Pontifícia Universidade Católica
do Rio de Janeiro

Paulo Roberto Gomes

Um
Estudo sobre Avaliação da Execução do BLAST em Ambientes Distribuídos

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para Obtenção do título de Mestre pelo Programa de Pós-Graduação em Informática da PUC-Rio.

Orientador: Sérgio Lifschitz

Rio de Janeiro

Abril de 2009

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Paulo Roberto Gomes

Um Estudo sobre Avaliação da Execução do BLAST em Ambientes Distribuídos

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática do Departamento de Informática do Centro Técnico e Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Sergio Lifschitz
Orientador
Departamento de Informática - PUC-Rio

Noemi de La Rocque Rodriguez
Departamento de Informática - PUC-Rio

Luiz Fernando Bessa Seibel
Departamento de Informática - PUC-Rio

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico – PUC-Rio

Rio de janeiro, 08 de abril de 2009

Paulo Roberto Gomes

Graduou-se em Ciências Navais e Engenharia Operativa em Mecânica pela Escola Naval no ano de 1973. Cursou Análise de Sistemas pela Marinha através o seu Curso Especial de Formação em Técnicos em Análise de Sistemas em 1981. Desempenhou diversas funções voltadas à área de informática na Marinha, tais como : Chefe do Departamento de Sistemas da Diretoria do Pessoal Militar da Marinha, Chefe do Departamento de Informática na Base Naval de Aratu e Acessor de Informática no Segundo Distrito Naval. Ministrou as disciplinas de Bancos de Dados e Estrutura de Dados na Sociedade de Ensino Superior e Assessoria Técnica, (Faculdade Anglo-Americano) em 1986 e 1987, e da disciplina de Bancos de Dados na Universidade Nuno Lisboa (UniverCidade) no mesmo período.

Ficha Catalográfica

Gomes, Paulo Roberto

Um estudo sobre avaliação da execução do BLAST em ambientes distribuídos / Paulo Roberto Gomes ; orientador: Sérgio Lifschitz. – 2009.

128 f. : il. (color.) ; 30 cm

Dissertação (Mestrado em Informática)– Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2009.

Inclui bibliografia

CDD: 004

À minha mãe Lucy e a minha esposa Glória
pelo seu apoio, mesmo em horas difíceis.

Agradecimentos

Ao Professor Sérgio Lifschitz por sua orientação segura e incentivo, que possibilitou a elaboração da presente dissertação.

Ao amigo Daniel Xavier de Sousa, que generosamente compartilhou seus conhecimentos sobre ambientes distribuídos e ferramentas BLAST, sempre que foi solicitado.

Ao amigo Jacques Alves da Silva que viabilizou a utilização dos equipamentos da UFF para elaboração dos testes em ambiente de *grid*, os quais compõem uma parcela significativa da presente dissertação.

Aos meus colegas do Laboratório de Bioinformática, por sua ajuda, sempre que necessário.

À minha esposa e meus filhos, pelo seu incentivo e compreensão.

À minha mãe que sempre me apoiou e incentivou, e sem a qual, não seria possível concluir o curso.

Resumo

Gomes, Paulo Roberto; Lifschitz, Sergio. **Um Estudo sobre Avaliação da Execução do BLAST em Ambientes Distribuídos**, Rio de Janeiro, 2009. 128p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Ferramentas BLAST são normalmente utilizadas para efetuar comparações entre seqüências de DNA, RNA e proteínas. No entanto, face ao crescimento exponencial das bases biológicas, existe uma preocupação quanto ao desempenho do BLAST, mesmo considerando os equipamentos de grande capacidade computacional hoje existente. Considerando tal fato, algumas ferramentas capazes de executar o BLAST em ambientes distribuídos, tais como clusters e *grids*, vêm sendo desenvolvidas de modo a acelerar consideravelmente a sua execução. No entanto, até o presente momento, não foi constatado, na literatura existente, nenhum estudo com o objetivo de comparar o desempenho entre essas ferramentas. A avaliação de desempenho dessas ferramentas é normalmente efetuada de forma isolada, considerando apenas o tempo de execução (*elapsed time*), em situações diversas, como, por exemplo, variando o número de nós em que a ferramenta BLAST é executada.. Almejando uma investigação mais detalhada, principalmente no que diz respeito a avaliação de desempenho do BLAST em ambientes distribuídos, a presente dissertação tem como um dos seus objetivos efetuar um estudo detalhado sobre como comparar o desempenho do BLAST em um ambiente distribuído, considerando para tal, a avaliação de três ferramentas BLAST, dentre elas o balaBLAST, desenvolvida no Laboratório de Bioinformática da PUC-Rio. O segundo objetivo é verificar a eficácia do balanceamento de carga efetuado pela ferramenta balaBLAST.

Palavras-chave

BLAST, Seqüência, DNA, RNA, Proteínas, Avaliação, Balanceamento de Carga, Desempenho.

Abstract

Gomes, Paulo Roberto; Lifschitz, Sergio (Advisor). **A Study on Evaluation of Implementation of BLAST in a Distributed Environment**, Rio de Janeiro, 2009. 128p. MsC. Dissertation – Departament of Informática, Pontifícia Universidade Católica do Rio de Janeiro.

BLAST tools are typically used to make comparisons between sequences of DNA, RNA and proteins. However, given the exponential growth of the biological databases, there is concern about the performance of BLAST, even considering the equipment of large computing power that exists today. Considering this fact, some tools to run BLAST in distributed environments such as clusters and grids, have been developed to greatly accelerate its performance. However, until now, has not been found in existing literature, no study in order to compare the performance between these tools. The performance evaluation of these tools is usually done in isolation, considering only the execution time (elapsed time) in different situations, for example, varying the number of nodes in the tool BLAST runs. Craving a more detailed investigation, especially with regard to performance evaluation of BLAST in distributed environments, this dissertation has as one of your goals make a detailed study to compare the performance of BLAST in a distributed environment, considering for such the evaluation of three tools BLAST, among them the balaBLAST developed in the Bioinformatics Laboratory of PUC-Rio. The second objective is to verify the effectiveness of load balancing performed by the tool balaBLAST.

Keywords:

BLAST, Sequence, DNA, RNA, Proteins, Evaluate, Load Balancing, Performance.

Sumário

Nenhum trabalho de qualidade pode ser feito sem concentração e auto-sacrifício, esforço e dúvida.

Max Beerbohm, *Frases e Pensamentos*