

Avaliações de monografias
MAC0499 - 2012
Trabalho de Formatura Supervisionado

Pedro Ivo Gomes de Faria
nº USP: 5893844

Sumário

1	Avaliação da monografia “SAGE <i>suite</i>: análise e desenvolvimento”	2
1.1	Resumo	2
1.2	Avaliação da parte técnica	2
1.3	Avaliação da parte subjetiva	3
1.4	Críticas, elogios e comentários	3
2	Avaliação da monografia “Anotação de sequências e expansão do sistema EGene/CoEd”	3
2.1	Resumo	4
2.2	Avaliação da parte técnica	4
2.3	Avaliação da parte subjetiva	4
2.4	Críticas, elogios e comentários	5

1 Avaliação da monografia “SAGE *suite*: análise e desenvolvimento”

- **Aluno:** Samuel Gales Guimarães
- **Ano:** 2007
- **Nota obtida:** não divulgada em <http://www.ime.usp.br/~cef/mac499-07/monografias/>

1.1 Resumo

O trabalho teve como objetivo fazer programas para analisar dados de expressão gênica obtidos a partir da técnica de LongSAGE (*Serial Analysis of Gene Expression*). O conjunto de programas para fazer a análise dos dados (chamado de SAGE Analysis) já existia e foi estendido pelo aluno, enquanto o programa para simular dados obtidos pela técnica (chamado de GenSuite) foi criado por ele para fazer a validação do funcionamento do SAGE Analysis.

A monografia começa com uma contextualização do trabalho e prossegue explicando em que consiste a técnica de LongSAGE. Algumas inconsistências nos resultados e como elas foram tratadas também são explicadas; e em seguida são descritas as atividades realizadas (extensão do SAGE Analysis e desenvolvimento do GenSuite). Por último, são apresentados os resultados de alguns testes (na maioria satisfatórios, segundo o aluno) e uma conclusão (segundo a qual foi possível entender melhor a técnica e, conseqüentemente, analisar os resultados de modo mais significativo).

1.2 Avaliação da parte técnica

A monografia já supõe que o leitor possui alguma familiaridade com biologia molecular, algo que não é necessariamente verdadeiro para um bacharelado em computação (apesar de conceitos mais básicos relativos ao assunto serem expostos no ensino médio). Assim sendo, acho que valeria a pena dedicar uma seção à exposição (mesmo que simplificada) de conceitos de biologia molecular necessários ao entendimento do trabalho (principalmente sobre estrutura de ácidos nucleicos, processamento eucariótico do pré-RNAm, enzimas de restrição e PCR). Apesar do uso de figuras, ficou um tanto quanto difícil saber quais sequências das figuras correspondiam às sequências mencionadas no texto (talvez isso pudesse ser detalhado um pouco mais nas legendas). Outra dificuldade no entendimento ficou por conta do uso de termos não definidos antes do uso (como “colisão” e “pontuação”).

Apesar disso, acho que o texto está coeso e explica em detalhes quais foram os problemas envolvidos (como a variação no tamanho das *ditags*), a abordagem utilizada para o tratamento dos dados (diminuição das *tags* para 17 pb) e como foi feita a validação do programa que iria analisá-los (o SAGE Analysis). Todos os itens descritos no [roteiro para preparação de monografias](#) (introdução, conceitos, atividades, resultados, conclusão e bibliografia) estão presentes.

1.3 Avaliação da parte subjetiva

A parte subjetiva ficou boa, pois o aluno falou de suas experiências (como a interação com pessoas de fora do IME - em particular com as vindas do ICB) e dificuldades (como a falta de interesse do professor tutor) no decorrer do curso e do trabalho. Ele também listou algumas disciplinas do BCC nas quais ele aprendeu práticas que foram úteis para o desenvolvimento do software (como "Introdução à computação" e "Princípios de desenvolvimento de algoritmos"). Acho que também poderiam ter sido citadas a disciplina de "Introdução à probabilidade e à estatística I" (pois alguns conceitos de probabilidade - distribuições normal, gama e Poisson - foram utilizados na criação do GenSuite) e as disciplinas que ele cursou no ICB (apesar de não fazerem parte do currículo, provavelmente foram importantes para a realização do trabalho).

1.4 Críticas, elogios e comentários

Como comentário final, acho que deveria ter sido feita uma revisão do texto da monografia, pois alguns erros ortográficos e gramaticais acabam desviando a atenção do conteúdo (ou até prejudicam o entendimento do texto). Citando o texto "[Forma versus conteúdo](#)", do professor Paulo Feofiloff: "toda vez que arranha a forma, o autor *impede que o leitor concentre sua atenção sobre o conteúdo*".

2 Avaliação da monografia "Anotação de sequências e expansão do sistema EGene/CoEd"

- **Aluno:** Ricardo Yamamoto Abe
- **Ano:** 2006
- **Nota obtida:** não divulgada em <http://www.ime.usp.br/~cef/mac499-06/>

2.1 Resumo

O objetivo do trabalho foi estender um sistema gerador de *pipelines* (EGene/CoEd) para processamento de sequências, de modo que ele permitisse a criação de ramificações (*forks*) e seletores (encapsulam componentes que processam sequências de modo condicional). O aluno também modelou um banco de dados de evidências para auxílio no processo de anotação.

A monografia começa explicando que sequências de DNA são tipicamente processadas através de *pipelines*, que são *scripts* que executam tarefas de modo sequencial. Esse processamento permite a obtenção de evidências, que são informações que podem ser utilizadas para fazer anotações (isto é, atribuir informações às sequências, como identificação de genes e suas funções). Em seguida, é descrita qual foi a extensão feita no modelo entidade-relacionamento do sistema original para permitir a inclusão dos dados de evidências. Também são descritos os programas CoEd (interface gráfica para gerar o arquivo de configuração do EGene) e EGene (gera o *pipeline* descrito no arquivo de configuração, usando *pipes* do UNIX). Por último, o aluno descreve quais as alterações feitas para permitir ramificações e seletores (implementados usando *named pipes*) e conclui que elas permitem processar qualquer conjunto de programas que se relacionem segundo um grafo orientado acíclico (exceto para o caso em que dois ou mais programas enviam dados para um mesmo programa - em outras palavras, o grau de entrada de todos os vértices do grafo é no máximo igual a um).

2.2 Avaliação da parte técnica

A parte técnica está bem escrita, incluindo várias figuras e exemplos de utilização do CoEd e do EGene, o que facilita o entendimento do funcionamento de ambos. O aluno também explica bem qual foi a estratégia utilizada para a implementação das ramificações e dos seletores (a utilização dos *named pipes*), que facilitou bastante a mesma (já que o próprio sistema operacional fica responsável pelo controle de concorrência e paralelismo).

Quase todos os itens descritos no **roteiro para preparação de monografias** (introdução, conceitos, atividades, resultados, conclusão e bibliografia) estão presentes, exceto pela parte de resultados (talvez porque ainda faltava implementar algumas componentes, conforme descrito na conclusão).

2.3 Avaliação da parte subjetiva

A parte subjetiva ficou boa (aliás, identifiquei-me bastante com ela), pois o aluno relata as dificuldades relativas ao contato com outra área e à utilização

da linguagem Perl (que permite facilmente a criação de códigos de difícil entendimento). Ele também relata ter desenvolvido parte do trabalho em conjunto com uma aluna de mestrado do ICB, e que a interação entre os dois foi proveitosa para ambos.

Na listagem das disciplinas mais relevantes, ele optou por citar não só as que tiveram relação com o trabalho mas também as que ele julgou importantes para sua formação (como “álgebra linear” e “programação linear”).

2.4 Críticas, elogios e comentários

Acho que poderia ter sido explicado um pouco mais o que cada um dos componentes desenvolvidos pelo aluno faz (o capítulo 4 - geração de evidências - apenas descreve os nomes dos componentes desenvolvidos). Além disso, acho que a modelagem do banco de dados como um todo poderia ser melhor detalhada, e não apenas a parte adicionada pelo aluno.