões
- d) Em seguida, com base no aprendizado teórico e computacional adquirido, a aluna deverá desenvolver e implementar códigos computacionais, simulando e reproduzindo, inicialmente, problemas modelos (*benchmarks*) e, posteriormente, modelos mais avançados na área de interesse desse projeto.

## Cronograma de Atividades (separado por bimestres)

	1 <sup>o</sup>	2 <sup>o</sup>	3 <sup>o</sup>	4 <sup>o</sup>	5 <sup>o</sup>	6 <sup>o</sup>
a)	X	X				
b)	X	X	X			
c)			X	X	X	X
d)			X	X	X	X

## Resultados e Discussões

O projeto está apenas começando. A aluna até o momento está estudando a solução de sistemas lineares e aprendendo conceitos básicos de computação. Com isto pretende-se ter uma base inicial para o desenvolvimento do projeto.

## Conclusões

Ainda não temos conclusões.

## Referências Bibliográficas

Cálculo Numérico, Aspectos Teóricos e Computacionais, Márcia A. Gomes Ruggiero, Vera Lúcia da R. Lopes, McGraw-Hill, 1988.

Elementary Differential Equations and Boundary Value Problems, William E. Boyce, Richard C. Di Prima, John Wiley & Sons

Equações Diferenciais Parciais I/II – Rodney J. Biezuner, UFMG, Notas de Aula, 2010 –  
Notas de Aula - <http://www.mat.ufmg.br/~rodney/>

Métodos Numéricos para Equações Diferenciais Parciais Elípticas, UFMG, Rodney J.  
Biezuner, 2017 - Notas de Aula - <http://www.mat.ufmg.br/~rodney/>

Técnicas computacionais para dinâmica dos fluidos: conceitos básicos e aplicações,  
Fortuna, A. O., 2000, Edusp.

**Relatório de Atividades (01/07/17 a 31/07/17)**  
**Pibit/LNCC**

**Título do Projeto:**

*Processamento e Visualização de Imagens de Microscopia  
Confocal*

**Bolsista**

**MATHEUS ALVES DE MOURA**

**Orientador**

Gilson Antonio Giraldi  
([gilson@lncc.br](mailto:gilson@lncc.br))

**Co-Orientador**

Francisco José Pereira Lopes  
([flopes@ufrj.br](mailto:flopes@ufrj.br))

Laboratório Nacional de Computação Científica – LNCC/MCTI

Julho de 2017

## 1. Objetivos

O objetivo deste projeto visa a aplicação de métodos de interpolação e reconstrução 3D de imagens de microscopia confocal. Numa primeira abordagem, o estudo de técnicas de processamento de imagens faz-se necessário para o entendimento de protocolos que serão realizados no futuro, com o intuito de viabilizar o objetivo final.

## 2. Introdução

Recentemente, a introdução da microscopia digital de alta resolução, juntamente com o emprego de compostos químicos denominados fluoróforos, levou a avanços significativos na pesquisa de organismos vivos [1]. Essa combinação de princípios da óptica e da físico-química, levou ao desenvolvimento do microscópio confocal por varredura laser, o qual vem possibilitando a análise visual mais detalhada de células bem como de fenômenos em tempo real [2].

Um estudo feito com um microscópio confocal é, em geral, constituído por uma sequência de imagens coloridas, bidimensionais, da região de interesse do organismo vivo [3,11]. O volume de dados gerado pode ser tratado via técnicas de processamento de imagens e computação gráfica, com o objetivo de extrair informações ou gerar modelos geométricos das estruturas de interesse [4].

Do ponto de vista de processamento digital de imagens, trata-se de uma área com dificuldades específicas, inerentes ao tipo de imagens, formatos digitais, protocolos de aquisição e diferentes escalas dos objetos de interesse. Softwares disponíveis nesta área, como o ImageJ (<https://imagej.net/Fiji>), ainda são incipientes para as necessidades dos pesquisadores, tanto do ponto de vista do processamento como do ponto de vista da visualização 3D.

Por um lado, observa-se a necessidade de novas técnicas para processamento e visualização de imagens de microscopia confocal. Assim, é necessário o desenvolvimento de uma plataforma de software que permite disponibilizar as técnicas pesquisadas para a comunidade científica.

A equipe de processamento de imagens do LNCC apresentou em [5] o protótipo do software *PyImageVis*, desenvolvido na linguagem Python [6], para processamento e visualização de imagens ([virtual01.lncc.br/pyimagevis/](http://virtual01.lncc.br/pyimagevis/)). Este protótipo foi desenvolvido principalmente para uso acadêmico, no intuito de proporcionar a estudiosos da área de visualização e processamento de imagens, um ambiente bem completo e eficiente. Atualmente, o grupo do LNCC vem atualizando este software para melhorar suas funcionalidades. Embora o foco inicial do *PyImageVis* seja imagens médicas e odontológicas, os fundamentos e plataformas utilizadas em seu desenvolvimento darão um núcleo para gerar o sistema necessário para o presente projeto, com suporte da equipe do LNCC.

Desta forma, foca-se as atividades deste projeto no estudo e pesquisa de técnicas para processamento e visualização das imagens. Serão utilizadas bibliotecas gratuitas e de código livre [7-9], seguindo a filosofia do sistema *PyImageVis*.

## 3. Metodologia

Inicialmente, será realizado um estudo dirigido sobre técnicas de processamento e análise de imagens, baseado no material disponível em [10], preparado pelo orientador deste projeto. Este estudo será mesclado com implementações e testes, utilizando imagens de microscopia confocal, para o candidato se familiarizar com a linguagem Python e seus recursos, bem como com os problemas do tipo de imagem de interesse. Em particular, será utilizado o PyMorph [8], que é uma coleção completa de funções de morfologia matemática para imagens. Suas funções podem ser aplicadas para segmentação de imagens, aplicação de filtros, etc. Este pacote é de código livre e foi desenvolvido totalmente em Python.



Em seguida, serão aplicados e desenvolvidos métodos de realce e filtragem para as imagens de interesse. Estes métodos permitirão uma definição melhor dos diferentes padrões de intensidade das estruturas presentes na imagem.

Inicialmente, serão testados métodos de limiarização para segmentação dos objetos da imagem. Serão testados também métodos baseados em morfologia matemática, entropia e modelos deformáveis 2D, aproveitando a experiência do grupo do LNCC nestes temas.

#### 4. Resultados e Discussão

Neste momento, o aprendizado está sendo encaminhado com o objetivo aplicar técnicas baseadas em extração contorno de objetos e limiarização de imagens de microscopia confocal, assim como em [12]. Para isso, estão sendo realizados estudos de bibliotecas do Python [6,7,8,9], bem como de funções disponíveis no MatLab para aquisição dos fundamentos necessários para os testes futuros.

#### 5. Bibliografia

[1] Wanderley De Souza W. (ed), 2010. Microscopia Óptica: fundamentos e Aplicações às Ciências Biomédicas, SBME, Rio de Janeiro.

[2] David J. Stephens, Victoria J. Allan. Light Microscopy Techniques for Live Cell Imaging, Science, 2003:300, 5616, pp. 82-86. DOI: 10.1126/science.1082160

[3] B. Pawley (ed), 2006, Handbook of Biological Confocal Microscopy, LLC, Third edition, New York.

[4] L. Rosenblum, R. Earnshaw, J. L. Encarnação, H. Hagen, A. Kaufman, G. Klimenko, S. and Nielson, F. Post, and D. Thalmann. Scientific visualization : advances and challenges. New York Academic Press 1994, 1994.

[5] Diego Augusto Thomaz Quadrado Leite and GIRALDI, G. A. and P. H. M. Lira and KNEIPP, R. E. "PyImageVis: Processamento e Visualização de Imagens Médicas em Python". Revista Eletrônica de Iniciação Científica. Volume 11, No. 1, paginas 16-35, 2011.

[6] Python programming language. In <http://www.python.org>

[7] Numpy. In <http://numpy.scipy.org/>

[8] PyMorph Morphology Toolbox. In <http://www.mmorph.com/pymorph/>

[9] Python Imaging Library (PIL). In <http://www.pythonware.com/products/pil/>

[10] Gilson A. Giraldi. Curos GA-023. Notas de Aula: <http://virtual01.Incc.br/~giraldi/GA023-2017/>

[11] Handbook of Biological Confocal Microscopy, Third edition, edited by B. Pawley, Science+Business Media, LLC, New York, 2006.

[12] Jin Qi, B. Wang, N. Pelaez, I. Rebay, R. W. Carthew, A. K. Katsaggelos, and L. A. Nunes Amaral. "Drosophila Eye Nuclei Segmentation Based on Graph Cut and Convex Shape Prior". Int Conf Signal Process Proc. 2013 Sep 18 : 670-674. DOI 10.1109/ICIP.2013.6738138.

## RELATÓRIO DE ATIVIDADES

**Programa de bolsas de iniciação científica e tecnológica dos programas PIBIC-PIBITI/LNCC**

**Área:** Biologia Computacional. **Sub-área:** Modelagem Molecular de Sistemas Biológicos.

### ***ESTUDO DE ESTRATÉGIAS AUXILIARES PARA PREDIÇÃO DE ESTRUTURAS DE PROTEÍNAS BASEADA EM MOLDES: ANÁLISE DE PREDITORES DE ATRACAMENTO MOLECULAR PROTEÍNA-PROTEÍNA***

---

#### **1. Dados Gerais**

**Título do Projeto:** “Estudo De Estratégias Auxiliares Para Predição De Estruturas De Proteínas Baseada Em Moldes: Análise De Preditores De Atracamento Molecular Proteína-Proteína”.

**Nome do Bolsista:** Matheus Miloski.

**Nível da Bolsa:** Iniciação Tecnológica.

**Orientador / Pesquisador Responsável:** Fábio Lima Custódio – LNCC/MCTIC.

**Co-orientador:** Gregório Kappaun Rocha – LNCC/MCTIC.

**Período do Relatório:** 01/03/2017 – 30/07/2017.

O projeto é coordenado pelos Professores Fábio L. Custódio, D.Sc. e Laurent E. Dardenne, D.Sc., pesquisadores do Laboratório Nacional de Computação Científica e membros do Grupo de Modelagem Molecular de Sistemas Biológicos (GMMSB/LNCC). Conta ainda com a participação de outro membro do GMMSB/LNCC, Gregório Kappaun Rocha, D.Sc., bolsista do Programa de Capacitação Institucional (PCI). O projeto é desenvolvido na área de Modelagem Molecular. O tema investigado no projeto – o Atracamento Molecular Proteína-Proteína – é parte do problema de predição de estrutura de proteínas, uma das áreas de pesquisa do GMMSB. O problema é altamente interdisciplinar, envolvendo conhecimentos avançados em modelagem molecular, aplicação de técnicas metaheurísticas de otimização, e necessitando ainda de alto poder computacional.

O bolsista vem realizando seus estudos no GMMSB/LNCC desde Novembro de 2015, tendo apresentado seminários e além disso, resultados parciais já foram alcançados em duas abordagens auxiliares fundamentais para a predição de proteínas mais complexas, que são: (i) a identificação de

domínios e; (ii) identificação e avaliação de preditores de moldes. O bolsista contribui, ainda, para a participação do GMMSB/LNCC na 12<sup>a</sup> ed. do CASP (*Critical Assessment of Techniques for Protein Structure Prediction*), realizada em Dez. de 2016 na Itália. O CASP é um dos principais encontros da área e avalia o estado da arte de grupos de todo o mundo, ocorrendo de maneira bianual.

## 2. Objetivos

**Objetivo geral:** investigar estratégias auxiliares para a predição de estruturas de proteínas, permitindo ao programa GAPF atacar proteínas maiores e mais complexas.

**Objetivo específico:** (i) Analisar o desempenho de metodologias para predição de atracamento molecular proteína-proteína. Isto é, avaliar programas que buscam prever a estrutura tridimensional de complexos moleculares formados à partir da interação entre duas ou mais proteínas. (ii) Implementar métricas de avaliação da qualidade de complexos proteicos preditos.

## 3. Introdução

Interações entre proteínas e outras moléculas, e mais especificamente, entre proteínas e outras proteínas, são essenciais para o funcionamento dos organismos, e conseqüentemente, para a sua sobrevivência (Leninski et al., 2016). Para expandir o conhecimento à respeito de processos biológicos (especialmente processos celulares), é necessário portanto, elucidar aspectos das propriedades biofísicas, bioquímicas e estruturais destes tipos de reações, bem como os complexos macromoleculares que são formados por elas ao longo dos processos (Zhang et al., 2016).

Devido ao avanço das técnicas de proteômica, a quantidade de dados de proteínas disponíveis se mostra enorme e crescente. Exemplos dessas informações descobertas e disponibilizadas em bancos de dados públicos, são: redes de interações entre proteínas, mostrando quais proteínas interagem entre si; estruturas primárias de proteínas (sequencia linear de aminoácidos que as constituem); e estruturas terciárias de proteínas (proteínas enoveladas em sua conformação tridimensional nativa). Embora estes dados gerem importantes *insights* ao revelar quais proteínas interagem com quais, ainda revelam poucas informações à respeito de como elas interagem, ou seja, características físico-químicas e estruturais das interações (Méndez et al, 2003).

Através de técnicas experimentais de biologia estrutural, como cristalografia de raios x e ressonância magnética nuclear, é possível elucidar a conformação tridimensional de proteínas individuais e dos complexos que elas formam ao interagir entre si, porém, estas técnicas vêm se mostrando lentas e dependentes de um alto custo monetário (Méndez et al, 2003; Zhang et al, 2016).

Devido às limitações das metodologias experimentais, esforços vêm sendo tomados para se desenvolver ferramentas computacionais capazes de prever como as proteínas interagem entre si, bem como prever a conformação dos complexos macromoleculares tridimensionais que são gerados

à partir da interação entre duas ou mais proteínas (Leninski et al, 2010; Lensinki et al, 2016). Investigar e avaliar o desempenho dessas metodologias computacionais é o foco deste trabalho.

À cada seis meses, desde o ano 2001, ocorre um evento chamado *Critical Assessment of Prediction of Interactions* (CAPRI - <http://www.ebi.ac.uk/msd-srv/capri/>), cuja atividade vem estimulando o desenvolvimento e progresso dos métodos de modelagem estrutural de complexos proteicos, também conhecidos como *protein-protein docking*, ou, atracamento molecular proteína-proteína. No CAPRI são disponibilizados conjuntos de estruturas tridimensionais de complexos proteicos (que já foram elucidadas experimentalmente, mas ainda não divulgados) para avaliar o desempenho das metodologias de predição. Inicialmente, o CAPRI focava apenas em complexos formados por duas proteínas, porém se expandiu ao longo dos anos, e hoje também inclui complexos de proteínas com outras moléculas, predição da afinidade de ligação e predição da posição das moléculas de água na interface de interação entre duas proteínas. Tal fato mostra a evolução e a grande importância do desenvolvimento de ferramentas nesta área.

## 4. Metodologia

### 4.1. Seleção dos Preditores de Complexos Proteicos e do Conjunto Teste

Inicialmente, foi feita uma busca na literatura, afim identificar as melhores ferramentas atuais para a predição da estrutura de complexos proteicos, bem como para se ter uma ideia geral de suas metodologias.

Os seis preditores que obtiveram os melhores resultados na sexta edição do CAPRI (que engloba os *rounds* 28 ao 25, e ocorreu do ano 2013 ao 2016) foram identificados. Estes preditores são: *ClusPro* (Kozakov et al, 2017), *PyDockWeb* (Jiménez-García et al, 2013), *LzerD* (Venkatraman et al, 2009), *Haddock* (van Zundert et al, 2015), *SwarmDock* (Torchala et al, 2014) e *GalaxyPPDock* (Shin et al, 2014). Além destes seis preditores, também selecionou-se o preditor *ZDock* (Pierce et al, 2014), relativamente novo na área e que vêm obtendo bons resultados.

Quatro metodologias (*ClusPro*, *SwarmDock*, *ZDock*, *Haddock*) disponibilizam um serviço de predição na *web* aberto.

Como conjunto teste, foi selecionado um *benchmark* (Vreven et al., 2015) com 230 complexos compostos por duas proteínas, e cujas estruturas foram elucidadas através de técnicas experimentais. Este conjunto possui uma grande diversidade de complexos, que podem ser do tipo enzima-substrato, anticorpo-antígeno e outros, e que são classificados em fáceis de serem preditos, de dificuldade mediana e difíceis.

Preparação do conjunto teste: antes de submeter os complexos para a predição, as coordenadas tridimensionais das proteínas ligantes foram alteradas, uma vez que já se encontravam

com certo grau de atracamento às proteínas receptoras no *benchmark*. Para isto, usou-se o programa Gromacs, onde as coordenadas foram rotacionadas em 90° e às translacionadas também em 90°.

Para fase de automatização de todo o método (execução do preditor e análise da qualidade dos modelos preditos), utilizou-se nos estudos iniciais um subconjunto de 5 complexos, composto por dois alvos fáceis: 2A1A (enzima-substrato) e 4GXU (antígeno-anticorpo); dois alvos médios: 4LW4 (enzima-substrato) e 3V6Z (antígeno-anticorpo); e um alvo difícil: 3H11 (outros). Dois preditores - que apresentam servidores *web* - foram avaliados nesta etapa inicial: *ClusPro* e *ZDock*.

Após esta fase de automatização, a metodologia será aplicada em larga escala em todo o *benchmark* (composto de 230 complexos) e para um número maior de preditores.

## 4.2. Avaliando a Qualidade dos Modelos Preditos

Três critérios para ranquear a similaridade dos modelos preditos com a estrutura tridimensional obtida experimentalmente são aplicados no CAPRI: i) *f-nat*, que é a fração de contatos nativos entre resíduos de aminoácidos encontrados no modelo, sendo um contato entre dois resíduos determinado pela proximidade física entre eles na estrutura tridimensional; ii) *I-rmsd*, que é o RMSD dos resíduos de interface, sendo o RMSD uma medida de similaridade estrutural que é obtida através do alinhamento entre duas estruturas, e os resíduos de interface, sendo aminoácidos que estão na região de interação entre as proteínas; iii) *L-rmsd*, que é o RMSD da proteína ligante após a sobreposição da proteína receptora, sendo a proteína ligante a que possuir a menor quantidade de aminoácidos à constituindo, e a proteína receptora, a que possuir a maior quantidade.

Não encontra-se disponível nenhuma ferramenta para calcular tais critérios (*f-nat*, *i-rmsd*, *l-rmsd*). Assim, os programas para fazer tais análises também foram desenvolvidos neste trabalho.

Para calcular o *f-nat*, o programa recebe quatro estruturas de proteínas, sendo duas estruturas nativas experimentais (receptora e ligante) e dois modelos preditos (receptora e ligante). É preciso determinar até qual distância entre os resíduos um contato será considerado, bem como até qual distância a região de interface será tratada. O programa então calcula os contatos reais, encontrados na proteína nativa, depois calcula os contatos encontrados no modelo e divide o número de acertos pelo número total de contatos reais. O valor de 10 ângstrons é usado para definir a região de interface e o de 5 ângstrons para definir um contato. O programa retorna ainda: a lista de resíduos da interface que foi obtida através da análise das estruturas experimentais; a lista de resíduos em contato obtida através da análise das estruturas experimentais; a lista de contatos do modelo e; a fração de contatos nativos encontrada no modelo (*f-nat*). Quanto mais próximo de 1.0 for o valor do *f-nat*, melhor o modelo do complexo predito.

Para calcular o *i-rmsd*, um *script* utiliza a lista de resíduos de interface gerada anteriormente e a submete para o cálculo de RMSD dos resíduos especificados utilizando o programa Gromacs.

## 5. Resultados e Discussão

Utilizou-se, nos estudos iniciais, um subconjunto de 5 complexos, composto por dois alvos fáceis: 2A1A (enzima-substrato) e 4GXU (antígeno-anticorpo); dois alvos médios: 4LW4 (enzima-substrato) e 3V6Z (antígeno-anticorpo); e um alvo difícil: 3H11 (outros). Dois preditores - que apresentam servidores *web* - foram avaliados nesta etapa inicial: *ClusPro* e *ZDock*.

Os resultados da avaliação foram divididos em duas partes: utilizando o melhor modelo segundo o preditor; e escolhendo o melhor modelo, dentre os dez melhores segundo o preditor.

Utilizando o melhor modelo fornecido pelo preditor, observa-se um desempenho satisfatório nos alvos fáceis em ambos os preditores. Porém, nenhum preditor conseguiu prever a interface de contato nos alvos médios (Tabela 1). No alvo difícil o *ZDock* se sai melhor (Tabela 1 e Figura 1).

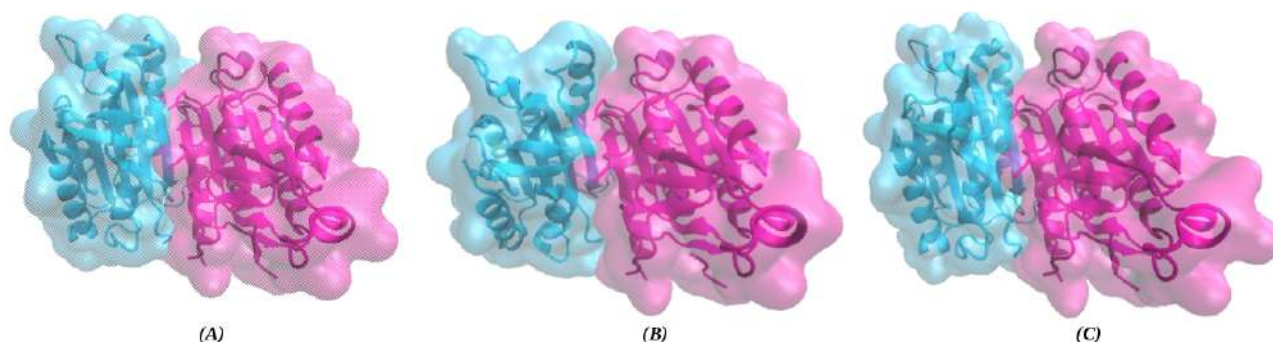
**Tabela 1:** Qualidade estrutural (via *f-nat*) dos modelos preditos pelos preditores *ClusPro* e *ZDock*.

Complexo Proteína-Proteína	Top 1		Top 10	
	<i>ClusPro</i>	<i>ZDock</i>	<i>ClusPro</i>	<i>ZDock</i>
2A1A – 8Cs - ES	0.50	0.50	0.50	0.88
4GXU – 41Cs - AA	0.88	0.73	0.88	0.73
3V6Z – 7Cs - AA	0.00	0.00	0.00	0.86
4LW4 – 6Cs - ES	0.00	0.00	1.00	0.83
3H11 – 12Cs - ECR	0.25	0.67	0.33	0.67

*Top 1*: *f-nat* para o modelo considerado melhor pelo preditor. *Top 10*: *f-nat* para o melhor resultado dentre os 10 modelos fornecidos pelo preditor. ES: enzima-substrato. AA: antígeno-anticorpo. ECR: enzima-cadeia regulatória.

Selecionando o melhor modelo dentre os dez melhores fornecidos pelos preditores, o *Zdock* apresenta resultados maiores que 0.67 em todos os alvos, enquanto que o *ClusPro* apresenta apenas dois resultados com *f-nat* maior que 0.67 (Tabela 1).

Assim, o *ZDock* coloca-se como uma boa opção para predição de complexos proteicos, embora ainda apresente problemas na seleção dos melhores modelos.



**Figura 1.** (A) Complexo nativo para alvo 3H11. (B) Modelo predito pelo *ClusPro* (*f-nat* 0.25). (C) Modelo predito pelo *ZDock* (*f-nat* 0.67).

## 6. Conclusões

Prever a organização tridimensional de complexos proteicos não é tarefa trivial, porém é fundamental para estudos que envolvam análises das interações entre proteínas. O preditor *ZDock* apresenta-se como um bom candidato para a predição de complexos gerados a partir da interação entre duas proteínas. Porém, avanços ainda precisam ser alcançados por essa metodologia uma vez que nem todo alvo consegue ser predito com alto sucesso, além de apresentar problemas na seleção de seu melhor modelo. Outros preditores como *PyDockWeb*, *LzerD*, *Haddock*, *SwarmDock* e *GalaxyPPDock* colocam-se, aparentemente, como boas alternativas para o problema e já estão sendo avaliados.

Após a análise completa dos preditores de complexos proteicos, o melhor deles será adicionado ao *workflow* do programa de predição de estrutura de proteínas GAPF (Custodio et al., 2014; Rocha et al., 2015), afim de melhorar sua capacidade preditiva.

## 7. Referências Bibliográficas

- Lensink, Marc F., Sameer Velankar, and Shoshana J. Wodak. "Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition." *Proteins: Structure, Function, and Bioinformatics* 85.3 (2016): 359-377
- Lensink, M.F. & Wodak, S.J. Docking and scoring protein interactions: CAPRI 2009. *Proteins* 78, 3073–3084 (2010)
- Kozakov D, Hall DR, Xia B, Porter KA, Pothorny D, Yueh C, Beglov D, Vajda S. The ClusPro web server for protein-protein docking. *Nature Protocols*. 2017 Feb;12(2):255-278
- Jimenez-Garcia B., Pons C. and Fernandez-Recio J. "pyDockWEB: a web server for rigid-body protein-protein docking using electrostatics and desolvation scoring". *Bioinformatics* (2013) 29(13):1698-1699.
- Vishwesh Venkatraman, Yifeng D Yang, Lee Sael, and Daisuke Kihara. Protein-protein docking using region-based 3D Zernike descriptors. *BMC Bioinformatics* 10:407(2009)
- G.C.P van Zundert, J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastiritis, E. Karaca, A.S.J. Melquiond, M. van Dijk, S.J. de Vries and A.M.J.J. Bonvin (2016). "The HADDOCK2.2 webserver: User-friendly integrative modeling of biomolecular complexes." *J. Mol. Biol.*, 428, 720 725 (2015)
- M. Torchala, P. A. Bates, 'Predicting the Structure of Protein-Protein Complexes Using the SwarmDock Web Server', in: *Protein Structure Prediction 3rd Edition (Methods in Molecular Biology, Vol. 1137)*, Humana Press, 2014
- W. -H. Shin, G. R. Lee, L. Heo, H. Lee, and C. Seok\*, Prediction of Protein Structure and Interaction by GALAXY protein modeling programs, *Bio Design*, 2 (1), 1-11 (2014)
- Pierce BG, Hourai Y, Weng Z. (2011) Accelerating Protein Docking in ZDOCK Using an Advanced 3D Convolution Library. *PLoS One* 6(9): e24657
- Vreven, Thom, et al. "Updates to the integrated protein-protein interaction benchmarks: docking benchmark version 5 and affinity benchmark version 2." *Journal of molecular biology* 427.19 (2015): 3031-3041
- Zhang, Qian, et al. "Recent Advances in Protein-Protein Docking." *Current drug targets* 17.14 (2016): 1586-1594
- Méndez, Raúl, et al. "Assessment of blind predictions of protein-protein interactions: current status of docking methods." *Proteins: Structure, Function, and Bioinformatics* 52.1 (2003): 51-67.
- Custodio, F.L., Barbosa, H.J.C. and Dardenne, L.E. "A multiple minima genetic algorithm for protein structure prediction". *Applied Soft Computing*, vol. 15, pp. 88–99, 2014.
- Rocha, G.K., Custodio, F.L., Barbosa, H.J.C. and Dardenne, L.E. "A multiobjective approach for protein structure prediction using a steady-state genetic algorithm with phenotypic crowding." in *Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 2015 IEEE Conference on. IEEE, 2015, pp. 1–8.