# Selected Readings in COMPUTING AND TELECOMMUNICATIONS

Editors
Mira Kartiwi
Teddy Surya Gunawan
Aisha Hassan Abdalla Hashim

IIUM Press
INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

# SELECTED READINGS IN COMPUTING AND TELECOMMUNICATIONS

**Editors**
Mira Kartiwi
Teddy Surya Gunawan
Aisha Hassan Abdalla Hashim

**IIUM Press**

Published by:
IIUM Press
International Islamic University Malaysia

First Edition, 2011


Perpustakaan Negara Malaysia Cataloguing-in-Publication Data

Mira Kartiwi, Teddy Surya Gunawan, Aisha Hassan Abdalla Hashim:
Selected Readings in Computing and Telecommunications

**ISBN: 978-967-0225-81-4**

Member of Majlis Penerbitan Ilmiah Malaysia – MAPIM
(Malaysian Scholarly Publishing Council)

Printed by:
**IIUM PRINTING SDN. BHD.**
No. 1, Jalan Industri Batu Caves 1/3
Taman Perindustrian Batu Caves
Batu Caves Centre Point
68100 Batu Caves
Selangor Darul Ehsan

# TABLE OF CONTENT

# 19. PROTEIN CODING IDENTIFICATION OF HUMAN DNA SEQUENCE USING WAVELET

Teddy Surya Gunawan and Norashikin Abdul Wahab

## ABSTRACT

An important topic in genomic sequence analysis is the identification of protein coding regions. In this context, several coding DNA model-independent methods based on the occurrence of specific patterns of nucleotides at coding regions have been proposed. Nonetheless, these methods have not been completely suitable due to their dependence on an empirically predefined window length required for a local analysis of a DNA region. This project introduces a method based on a Wavelet for the identification of protein coding regions. This novel transform is tuned to analyze periodic signal components and presents the advantage of being independent of the window length. Here, this paper compared the performance of the Wavelet with FFT methods by using eukaryote data sets. The results show that Wavelet outperforms all assessed model-independent methods with respect to identification accuracy. These results indicate that the source of at least part of the identification errors produced by the previous methods is the fixed working scale. The new method not only avoids this source of errors but also makes a tool available for detailed exploration of the nucleotide occurrence.

## 19.1 INTRODUCTION

The evolution of genetic information and the generation of genes is one of the most challenging problems facing evolutionary and molecular biologists. The computational identification of genes and coding regions in DNA