

CROSSOVER SYSTEMS. A GENERALIZATION OF SPLICING SYSTEMS

VICTOR MITRANA

*Faculty of Mathematics, University of Bucharest
Str. Academiei 14, R-70109 Bucharest, Romania
e-mail: mitrana@funinf.math.unibuc.ro*

ABSTRACT

In this paper we consider the crossover operation as an operation on strings and languages and study the relationships between this operation and other operations in formal language theory. Roughly speaking, crossover produces a new sequence starting from two parent sequences, cutting them in several fragments and crossing these fragments by ligation. Thus, crossover is a generalization of the splicing operation viewed as a formal model of DNA recombination.

Keywords: formal languages, crossover systems and languages.

1. Introduction

The fundamental mechanism by which genetic material is merged is *recombination*. DNA sequences are recombined under the effect of enzymatic activities. In [4] T. HEAD introduced an operation on strings called *splicing*, with motivations closely related by DNA recombinant behaviours; we refer this paper and also [5] for motivations and further references.

The main idea of the splicing operation is that two sequences are cut at specified sites, and the first subword of one is pasted to the second segment of the other.

Roughly speaking, *crossover* produces a new sequence starting from two parent sequences, cutting them in several fragments and crossing these fragments by ligation. The new sequence is formed by starting at the left end of one parent, copying a substring, crossing over to some site in the other parent, copying a substring, crossing back to some site in the first parent and so on. The process ends when the right end of one parent is reached. As one can easily see, splicing is a particular case of crossover. The mechanism of crossover was considered from different points of view, in this respect, and for more details, the reader is referred to [3, 6].

In accordance with [2, 9] the languages defined by splicing systems starting from regular languages are regular, too. Moreover, each splicing language generated non-iteratively from a context-free language is also context-free. On the other hand, it is commonly asserted [1, 11] that the structural linguistics of the DNA language cannot be modelled by context-free languages. Crossover systems seem to be more