A Comparative Literature Survey of Splicing System

F. Amjad Basha
Assistant Professor
Department of Mathematics
The New College, Chennai, India.

Abstract - This paper describes an extensive literature survey about splicing system. The operation of splicing have recently become of interest in many fields such as molecular computing and cryptography etc. The purpose of this article is to present a comprehensive survey about the recent developments on splicing system, including a clear description of study purpose and key findings in this area, with the hope of spurring further research.

Keywords - Splicing System, DNA Sequence, Recursively Enumerable Languages.

I. INTRODUCTION

A new operation on strings, called splicing is introduced in [1], as a formal model of certain cut and paste biochemical transformation processes of an initial collection of DNA strands under the simultaneous influence of enzymes. Two strands of DNA are cut at specified substrings (sites) by restriction enzymes that recognize a pattern inside the molecule and then the fragments are pasted by ligase enzymes. The system is described by an initial set of words and a finite set of rules. The language generated is the closure of the initial set under the application of splicing rules. A splicing rule is a quadruplet of words, usually written as α # β $ δ # γ. The words α, β, γ, δ are called the handles of the rule. A rule indicates where to cut and what to paste. Since 1987, his basic idea has been formalized in terms of generative mechanisms for formal languages, called splicing systems.

Requests for new applications [7] from the world of computing and technology are stimulating interest towards computational models inspired by biological mechanisms. Molecular computing is one of the research directions in this framework which is strictly related to formal languages theory, is based on the notion of splicing systems.

II. TYPES OF SPLICING

2.1 Linear splicing

A Paun splicing system is a triple $S = (A, I, R)$, where $I \subset A^*$ is a set of strings, called initial language, R is a set of rules $r = u_1 | u_2 \$ u_3 | u_4$, with $u_i \in A^*$, $i = 1,2,3,4$ and |, $ $\in A$. Given two words $x = x_1 u_1 u_2 x_2$, $y = y_1 u_3 u_4 y_2$, x_1 , x_2 , y_1, $y_2 \in A^*$ and the rule r = u_1 | u_2 $ u_3 | u_4, the splicing operation produces $w = x_1 u_1 u_4 y_2$ and $w = y_1 u_3 u_2 x_2$.

2.2 Circular splicing

A Paun circular splicing system is a triple $S = (A, I, R)$, where A is a finite alphabet, I is the initial circular language, with $I \subseteq \sim A^*$ and R is the set of the rules, with $R \subseteq A^* \# A^* \$ A^* \# A^*$ and #, $ $\notin A$. Then, given a rule $r = u_1 \# u_2 \$ u_3 \# u_4$ and circular words $\sim w'$, $\sim w''$, $\sim w$, we set $(\sim w', \sim w'') \vdash_r \sim w$ if there are linearizations w' of $\sim w'$, w'' of $\sim w''$, w of $\sim w$ such that $w' = u_2 x u_1$, $w'' = u_4 y u_3$ and $w = u_2 x u_1 u_4 y u_3$. If $(\sim w' \sim w'') \vdash_r \sim w$ we say that $\sim w$ is generated (or spliced) starting with w', $\sim w''$ and by using a rule r. We also say that $u_1 u_2$, $u_3 u_4$ are sites of splicing and we will use SITES (R) to denote the set of sites of the rules in R.

Example 2.2.1

Consider the circular splicing system over $A = \{a, b\}$, with initial set $I = \{\sim ab\}$ and with the single rule $\langle a \mid a - b \mid b \rangle$. The rule expresses the fact that a word starting with the letter a and ending with a letter b can be inserted, in a circular word, between a letter a followed by a letter b. As a consequence, the set generated by the system is the $\sim \{a^n b^n | n \geq 1\}$.

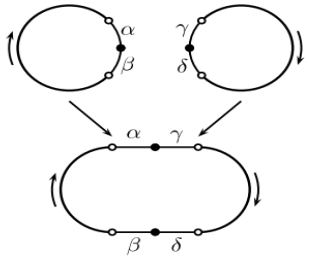

Figure 1. Circular Splicing

2.3 Flat splicing

A flat splicing system, or a splicing system for short, is a triplet $S = (A, I, R)$, where A is an alphabet, I is a set of words over A, called the initial set and R is a finite set of splicing rules, which are quadruplets $\langle \alpha \mid \gamma - \delta \mid \beta \rangle$ of words over A. The words α, β, γ and δ are called the handles of the rule. Let $r = \langle \alpha \mid \gamma - \delta \mid \beta \rangle$ be a splicing rule. Given two words $u = x\alpha \cdot \beta y$ and $v = \gamma z \delta$, applying r to the pair (u, v) yields the word $w = x\alpha \cdot \gamma z \delta \cdot \beta y$. (The dots are used only to mark the places of cutting and pasting, they are not parts of the words.) This operation is denoted by $u, v \vdash_r w$ and is called a production. Note that the first word (here u) is always the one in which the second word (here v) is inserted.

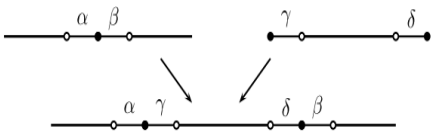

Figure 2. Flat Splicing

Remark: The difference between the above definitions depends on the biological phenomena that needs to model.

Table: 1 Literature Review Summary Table

Topic	Ref	Year	Study Purpose	Key Findings
On the splicing operations.	[2]	1995	To consider both simple and iterated splicing, with respect to a finite or an infinite set of splicing rules.	Relations between these operations and usual operations with languages are investigated.
Computing by splicing.	[3]	1996	To introduce a new model of splicing.	Characterization of recursively languages.
Simple splicing system.	[4]	1998	To start with an opposite approach in splicing rules which are as simple as possible.	A series of language theoretic properties of languages are investigated.
Splicing on tree like structures.	[5]	1999	To provide a method to increase the power of splicing system.	Splicing system on trees with finite set of axioms and rules can generate the class of context free languages.
Multiple splicing systems and the universal computability.	[6]	2001	To extend the original splicing system based on a kind of logic grammars, called elementary formal systems (EFS).	A new extension of splicing systems, called multiple splicing systems has been introduced.
An alternate definition of splicing.	[8]	2006	To propose a new definition of the language generated by a splicing system.	Finite extended H systems can generate all recursively enumerable languages.
Complexity theory for splicing system.	[9]	2007	To propose a notion of time complexity in splicing system.	Regular and context free languages are in special time (SPLTIME).
A characterization of regular circular languages generated by marked splicing system.	[10]	2009	Some unanswered questions are related to the computational power of circular splicing systems.	Introduced a special class of the latter systems called marked systems.

A characterization of (regular) circular languages generated by monotone complete splicing systems.	[11]	2010	To characterize (regular) circular languages generated by monotone complete systems.	To point out that monotone complete systems have the same computational power as finite simple systems.
On the regularity of circular splicing languages: a survey and new developments.	[12]	2010	To focus on the relationship between regular circular languages and languages generated by finite circular splicing systems.	To prove that marked systems with self-splicing generate only regular circular languages.
Splicing systems and the Chomsky hierarchy.	[13]	2012	To prove decidability properties and new results on the position of the family of languages generated by (circular) splicing systems within the Chomsky hierarchy.	A new variant in circular splicing called flat splicing has been introduced.
Splicing System in Automata Theory: A Review.	[14]	2019	To study the design of automated enzymatic processes in Automata Theory.	The regular language can be implemented in the splicing system to show the DFA structure in the splicing system.
Descriptional Complexity of semi-simple splicing systems.	[15]	2020	To focus on restricted finite splicing systems.	To investigate the descriptional complexity of semi-simple splicing systems.
Characterization of central splicing.	[16]	2021	To propose a new extension of splicing systems, called central splicing system.	The relationship between non synchronized pure pattern languages and recursively enumerable languages are discussed.

III. IMPORTANT RESULTS

- The language generated by a context-sensitive flat splicing system is context-sensitive.
- The language generated by a circular alphabetic context-free splicing system is context-free.
- The language generated by a flat alphabetic context-free splicing system is context-free.
- If a family F is closed under union, concatenation with symbols, intersection with regular sets, and unary splicing operation, then F is closed under concatenation.
- Extended non-reflexively evolving H systems generate all recursively enumerable languages.
- *Let* $S = (A_1, I_1, R_1)$, $S' = (A_2, I_2, R_2)$ be two Paun circular splicing systems. If $A_1 \subseteq A_2$, $I_1 \subseteq I_2$ and $R_1 \subseteq R_2$ then $L(S) \subseteq L(S')$.
- Given a marked system S, we can decide whether $L(I, R)$ is a regular circular language. Given a regular circular language C (over a finite alphabet I) we can decide whether a marked system $S = (I, R)$ exists such that $C = L(I, R)$.

IV. CONCLUSIONS

Splicing is a powerful tool applied in various fields such as Biology, Computer Science and Mathematics etc. This paper describes the scenario of the review papers which helps the researchers to proceed further. Here I conclude that some open problems are yet to solve in the above mentioned papers.

REFERENCES

[1] T. Head, Formal language theory and DNA: an analysis of the generative capacity of specific recombinant behaviors, Bull. Math. Biol., 49, No. 5, (1987), 737- 759.

[2] Gheorghe Paun, On the splicing operations, Discrete Applied Mathematics 70 (1996) 57- 79.

[3] Gheorghe Paun, Grzegorz Rozenberg and Arto Salomaa, Computing by splicing, Theoretical Computer Science 168 (1996) 321 - 336.

[4] A. Mateescu, Gh. Paun and G. Rozenberg, Simple splicing system, Discrete Applied Mathematics 84 (1998) 145 - l63.

[5] Yasubum Sakakibar and Claudio Ferretti, Splicing on tree like structures, Theoretical Computer Science 210 (1999) 227 - 243.

[6] Satoshi Kobayashi and Yasubumi Sakakibara, Multiple splicing systems and the universal computability, Theoretical Computer Science 264 (2001) 3 - 23.

[7] Paola Bonizzoni, Clelia De Felice, Giancarlo Mauri and Rosalba Zizza, Circular splicing and regularity, Rairo -Theoretical informatics and applications 38 (2004) 189-228.

[8] Remco Loos, An alternate definition of splicing, Theoretical Computer Science 358 (2006), 75 - 87.

[9] Remco Loos and Mitsunori Ogihara, Complexity theory for splicing system, Theoretical Computer Science 386 (2007) 132 - 150.

[10] Clelia De Felice, Gabriele Fici and Rosalba Zizza, A characterization of regular circular languages generated by marked splicing system, Theoretical Computer Science 410 (2009), 4937- 4960.

[11] Paola Bonizzoni, Clelia De Felice and Rosalba Zizza, A characterization of (regular) circular languages generated by monotone complete splicing systems, Theoretical Computer Science 411 (2010) 4149 - 4161.

[12] Paola Bonizzoni, Clelia De Felice, Gabriele Fici and Rosalba Zizza, On the regularity of circular splicing languages: a survey and new developments, Natural Computing, June 2010.

[13] Jean Berstel, Luc Boasson and Isabelle Fagnot, Splicing systems and the Chomsky hierarchy, Theoretical Computer Science 436 (2012) 2 - 22.

[14] S H Khairuddin, M A Ahmad and N Adzhar, Splicing System in Automata Theory : A Review , Journal of Physics: Conference Series 1366, (2019).

[15] Nataša Jonoska and Dmytro Savchuk, Descriptional Complexity of semi-simple splicing systems, Developments in Language Theory, (2020) May 26; 12086: 150 - 163.

[16] F. Amjad Basha, Sindhu J Kumaar and Faiz Ahmed, Characterization of central splicing, GIS Science Journal, Vol 8, Issue 5, (2021), 1052 – 1056.