

UNIVERSITI TEKNOLOGI MARA

**SPLICING SYSTEM AND THEIR
VARIANTS WITH FUZZY
ALGEBRAIC SUM**

MOHD PAWIRO SANTONO BIN OTHMAN

Thesis submitted in fulfillment
of the requirements for the degree of
Master of Science
(Mathematics)

Faculty of Computer and Mathematical Sciences

July 2023

ABSTRACT

This research discusses the concept of biological systems in Deoxyribonucleic Acid (DNA) which involves formal language theory namely splicing system. In DNA computing, a splicing system combines theoretical computer science and applied discrete mathematics. This splicing model is based on the recombinant behavior of DNA molecules where the model of splicing with finite components can only generate languages of lower generative power in the Chomsky hierarchy. Hence, in this research, an extension of the original models of splicing systems using fuzzy algebraic sum operation is introduced as a restriction in splicing systems and their variants to increase the generative power of the splicing languages. Moreover, the languages generated by algebraic sum fuzzy splicing systems and their variants are developed by considering the threshold and the languages generated are categorized according to the Chomsky hierarchy. An algorithm is then developed in JAVA using visual code studio software to replace the time-consuming manual computation of the languages generated by algebraic sum fuzzy splicing systems and their variants. As conclusion, it is shown that the generative power of the languages generated by algebraic sum fuzzy splicing systems and their variants are up to the context-free than those of the original splicing models. The results in a form of theorems, lemmas and corollaries are proven and presented.

ACKNOWLEDGEMENT

I would like to extend my gratitude to those who had helped me throughout my studies towards the successful completion of this thesis. First and foremost, I would like to thank my supervisor, Dr. Mathuri Selvarajoo, for her guidance, support, patience, suggestions and invaluable encouragement during this research. I am grateful for her supervision and would not have achieved this study without her.

My appreciation also goes to my co-supervisors, Prof. Dr. Isamiddin Rakhimov for his valuable criticism and suggestions throughout this research. Moreover, I would also like to thank Assoc. Prof. Dr. Fong Wan Heng for her significant contributions, suggestions and guidance. I particularly thank Mr Shafri Amir from UiTM for his guidance and advice on designing an algorithm for generating algebraic sum fuzzy splicing languages using JAVA programming.

Moreover, I am also indebted to Ministry of Higher Education and Universiti Teknologi MARA Malaysia for the financial funding through research grant FRGS-RACER (600-IRMI/FRGS-RACER 5/3 (050/2019)).

Last but not least, I wish to express my appreciation to my parents for their support and help throughout the tough time of my research. This piece of victory is dedicated to both of you. Alhamdulillah.

TABLE OF CONTENTS

	Page
CONFIRMATION BY PANEL OF EXAMINERS	ii
AUTHOR'S DECLARATION	iii
ABSTRACT	iv
ACKNOWLEDGEMENT	v
TABLE OF CONTENTS	vi
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF SYMBOLS	x
LIST OF ABBREVIATIONS	xiii
CHAPTER ONE: INTRODUCTION	1
1.1 Introduction	1
1.2 Research Background	2
1.3 Problem Statement	3
1.4 Research Questions	5
1.5 Objectives of Research	5
1.6 Scope of the Research	5
1.7 Significance of Research	6
CHAPTER TWO: LITERATURE REVIEW	8
2.1 Introduction	8
2.2 Background of Formal Language Theory	8
2.3 Splicing System	12
2.4 Fuzzy Set Theory	20
2.5 Conclusion	23
CHAPTER THREE: METHODOLOGY	24
3.1 Introduction	24

CHAPTER ONE

INTRODUCTION

1.1 Introduction

Every living organism has a unique deoxyribonucleic acid (DNA). The structure of DNA was firstly introduced in 1953 by Watson and Crick as a double-helical form [1]. DNA molecules are constructed from monomers called nucleotides. Nucleotide consist of three components: sugar, phosphate, and base [2]. These structures of nucleotide are different from each other by the sequence of their bases namely Adenine, Guanine, Cytosine, and Thymine abbreviated as A, G, C, and T respectively. These bases are tied together by hydrogen bonds using base-complementary rules, where A pairs with T, G pairs with C and vice versa [1].

Splicing systems were first proposed by Head in 1987 as a mathematical model of the recombinant behavior of double-stranded DNA (dsDNA) and the enzymes that cut and paste dsDNA [3]. Restriction enzymes, which are found naturally in bacteria, can cut DNA fragments at certain sequences called restriction sites while ligases, on the other hand, can re-join DNA fragments with complementary ends [2]. This model consists of a finite alphabet, V , a finite set of initial strings over an alphabet, A , and a finite set of rules, R , that act upon the strings by iterative cut and paste operation in order to generate new strings [1]. Splicing language is a language generated by the splicing system and this languages are classified based on their generative power according to the Chomsky hierarchy [4]. It has been demonstrated that all splicing languages generated with finite sets of axioms and rules are regular where regular languages is the lowest level of languages in Chomsky hierarchy in term of the generative power [5].

A variety of studies have been done in order to increase the generative power of the languages generated by splicing systems up to the recursively enumerable languages. From the perspective of DNA computing, splicing systems that generate languages with higher generative power can be considered as theoretical models of universally programmable DNA-based computers. However, generating languages with generative power comparable to the Turing machine is still difficult [4]. Hence, in