Journal of mathematics and computer science 14 (2015), 250-257

Contents list available at JMCS
Journal of Mathematics and Computer Science
Journal Homepage: www.tjmcs.com

Some Algebraic Structures of Languages

Dasharath Singh[1,*], Ahmed Ibrahim Isah[1,+]

[1]Mathematics Department, Ahmadu Bello University, Zaria, Nigeria.
[*]*mathdss@yahoo.com*
[+]*aisah204@gmail.com*

Article history:
Received November, 2014
Accepted January, 2015
Available online January 2015

Abstract

In this paper, suitable operations are defined on the class of partitions of a language which give rise to certain monoids and semigroups. In particular, certain algebraic structures of a language defined over a string are described.

Keywords: Language, partition, semigroup, monoid

Mathematics Subject Classification: 11P83, 68Q45, 68Q70

1. Introduction

The applications of various algebraic structures abound (see [1, 2, 3, 6, 7, 8] for details and related references). In particular, certain algebraic structures have found applications in formal language theory (see [6] for details). Moreover, a number of algebraic structures of partitions of a set and that of an integer have been developed which have useful applications in computer arithmetic, formal languages and sequential machines (see [5, 7] for details). In this paper, suitable operations on the set of partitions of a language are defined which give rise to certain monoids and semigroups. In addition, certain algebraic structures of a language defined over a string are described.

2. Definitions

Definition 2.1 Union, Intersection and Concatenation of Languages

Let X be an alphabet and X^* denote the set of all strings over X. A *language* L is a subset of X^* i.e., $L \subseteq X^*$. Let L_1 and L_2 be any two languages over X. The union of L_1 and L_2, denoted $L_1 \cup L_2$, is the language $L_1 \cup L_2 = \{u \in X^* \mid u \in L_1$ or $u \in L_2\}$. The intersection of L_1 and L_2, denoted $L_1 \cap L_2$, is the language $L_1 \cap L_2 = \{u \in X^* \mid u \in L_1$ and $u \in L_2\}$. The concatenation (or simply, catenation) of L_1 and L_2, denoted L_1L_2, is the language $L_1L_2 = \{u = u_1u_2 \mid u_1 \in L_1$ and $u_2 \in L_2\}$. It is immediate to see that the union, intersection and catenation of languages are each associative because union, intersection and catenation of strings are each associative and hence X^* with catenation is a non-commutative monoid (see [2, 4], for details). In the same vein, let u^* be defined as the set of all strings over $u \in X^*$, then u^* with catenation is a commutative monoid.

Definition 2.2 Cardinality bounded languages

Let X^{n*}, henceforth called a *cardinality bounded language* over X, denote the set of all strings of length $\leq n$ over X. In other words, $\{X^{n*}\}$ is a strictly monotonic increasing nested sequence, and obviously $X^* = X^{0*} \cup X^{1*} \cup \ldots \cup X^{n*} \cup \ldots$. However, it gives an alternative form of representation of the usual one viz., $X^* = X^0 \cup X^1 \cup \ldots \cup X^n \cup \ldots$, where X^n is the set of all strings of length n over X. Moreover, $\bigcup_{n=0}^{\infty} X^n = \bigcup_{n=0}^{\infty} X^{n*}$, but $\bigcap_{n=0}^{\infty} X^n = \emptyset$, whereas $\bigcap_{n=0}^{\infty} X^{n*} = \{\varepsilon\}$.

It may also be observed that each of X^{n*} is a well-ordered set with $\subset$ (inclusion), and hence a finite *ordinal*, say $\alpha, \beta, \gamma, \ldots$, satisfying the following properties: (i) $\beta \in \alpha \Longrightarrow \beta \subset \alpha$ (ii) each α is well-ordered by $\subset$ and (iii) neither α nor its element is an element of itself.

For example, let $X = \{0,1\}$, then $X^{0*} = \{\varepsilon\}$, $X^{1*} = \{\varepsilon, 0,1\}$, $X^{2*} = \{\varepsilon, 0,1,00,01,10,11\}$, $X^{3*} = \{\varepsilon, 0,1,00,01,10,11,000,001,010,011,100,101,110,111\}$, and so on.

Let us recapitulate that the cardinality of a language L, denoted $|L|$, is the number of strings in L. Thus X^* is countably infinite over any X. Moreover, $|X^{n*}| = |X^0| + |X^1| + |X^2| + \cdots + |X^n|$.

Examples

Let $X = \{0\}$. Then,

$$|X^{n*}| = |X^0| + |X^1| + |X^2| + \cdots + |X^n| = |\{\varepsilon\}| + |\{0\}| + |\{00\}| + |\{000\}| + \cdots + |\{0\}^n|$$
$$= 1^0 + 1^1 + 1^2 + \cdots + 1^n.$$

Let $X = \{0,1\}$. Then,

$|X^{n*}| = |X^0| + |X^1| + |X^2| + \cdots + |X^n| = |\{\varepsilon\}| + |\{0,1\}| + |\{00,01,10,11\}| +$
$+ |\{000,001,010,011,100,101,110,111\}| + \cdots + |\{0,1\}^n| = 2^0 + 2^1 + 2^2 + \cdots + 2^n$.

Let $X = \{0,1,2\}$. Then,

$|X^{n*}| = |X^0| + |X^1| + |X^2| + \cdots + |X^n| = |\{\varepsilon\}| + |\{0,1,2\}| +$
$|\{00,01,02,10,11,12,20,21,22\}| +$
$|\{000,001,002,010,011,012,020,021,022,100,101,102,110,111,112,120,121,122,200,201,$
$202,210,211,212,220,221,222\}| + \cdots + |\{0,1,2\}^n| = 3^0 + 3^1 + 3^2 + \cdots + 3^n$.

Let $X = \{0,1,2,3\}$. Then,

$|X^{n*}| = |X^0| + |X^1| + |X^2| + \cdots + |X^n| = |\{\varepsilon\}| + |\{0,1,2,3\}| + |\{00,01,02,03,10,11,12,13,$
$20,21,21,22,23,30,31,32,33\}| + |\{000,001,002,003,010,011,012,013,020,021,022,023,$
$030,031,032,033,100,101,102,103,110,111,112,113,120,121,122,123,130,131,132,133,$
$200,201,202,203,210,211,212,213,220,221,222,223,230,231,232,233,300,301,302,303,$
$310,311,312,313,320,321,322,323,330,331,332,333\}| = 4^0 + 4^1 + 4^2 + \cdots + 4^n$.

By induction, if X be a k −element set, we have

$|X^{n*}| = k^0 + k^1 + k^2 + \cdots + k^n$.

3. Some algebraic structures of languages

3.1 Monoids of equivalence classes of a partition of a language

Let R_a be a relation on X^* such that for $s, t \in X^*$, sR_at if and only if s and t are of equal length.

It is easy to see that R_a is an equivalence relation on X^* and hence, it partitions X^* into its equivalence classes. In other words, a partition of X^* can be viewed as a collection of disjoint languages of X^*, whose union is X^*.

Let the equivalence class generated by $S \in X^*$ be denoted $[S]_{R_a}$ or simply $[S]$, and the quotient set X^*/R_a denote the family of all equivalence classes of X^*.

Let us define an operation $*$ on X^*/R_a such that $[s]*[t] = [st]$ where st is the catenation of s and t. Then, $(X^*/R_a, *, [\varepsilon])$ is a monoid of the partition of X^* induced by R_a, where $[\varepsilon]$

is the identity of catenation. The operation $*$ is neither commutative nor idempotent, in general. However, the identity element $[\varepsilon]$ is the only idempotent element. Also, the operation $*$ is commutative if X is a singleton.

Moreover, as described in section 2 above, it is easy to see that $(L/R_a,\ *,\ [\varepsilon])$ is a commutative monoid where $L = \mathrm{u}^*,\ \mathrm{u} \in \mathrm{X}^*$.

Similarly, for each of the relations $R_b,\ R_c$, and R_d defined on X^* as

(i) sR_bt iff both s and t have the same number of occurrences of each symbol,
(ii) sR_ct iff s and t agree in their first symbols, and
(iii) sR_dt iff s and t agree in their last symbols;

the respective quotient set is a non-commutative and non-idempotent monoid.

Moreover, each of $R_b,\ R_c$, and R_d, similar to R_a, defined on a language u^*, partitions it, and the respective quotient set is a commutative monoid of the partitions of u^*.

3.2 Monoids of partitions of a language

We introduce further three operations on the class of all partitions of X^*.

Let $\mathcal{F}(\mathrm{X}^*)$ denote the collection of all partitions of X^* and $S = \{S_1, S_2, \dots\}$ and $T = \{T_1, T_2, \dots\}$ be two partitions of X^*. Observe that $S_i's$ and $T_i's$ are the blocks of S and T, respectively, and each block is a subset of X^*.

Let a binary operation $\circledast$ be defined on $\mathcal{F}(\mathrm{X}^*)$ as follows:

For any $S, T \in \mathcal{F}(\mathrm{X}^*)$, $S \circledast T$ consists of the set of nonempty intersections of every block of S with every block of T. It is clear that the operation $\circledast$ is both associative and commutative as intersection on languages is associative and commutative. The partition consisting of a unique single block is the identity of $\circledast$. It may be observed that $P \circledast P = P$ for all $P \in \mathcal{F}(\mathrm{X}^*)$ i.e., $\circledast$ is idempotent. Thus, $(\mathcal{F}(\mathrm{X}^*),\ \circledast)$ or $(\mathcal{F}(\mathrm{X}^*),\ \circledast,\ \{\overline{\mathrm{X}^*}\})$ is a commutative, idempotent monoid.

Let another binary operation $\bigoplus$ on $\mathcal{F}(\mathrm{X}^*)$ be defined as follows:

Let $S, T \in \mathcal{F}(\mathrm{X}^*)$. A subset P of X^* belongs to $S \bigoplus T$ if

(i) P is the union of one or more elements of S;
(ii) P is the union of one or more elements of T; and
(iii) No element of P satisfies (i) and (ii) except P itself.

Clearly, $\bigoplus$ is associative and commutative, and the partition consisting of singleton blocks is the identity of the operation $\bigoplus$ on $\mathcal{F}(\mathrm{X}^*)$. Thus, $(\mathcal{F}(\mathrm{X}^*),\ \bigoplus)$ or $(\mathcal{F}(\mathrm{X}^*),\ \bigoplus,\ \{\overline{\mathrm{x}_0}, \overline{\mathrm{x}_1}, \overline{\mathrm{x}_2}, \dots\})$, where $\mathrm{x}_i's$ are the elements of X^*, is a commutative, idempotent monoid.

Finally, let a binary operation $\odot$ be defined on $\mathcal{F}(\mathrm{X}^*)$ as follows:

For any $S, T \in \mathcal{F}(\mathrm{X}^*)$, $S \odot T$ is the union of every block of S with every block of T if no element of the block of S and/or T appears previously. In the case, a block has an element that appeared previously, it is not included in the union.

It is immediate to see that $\odot$ is associative but non-commutative and every element $P \in \mathcal{F}(\mathrm{X}^*)$ is idempotent as $P \odot P = P$ holds. Thus, $(\mathcal{F}(\mathrm{X}^*), \odot)$ is only a semigroup as there is no identity element.

It is immediate to see that all the foregoing constructions, described above, hold good for X^{n*} as well.

Examples

Let $X = \{0,1\}$ be an alphabet and $n = 2$. Then, $\mathrm{X}^{2*} = \{\varepsilon, 0,1,00,01,10,11\}$.

Let $S, T \in \mathcal{F}(\mathrm{X}^{2*})$ where $\mathrm{S} = \{\overline{\varepsilon, 0,1}, \overline{00,01,10,11}\}$ and $T = \{\overline{\varepsilon, 0,1}, \overline{00,01,10}, \overline{11}\}$. Then, the following hold:

(i) $S \circledast T = \{\overline{\varepsilon, 0,1}, \overline{00,01,10}, \overline{11}\} \in \mathcal{F}(\mathrm{X}^{2*})$, $\mathrm{S} \circledast S = \{\overline{\varepsilon, 0,1}, \overline{00,01,10,11}\} = S$, and $S \circledast T = T \circledast S$. Similarly, results could be computed to show associativity. Thus, $(\mathcal{F}(\mathrm{X}^{2*}), \circledast)$ is a commutative, idempotent monoid with $\{\overline{\varepsilon, 0,1,00,01,10,11}\}$ as the identity.

(ii) $S \oplus T = \{\overline{\varepsilon, 0,1}, \overline{00,01,10,11}\} \in \mathcal{F}(\mathrm{X}^{2*})$, $S \oplus S = \{\overline{\varepsilon, 0,1}, \overline{00,01,10,11}\} = S$, and $S \oplus T = T \oplus S$. Moreover, $I = \{\overline{\varepsilon}, \overline{0}, \overline{1}, \overline{00}, \overline{01}, \overline{10}, \overline{11}\}$ is the identity element since $I \oplus T = \{\overline{\varepsilon}, \overline{0}, \overline{1}, \overline{00}, \overline{01}, \overline{10}, \overline{11}\} \oplus \{\overline{\varepsilon, 0,1}, \overline{00,01,10}, \overline{11}\} = \{\overline{\varepsilon, 0,1}, \overline{00,01,10}, \overline{11}\} = T$, for any T. Results could be computed to show that $\oplus$ is associative. Thus, $(\mathcal{F}(\mathrm{X}^{2*}), \oplus)$ is a commutative, idempotent monoid with $\{\overline{\varepsilon}, \overline{0}, \overline{1}, \overline{00}, \overline{01}, \overline{10}, \overline{11}\}$ as the identity.

(iii) $S \odot T = \{\overline{\varepsilon, 0,1}, \overline{00,01,10}, \overline{10}\} \in \mathcal{F}(\mathrm{X}^{2*})$, $T \odot S = \{\overline{\varepsilon, 0,1}, \overline{00,01,10,11}\} \in \mathcal{F}(\mathrm{X}^{2*})$, and $T \odot S \neq S \odot T$. Moreover, $T \odot T = \{\overline{\varepsilon, 0,1}, \overline{00,01,10}, \overline{11}\} \odot \{\overline{\varepsilon, 0,1}, \overline{00,01,10}, \overline{11}\} = \{\overline{\varepsilon, 0,1}, \overline{00,01,10}, \overline{11}\} = T$. In order to show associativity, let $R = \{\overline{\varepsilon, 0,1,00}, \overline{01,10}, \overline{11}\}$. Then, $(S \odot T) \odot R = \{\overline{\varepsilon, 0,1}, \overline{00,01,10}, \overline{10}\} \odot \{\overline{\varepsilon, 0,1,00}, \overline{01,10}, \overline{11}\} = \{\overline{\varepsilon, 0,1,00}, \overline{01,10}, \overline{11}\}$, and $S \odot (T \odot R) = \{\overline{\varepsilon, 0,1}, \overline{00,01,10,11}\} \odot (\{\overline{\varepsilon, 0,1}, \overline{00,01,10}, \overline{11}\} \odot \{\overline{\varepsilon, 0,1,00}, \overline{01,10}, \overline{11}\}) = \{\overline{\varepsilon, 0,1}, \overline{00,01,10,11}\} \odot \{\overline{\varepsilon, 0,1,00}, \overline{01,10}, \overline{11}\} = \{\overline{\varepsilon, 0,1,00}, \overline{01,10}, \overline{11}\}$ i.e., $(S \odot T) \odot R = S \odot (T \odot R)$. Thus, $(\mathcal{F}(\mathrm{X}^{2*}), \odot)$ is an idempotent semigroup.

3.3 Some algebraic structures of a language over $u \in X^*$

Let u^* denote the set of all strings over $u \in X^*$. Then u^* is a commutative monoid under catenation. Moreover, the monoid $C = (u^*, \circ)$ is isomorphic to the monoid $N = (\mathbb{N}, .)$, where $\circ$ and . denote catenation and multiplication, respectively.

Proof

The first part follows by definition.

For the second part, let $f: C \longrightarrow N$ be a function defined as

$$f(u) = \begin{cases} 1, \text{if } u = \varepsilon, \\ n, \text{if } u = u^n, \quad \forall u \in C, \end{cases}$$

where u^n is the $n-$ times catenation of u itself.

It is easy to see that $\forall u, v \in C$, since $f(uv) = f(u_1u_2 \dots u_nv_1v_2 \dots v_n) =$ $f(u_1)f(u_2) \dots f(u_n)f(v_1)f(v_2) \dots f(v_n) = f(u_1u_2 \dots u_n)f(v_1v_2 \dots v_n) = f(u)f(v)$, the function f is a monoid homomorphism.

Let $u, v \in C$ such that $f(u) = f(v)$ i.e., $f(u_1)f(u_2) \dots f(u_n) = f(v_1)f(v_2) \dots f(v_n)$. Then, as strings are ordered, we have $u_1 = v_1, u_2 = v_2, \dots, u_n = v_n$ i.e., $u = v$, which imply that f is injective. Moreover, by the definition of f, $\forall n \in \mathbb{N}$, $\exists u \in C$ such that $f(u) = n$ i.e., f is surjective.

Hence f is an isomorphism.

Proposition 3.3.1

A finite $C = (u^*, \circ)$ is a cyclic group of order n.

Proof

Let u^* be represented as $\{u^0, u^1, \dots, u^{n-1}, \dots\}$. A finite C can be represented as $(u_n^*, \circ)$ where u_n^* is the set of n elements of u^*. Let C be finite viz., $C = \{C^i, \circ\}$, $i = 0,1, \dots, n-1$, where

$$C^i = \begin{cases} C^{i+1}, 0 \leq i < n-1 \\ C^0, i = n-1. \end{cases}$$

Let $C^iC^j = C^{i+j}$, $i + j < n$ and $C^iC^j = C^{i+j-n}$, $i + j \geq n$. Then, it is easy to see that C is a cyclic group of order n.

Example

Let $u = bba$, $u^* = \{\varepsilon, bba, bbabba, \ldots\}$ and $n = 7$. Then, $C = \{C^0, C^1, C^2, C^3, C^4, C^5, C^6\} = \{\varepsilon, bba, bbabba, bbabbabba, bbabbabbabba, bbabbabbabbabba, bbabbabbabbabbabba\}$.

Observe that $C^1C^2 = C^3$, $C^4C^5 = C^2$, $C^6C^1 = C^0$, etc. Thus, C is a cyclic group of order 7.

Proposition 3.3.2

Languages of a finite u^* form a bounded distributive lattice.

Proof

Let a finite u^* be represented as $u^{n\circledast} = \mathrm{u}^0 \cup \mathrm{u}^1 \cup \ldots \cup \mathrm{u}^{\mathrm{n}-1}$, and G be the set of all possible languages of $\mathrm{u}^{n\circledast}$. Let H be a structure consisting of G with union and intersection representing the (join) $\vee$ and (meet) $\wedge$ operations, respectively. Let $L_1, L_2, L_3 \in G$. It is straightforward to see that $L_1 \vee L_2 = L_2 \vee L_1$, $L_1 \vee (L_2 \vee L_3) = (L_1 \vee L_2) \vee L_3$ and $L_1 \vee L_1 = L_1$ hold, as the union of languages is associative, commutative and idempotent. Thus, $(G, \vee)$ is a commutative, idempotent semigroup. Also, as the intersection of languages is commutative, associative and idempotent, $(G, \wedge)$ is a commutative, idempotent semigroup.

Moreover, as the absorption properties hold i.e., $L_1 \vee (L_1 \wedge L_2) = L_1$ and $L_1 \wedge (L_1 \vee L_2) = L_1$, and for all $L_1, L_2 \in G$, $L_1 \wedge L_2 = L_1$ and $L_1 \vee L_2 = L_2$ hold, $H = (G, \vee, \wedge)$ is a lattice.

Also, $\forall L \in G$, as $L \vee \emptyset = L$, $\emptyset$ is the identity element of the join operation and, as $L \wedge G = L$, G is the identity of the meet operation. Thus, H is a bounded lattice.

In addition, as $L_1 \vee (L_2 \wedge L_3) = (L_1 \vee L_2) \wedge (L_1 \vee L_3)$ and $L_1 \wedge (L_2 \vee L_3) = (L_1 \wedge L_2) \vee (L_1 \wedge L_3)$ hold, H is a bounded distributive lattice.

Example

Let $u = 01 \in X^*$ over an alphabet $X = \{0,1\}$, and $u^{3\circledast} = \{\varepsilon, 01, 0101\}$. The set G of all possible languages of $\mathrm{u}^{3\circledast}$ is $\{\emptyset, \{\varepsilon\}, \{01\}, \{0101\}, \{\varepsilon, 01\}, \{\varepsilon, 0101\}, \{01,0101\}, \{\varepsilon, 01, 0101\}\}$.

Observe that $\{01\} \vee (\{01\} \wedge \{01,0101\}) = \{01\}$, $\{01\} \wedge (\{01\} \vee \{01,0101\}) = \{01\}$ i.e., absorption properties hold. Also, $\emptyset \vee \{01,0101\} = \{01,0101\}$ and $\{\varepsilon, 0101\} \wedge \{\varepsilon, 01, 0101\} = \{\varepsilon, 0101\}$ i.e., $\emptyset$ is the identity for $\vee$, and $\{\varepsilon, 01, 0101\}$ is the identity for $\wedge$. Similarly, results for various other combinations could be computed.

Thus, $(G, \vee, \wedge,)$ is a bounded distributive lattice.

4. Concluding Remarks

A number of operations were introduced on the class of partitions of a language which gave rise to certain monoids and semigroups. Moreover, cyclic group, commutative monoid and bounded distributive lattice of a language over a string were introduced. It may be emphasized at this end that the constructions provided in this paper, specially defined on X^{n*}, may be found useful to Network segmentation, analysis of large databases, finite state machines, etc. In particular, an alternative representation of a language, developed in definition 2.2, may be exploited for further research.

5. Acknowledgements

The authors are thankful to the Editor of The Journal of Mathematics and Computer Science for his suggestion to improve upon the references which has been incorporated.

References

[1] B. Ahmadi, C. M. Campbell and H. Doostie, "*Non-commutative finite monoids of a given order* $n \geq 4$ " VERSITA, 22 (2) (2014) 29-35.
[2] J. Gallier, "*Introduction to the Theory of Computation*", Formal Languages and Automata Models of Computation, Lecture Notes, (2010) 1- 60.
[3] E. Hosseinpour, "*T-Rough Fuzzy Subgroups of Groups*", The Journal of Mathematics and Computer Science 12 (3) (2014) 186-195.
[4] J. Kari, "*Automata and Formal Languages*", Lecture Notes, University of Turku, Finland, (2013) 1 – 150.
[5] D.E Knuth, "*The Art of Computer Programming*", Semi-numerical Algorithms, Vol. II, 2nd, Addison-Wesley, (1981).
[6] U. Priss, L. J. Old, "*Conceptual Structures: Inspiration and Application*", Proceedings of the 14th International Conference on Conceptual Structures ICCS, Denmark, 4068 (2006) 388-400.
[7] J.P. Tremblay, R. Manohar, "*Discrete Mathematical Structures with Applications to Computer Science*", Tata McGraw-Hill Edition, (1997).
[8] S.A. N. Zadeh, A. Radfar, "*A. B. Saied, On BP-algebras and QS-algebras*", The Journal of Mathematics and Computer Science, 5 (1) (2012) 17-21.