Author: Rosalba Zizza
Title: On the power of classes of splicing systems
Language of presentation: English
Supervisors: Prof. Giancarlo Mauri (Univ. of Milano-Bicocca, Italy),
Prof. Clelia De Felice (Univ. of Salerno, Italy)
Date of defence: January 11, 2002
Institute granting degree: University of Milano, Italy

Abstract

"It is hoped that the line of research initiated here will interest biologists who have not previously found reason to be concerned with formal language theory... The results discussed here are certainly only modest formal achievements. It is hoped that future developments will provide much stronger results following the given paradigm and yield wholly new paradigms as well (T. Head, 1987)"

My thesis provides an overview and investigations on Molecular Computing and Splicing Systems. Molecular Computing is a novel branch connecting Computer Science, Mathematics and Biology: computation is carried on by DNA molecules and the operations are performed by using enzymes and other biological processes. Splicing Systems were introduced by Tom Head in 1987, as formal models for the *splicing*, a recombinant generative process among molecules which are cut and pasted by a simultaneous action of restriction and ligase enzymes on DNA molecule. There is a natural definition of this phenomenon in a formal language framework, defining "splicing" as an operator on strings. As DNA occurs in both linear and circular form, two kinds of splicing operation were defined to model different biological phenomena. Even if different definitions of splicing have been given in the literature, the linear splicing operation concatenates a prefix of one string with a suffix of another string, under some conditions, represented into a (splicing) rule. On the contrary, the circular splicing opens two circular strings and pastes them together, obtaining a circular string, under some conditions dictated by a (splicing) rule. A Splicing System is defined by giving the initial set of strings (or axioms) I and the set of rules R (or triples, according to which definition we use). Thus the set of all possible molecules generated in the test tube by splicing is nothing but the language (called *splicing language*) containing I and closed under the splicing operation. The generative power of splicing systems, i.e., the type of languages that are generated by it, has been investigated thoroughly, and depends on which level of the Chomsky hierarchy I, R belong to. In spite of a vast literature on splicing systems, a few problems related to their computational power are still open.

In this thesis we intend to evidence how classical techniques and definitions in automata theory are a legitimate tool for investigating a few of these open problems. We study finite linear and circular splicing systems, i.e., splicing systems with both I, R finite sets, in a perspective which has never been explored so far. It is known that the class of languages generated by finite splicing systems is a proper subclass of the class of regular languages. One of the aims is the characterization of this subclass, which is still unknown, except for some special cases. This has allowed us to compare splicing systems and finite state automata and find new models generating infinite (but regular) languages, starting with finite languages. We have compared the different definitions of splicing and several examples of (formal or circular) regular languages, generated or not generated by finite splicing. The research has succeded in defining new families of regular languages by means of automata or combinatorial properties and proving them to be splicing languages. Decidability and descriptional complexity questions are answered by using combinatorics on words, classical issues in Formal Languages and novel ideas on automata. In order to characterize regular languages generated by linear splicing, we have used the syntactic monoid, so stating a sufficient condition for being linear splicing generated. We argue that this approach, we are already using, can result effective for bounding the length of splicing rules and for obtaining the characterization of the sublclass of regular languages generated by finite linear splicing. Deep connections with Theory of (variable length) Codes have been discovered and used for providing circular regular languages generated by finite circular splicing. A circular regular language not generated by splicing has been presented, showing that, at least in the simplest formulation, circular splicing cannot be used to characterize circular regularity, which is still an open problem. The present study highlights that, up to now, no circular regular splicing language exist which is not contained in one of the classes we have defined. In the case of one-letter alphabet it has been proved to be decidable whether or not a regular language is generated by a finite circular splicing system. Problems and conjectures for future work are also proposed. A flash of biochemical background, thought out of our scope, is provided together with an extensive bibliography.

Table of contents

Author's correspondence address

Rosalba Zizza
Dipartimento di Informatica ed Applicazioni "R.M. Capocelli"
Università di Salerno
Via Allende - 84081 Baronissi (Salerno)
E-mail: `zizza@unisa.it`
URL: `http://web.tiscalinet.it/RosalbaZizza`