



**Universidade de São Paulo**  
Instituto de Matemática e Estatística



# **Parte Subjetiva**

## **Adequabilidade do Kepler para pacotes de Bioinformática do R**

---

**ALUNO:**

**Bruno Henrique Yoshimura**

**ORIENTADOR**

**Prof. Dr. Roberto Hirata Jr.**

**2/19/2008**

## Desafios e Frustrações

A interdisciplinaridade deste trabalho não nos limitou a aplicar apenas os conhecimentos obtidos durante o curso de ciência da computação, como também estudar a Biologia Molecular e novas ferramentas computacionais.

Por se tratar de um tema com o qual não tinha familiaridade, o principal desafio foi conseguir conhecimento suficiente para ganhar mais confiança. Mas assim como não é fácil para um biólogo aprender computação, gastamos o precioso tempo inicial no entendimento da parte biológica da análise de *microarray* e problemas de infraestrutura.

A idéia inicial do projeto era a reengenharia do pacote de Bioinformática MAIGES, e para isso, deveríamos antes de tudo, estudar a linguagem R e entender o processo biológico envolvido. Nossa primeira dificuldade foi entender a organização do pacote, assim como aprender uma nova linguagem de programação. Um pequeno desafio foi conseguir um ambiente no qual pudéssemos simular e instalar pacotes de R, uma vez que os laboratórios nos quais trabalhamos tinham não ofereciam permissão suficiente para tais alterações.

Não sentíamos em situação confortável trabalhar em duas áreas até então desconhecidas por nós: biologia e programação R. Depois de algumas reuniões, decidimos focar nosso trabalho na análise de um software de programação visual que pudesse ser integrado com o Kepler.

Este novo tema nos deu confiança por envolver a pesquisa em uma área de extrema utilidade para a comunidade científica atual. Apesar de parecer um estudo específico sobre a utilização do Kepler para o MAIGES, conseguimos ir muito além.

Estudamos as dificuldades e vantagens na utilização desse tipo de software para cientistas e abrimos portas para novos estudos na área.

No início ficamos frustrados em descobrir que não seria possível utilizar o pacote MAIGES integrado ao Kepler, mas agora sabemos que deixamos uma contribuição importante. Sabemos os motivos pelos quais não foi possível tal integração e também descrevemos algumas formas de fazê-la viável, tanto com mudanças no Kepler, como na alteração do próprio pacote.

## Disciplinas cursadas no BCC mais relevantes para o trabalho

### **MAC0110 - Introdução à Computação**

### **MAC0122 - Princípios de Desenvolvimento de Algoritmos**

### **MAC0323 - Estruturas de Dados**

Mais do que qualquer outra, considero as três disciplinas acima as bases mais importantes para o curso de computação. Este conhecimento foi essencial para a análise e entendimento dos códigos do pacote MAIGES.

### **MAE0121 - Introdução a Probabilidade e a Estatística I**

### **MAE0212 - Introdução à Probabilidade e à Estatística II**

Não apenas a teoria, mas a familiarização com a análise de dados e leitura relatórios/gráficos é muito importante na análise de *microarray*. A análise de dados brutos e análise exploratória utiliza muitos modelos estatísticos. Até mesmo a leitura das lâminas de *microarray* utilizam conhecimento básico de estatística (média, mediana).

Para medir a precisão dos dados obtidos e para analisar possíveis ruídos de leitura, também utilizamos a estatística.

### **FLC0474 - Língua Portuguesa**

Não listei esta matéria apenas por melhorar a escrita, pois acredito que o aprendizado nesta área acontece ao longo de uma vida. Acredito que o foco da matéria deveria ser apenas na escrita e não a considero suficiente para um curso de exatas.

### **MAC0338 - Análise de Algoritmos**

Dado o volume de dados na análise de *microarray*, considero importante noções sobre complexidade de algoritmos.

### **MAC0211 - Laboratório de Programação I**

Pela utilização de ferramentas computacionais que são úteis para o desenvolvimento de projetos e no dia-a-dia de um programador.

### **MAC0332 - Engenharia de Software**

O nosso projeto inicial era a reengenharia de um pacote em bioinformática, e esta disciplina é essencial para este tipo de estudo.

### **MAC0441 - Programação Orientada a Objetos**

O idêa de encapsulamento de dados foi uma das soluções que encontramos para nosso problema de integrar um pacote de bioinformática em um sistema de programação visual.

### **Outras matérias**

Duas matérias que não cursei, mas seriam importantes para este trabalho seriam “MAC0447 Análise e Reconhecimento de Formas: Teoria e Prática” pela aplicação no reconhecimento da imagem escaneada da lâmina *microarray* e “IBM1029 Introdução à Bioinformática” para aprender as bases da bioinformática.

### **Continuidade do Trabalho**

Para dar continuidade a este trabalho, seria preciso abranger nosso conhecimento tanto na área de computação e biologia molecular. Na primeira, deveríamos estudar outros ambientes de programação visual para entender as limitações na utilização dos mesmos para a bioinformática. Já na segunda, deveríamos estudar a fundo o processo biológico a fim de entender e modificar os pacotes.

Apenas assim, poderíamos analisar e eleger uma solução de integração. Acredito que seja mais fácil conseguir tal solução fazendo alterações nos pacotes de bioinformática, principalmente no encapsulamento dos dados.