Advances in Applied Mathematics 75 (2016) 56–93

Contents lists available at ScienceDirect

Advances in Applied Mathematics

www.elsevier.com/locate/yaama

Operads, quasiorders, and regular languages

Samuele Giraudo [a,*], Jean-Gabriel Luque [b], Ludovic Mignot [b], Florent Nicart [b]

[a] *Laboratoire d'Informatique Gaspard-Monge, Université Paris-Est Marne-la-Vallée, 5 boulevard Descartes, Champs-sur-Marne, 77454 Marne-la-Vallée cedex 2, France*

[b] *Laboratoire LITIS – EA 4108, Université de Rouen, Avenue de l'Université – BP 8, 76801 Saint-Étienne-du-Rouvray Cedex, France*

ARTICLE INFO

Article history:
Received 20 October 2015
Accepted 5 January 2016
Available online 19 January 2016

MSC:
05E99
68Q45
18D50

Keywords:
Operad
Multi-tilde
Language
Regular language
Quasiorder

ABSTRACT

We generalize the construction of multi-tildes in the aim to provide double multi-tilde operators for regular languages. We show that the underlying algebraic structure involves the action of some operads. An operad is an algebraic structure that mimics the composition of the functions. The involved operads are described in terms of combinatorial objects. These operads are obtained from more primitive objects, namely precompositions, whose algebraic counter-parts are investigated. One of these operads acts faithfully on languages in the sense that two different operators act in two different ways.

0. Introduction

Following the Chomsky–Schützenberger hierarchy [9], regular languages are defined to be the formal languages that are generated by Type-3 grammars (also called reg-

* Corresponding author.
E-mail addresses: samuele.giraudo@univ-mlv.fr (S. Giraudo), jean-gabriel.luque@univ-rouen.fr (J.-G. Luque), ludovic.mignot@univ-rouen.fr (L. Mignot), florent.nicart@univ-rouen.fr (F. Nicart).

http://dx.doi.org/10.1016/j.aam.2016.01.002
0196-8858/

ular grammars). These particular languages have been studied for several years since they have many applications in several areas such as pattern matching, compilation, verification, and bioinformatics. Their generalization as rational series links them to various algebraic or combinatorial topics like enumeration (manipulation of generating functions), rational approximation (for instance Pade approximation), representation theory (modules viewed as automata), and combinatorial optimization ($(\max, +)$-automata).

One of the main specificities of regular languages is that they can be represented by various tools: regular grammars, automata, regular expressions, etc. Whilst they can be represented by both automata and regular expressions [13], these two tools are not equivalent. Indeed, Ehrenfeucht and Zeiger [10] showed a one parameter family of automata whose shortest equivalent regular expressions have a width exponentially growing with the numbers of states. Note that it is possible to compute an automaton from a regular expression E such that the number of its states is a linear function of the alphabet width (i.e., the number of occurrences of alphabet symbols) of E [2,8,11,18].

In the aim to increase expressiveness of regular expressions for a bounded length, Caron et al. [6] introduced the so-called multi-tilde operators and applied these to represent finite languages. Investigating the equivalence of two multi-tilde expressions, they define a natural notion of composition which endows the set of multi-tilde operators with a structure of operad. This structure has been investigated in [14].

Originating from the algebraic topology [17,3], operad theory has been developed as a field of abstract algebra concerned by prototypical algebras that model classical properties such as commutativity and associativity [15]. Generally defined in terms of categories, this notion can be naturally applied to computer science. Indeed, an operad is just a set of operations, each one having exactly one output and a fixed finite number of inputs, endowed with the composition operation. An operad can then model the compositions of functions occurring during the execution of a program. In terms of theoretical computer science, this can be represented by trees with branching rules. The whole point of the operads in the context of the computer science is that this allows to use different tools and concepts from algebra (such as morphisms, quotients, substructures, generating sets).

In order to illustrate this point of view, let us recall the main results of our previous paper [14]. In this paper, we first showed that the set of multi-tilde operators has a structure of operad. We used the concept of morphism in the aim to choose the operad allowing us to describe in the simplest way a given operation or a property. For instance, the original definition of the action of the multi-tildes on languages is rather complicated. But, *via* an intermediate operad based on set of boolean vectors, the action was described in a more natural way. In the same way, the equivalence problem is clearer when asked in a operad based on antisymmetric and reflexive relations which is isomorphic to the operad