

## INTISARI

### **Pengkodean atas Himpunan Berbasis Jarak Subset Sekuen**

Oleh

MUHAMMAD RAID AKRAM

19/440076/PA/19065

Teori pengkodean adalah ilmu yang mempelajari cara mengoreksi informasi yang ditransmisikan melalui *kanal* secara matematis. Kanal yang dikaji di dalam skripsi ini adalah penyimpanan DNA. Berbeda dengan kode konvensional, informasi yang digunakan pada kanal ini dapat dimodelkan sebagai himpunan dari beberapa sekuen. Di dalam model yang digunakan, galat yang terjadi dapat berupa perubahan simbol, penghapusan sekuen, dan penambahan sekuen. Untuk itu, dibentuk sistem metrik baru yaitu *jarak subset sekuen* antara 2 buah himpunan sekuen sama panjang. Berdasarkan metrik ini, dikonstruksi kode subset sekuen beserta algoritma *encoding* dan *decoding* untuk mengoreksi galat yang terjadi. Skripsi ini membahas sifat-sifat dari jarak subset sekuen beserta konstruksi kodenya. Masing-masing kode dikonstruksi dengan memanfaatkan kode konvensional. Selain itu, dibahas pula beberapa batas atas banyak anggota dari kode subset sekuen dengan ukuran *codeword* tetap. Meskipun begitu, kode-kode yang dibentuk masih belum dapat mencapai batas tersebut. Lebih lanjut, disajikan pula contoh algoritma *encoding* dan *decoding* dari kode-kode tersebut beserta kalkulasinya menggunakan program. Algoritma tersebut dibuat dengan memanfaatkan kode Reed-Solomon.

## ABSTRACT

### Coding Over A Set Based on Subset Sequence Distance

By

MUHAMMAD RAID AKRAM

19/440076/PA/19065

Coding theory studies the correction of transmitted or stored information mathematically. Information can be transmitted in various forms, depending on the channel. This undergraduate thesis mainly focuses on the DNA storage channel. Unlike conventional code, information in this channel can be modeled as a set of DNA sequences. In this model, we consider 3 types of error: symbol substitution, sequence deletion, and sequence insertion. A new metric, termed *sequence-subset distance*, is introduced to combat those errors. This metric measures the distance between 2 sets of sequences with equal length. Based on this metric, a sequence-subset code is constructed. This undergraduate thesis studies the property of subset sequence distance and code construction based on that distance. The subset sequence code is constructed using conventional codes. We also study some upper bounds for the cardinality of subset sequence code. However, the constructed code does not achieve these upper bounds. The author further proposes encoding and decoding algorithms of the constructed codes and their calculation using a program. The algorithm is constructed using Reed-Solomon code.