Journal of Theoretical and Applied Information Technology
30th November 2017. Vol.95. No 22
© 2005 – ongoing JATIT & LLS

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MULTISET CONTROLLED GRAMMARS

[1]SALBIAH ASHAARI, [1]SHERZOD TURAEV, [1]M. IZZUDDIN M. TAMRIN
[2]ABDURAHIM OKHUNOV, [3]TAMARA ZHUKABAYEVA

[1]Kulliyyah of Information and Communication Technology, International Islamic University Malaysia

53100 Kuala Lumpur, Malaysia

[2]Kulliyyah of Engineering, International Islamic University Malaysia

53100 Kuala Lumpur, Malaysia

[3]Faculty of Information Technology, L.N. Gumilyov Eurasian National University

010008 Astana, Kazakhstan

E-mail: [1]salbiah.ash@gmail.com, [1]sherzod@iium.edu.my, [1]izzuddin@iium.edu.my
[2]abdurahimokhun@iium.edu.my, [3]tamara_kokenovna@mail.ru

ABSTRACT

This study focusses on defining a new variant of regulated grammars called *multiset controlled grammars* as well as investigating their computational power. In general, a multiset controlled grammar is a grammar equipped with an arithmetic expression over multisets terminals where to every production in the grammar a multiset is assigned, which represents the number of the occurrences of terminals on the right-hand side of the production. Then a derivation in the grammar is said to be successful if only if its multiset value satisfies a certain relational condition. In the study, we have found that control by multisets is powerful tool and yet a simple method in regulation of generative processes in grammars. We have shown that multiset controlled grammars are at least as powerful as additive valence grammars, and they are at most powerful as matrix grammars.

Keywords: *Multisets, Context-free Grammars, Regulated Grammars, Generative Capacity.*

1. INTRODUCTION

Grammars are language generation models that define their language strings so their process of rewriting will generate them starting from a special start symbol. Basically, they can be classified into two fundamental categories which are context-free grammars and non-context-free grammars. Among those two, context-free grammars are the most developed and well examined grammar class in Chomsky hierarchy since they have a lot of good sides in terms of computational properties and complexity problems. Additionally, they are also undoubtedly easy to be implemented in many formal languages applications. However, unfortunately, context-free grammars are not able to cover all aspects occurred in modeling of phenomena and it is also already well-known that many real-world problems cannot be described accurately by context-free languages which have been discussed in [1].

Thus, we need to go beyond context-free grammars where one of the solutions is to consider the context-sensitive grammars which are more powerful. Nevertheless, in spite of their great power, they have some serious problems in the practical usage, where they have several adverse features regarding decidability problems in which whether they are undecidable or having exponential algorithms. Furthermore, it is hard or impossible to describe the derivations of context sensitive grammars by a graph or tree structure which is an essential tool in analyzing the structure of the problems [1,2]. These are the reasons why many researchers are looking for intermediate grammars between context-free and context-sensitive grammars, called regulated or controlled grammars, where they can combine the beauty and simplicity of context-free, at the same time possess the power of context-sensitive grammars.

A regulated or controlled grammar is portrayed as a grammar with some additional mechanisms where the applications of certain rules are being restricted in order to avoid certain derivations. In [1-10], we can find a large number of old and new as well as well-known of various types of regulated grammar that preserve the nature of context-free such as matrix grammars, regularly controlled grammars, vector grammars, random context grammars, tree controlled grammars, semi-conditional grammars, global indexed grammars, Petri net controlled grammars, string-regulated graph grammars, Parikh vector controlled grammars and many more. All of those grammars have achieved plentiful remarkable results within formal language theory and are different from each other depending on their restrictions either based upon the variety of context related or on the use of rules during the process of generating the languages. However, under certain circumstances they are too complicated or not computationally complete or correlate to a group of grammars with too many unsolvable decision problems which have lessen the practical interest. Therefore, here we introduce a new controlled grammar called multiset controlled grammar where its restriction of rules is based on terminal multisets.

Multiset was defined by [11] is a collection of unordered objects called elements in which it is allowed to have repeated occurrences of identical elements. It is important to consider the term multiset since there exist circumstances such there are repeated hydrogen and oxygen atoms in a sulfuric acid molecule (H_2SO_4), repeated roots of polynomial equations, repeated observations in statistical samples and so on where those repeated elements need to be counted in order to attain their definiteness and adequacy [9]. On account to its aptness, it has been used interchangeably with a variety of term which carrying synonymy with multiset even in different contexts like heap, bag, occurrence set, fireset, list, weighted set, sample and bunch [13,14].

The notion of relating multiset rewriting with Chomsky grammars was initiated by Kudlek, Martin and Paun in 2001 with the name of "multiset grammars" where they considered the applicability of rules as multiset in restricting the use of productions of grammars [14]. Additionally, the lower and upper bound of computational power as well as the closure properties of multiset grammars have been investigated in [14,15]. In both papers, it was found that multiset grammars are more powerful than context-free grammars and have at most computational power of matrix grammars.

Interestingly, in the same year where multiset grammar was discovered, authors' in [16] immediately came up with a Chomsky hierarchy characterization of multiset grammars in term of multiset automata where their working mode is according to the addition and subtraction of multiset. In that paper [16], they defined three types of multiset automata known as multiset finite automata (MFA), multiset linear bounded automata (MLBA) and multiset Turing machine (MTM). It was shown that MFAs are powerful as multiset regular and context free grammars, Parikh set of regular and context free grammars as well as semilinear grammars while MLBAs are powerful as monotone grammars and MTMs are powerful as arbitrary grammars. Besides, it was proven that the deterministic variants of all of those automata are firmly less powerful than the non-deterministic variants unlike the grammar cases [16].

Since then, after a while, the direction of multiset study are continuing expanded where in 2007, the researchers' in [17] developed a new grammar model known as random context multiset grammars which based on relation of partial order on the objects the grammars contend with together with the multiset random context checkers and transducers concept. In that study, they showed how those grammars can generate set of recursively enumerable of finite multiset and also can be easily enhanced to antiport P system [17].

Further, the authors' in [18] made use of fuzzy concept to introduce two new extensions of multiset grammars and automata called fuzzy multiset grammars and fuzzy multiset finite automata with discussion of the relationship between fuzzy multiset regular grammars with fuzzy multiset finite automata in 2013. They demonstrated that if a fuzzy multiset language is accepted by a fuzzy multiset finite automaton, it can also be generated by fuzzy multiset regular grammar and vice versa [18]. In addition, they investigated closure properties of fuzzy multiset finite languages family under certain regular operations such union, addition and Kleene star [18]. Shortly thereafter, in 2015, the researchers' in [19] widened the study done by [18] by associating a deterministic fuzzy multiset finite automaton with a given fuzzy multiset finite automaton and showing that both automata are equally powerful in the sense of fuzzy multiset language acceptance. They as well studied and presented two minimal realizations of fuzzy multiset language where they proved that both of them are isomorphic [19].

In 2016, Sharma et al. conferred an in-depth study of fuzzy multiset grammars where they introduced two new variant of grammars called fuzzy multiset left linear grammars and fuzzy multiset right linear grammars. They proved that both grammars along with fuzzy multiset grammars and fuzzy multiset regular grammars are equivalent to each other in normal formal except for the case of empty string. They also showed that all languages of fuzzy multiset automaton class is closed under homomorphism, inverse homomorphism, right quotient by any multiset, quotient with arbitrary multisets and reversal of a fuzzy multiset finite language [22].

This paper is outlined as follows. First, we recall some well-known basic notations, terminologies and concepts related to the formal languages theories, multisets and regulated rewriting grammars which will be used throughout this paper. Then, we introduce a new type of controlled grammars called *multiset controlled grammar*s. Further, we investigate their generative power and closure properties. Lastly, we give a brief summarization of all materials discussed in this paper together with some open problems regarding this multiset controlled grammar.

2. PRELIMINARIES

In this section, we only recall some well-known basic notations, terminologies and concepts related to the formal languages theories, multiset and regulated rewriting grammars that will be used in the next sections. For more details information, the reader can refer to [1,2,20-21].

2.1 General Notations

Throughout the paper, we use the following notations. The symbols $\in$ and $\notin$ are used to represent the set membership and negation of set membership of an element to a set, respectively. The symbol $\subseteq$ signifies the set inclusion which is not necessarily proper, and the symbol $\subset$ marks the strict inclusion. $|A|$ portrays the cardinality of a set A, which is the number of elements in A, and 2^A depicts the power set of a set A. The symbol $\emptyset$ denotes the empty set which is the set without elements. The sets of integer, natural, real and rational numbers are denoted by $\mathbb{Z}, \mathbb{N}, \mathbb{R}$ and $\mathbb{Q}$, respectively. An alphabet Σ is a finite and nonempty set of elements known as symbols or letters, and a string (sometimes referred as a word) over the alphabet Σ is a finite sequence of symbols of Σ. The string without symbols is called the null or empty string, and it is denoted by λ. The set of all strings (including λ) over the alphabet, Σ is represented by Σ^*, and the set of all strings except the empty string is denoted by Σ^+, i.e., $\Sigma^+ = \Sigma^* - \{\lambda\}$. A language L is a subset of Σ^*. For a set A, a mapping $\mu : A \to \mathbb{N}$ is called a multiset. The set of all multisets over A is denoted by $A^{\oplus}$. Then, the set A is a called the basic set of $A^{\oplus}$. For a multiset $\mu \in A^{\oplus}$ and element $a \in A$, $\mu(a)$ represents the number of a's in μ.

2.2 Grammars

A phrase structure grammar is a quadruple $G = (N, T, S, P)$ where N is an alphabet of nonterminals, T is an alphabet of terminals and $N \cap T = \emptyset$, $S \in N$ is the start symbol, P is a finite set of production of the form $A \to w$ where $A \in (N \cup T)^* N (N \cup T)^*$, $w \in (N \cup T)^*$. If a production is in form of $A \to \lambda$, then it is called an erasing rule.

For a grammar $G = (N, T, S, P)$, a direct derivation relation over $(N \cup T)^*$ which denoted by $\Rightarrow$ and defined as $u \Rightarrow v$ provided if and only if there is a rule $A \to w \in P$ such that $u = x_1 A x_2$ and $v = x_1 w x_2$ for $x_1, x_2 \in (N \cup T)^*$. Since $\Rightarrow$ is a relation, then its nth, $n \geq 0$, power is denoted by $\Rightarrow^n$, its transitive closure by $\Rightarrow^+$, and its reflexive and transitive closure by $\Rightarrow^*$. A string $w \in (N \cup T)^*$ is a sentential form if $S \Rightarrow^* w$. If $w \in T^*$, then w is called a sentence or a terminal string and $S \Rightarrow^* w$ is said to be a successful derivation. We also use the notations $\overset{m}{\Rightarrow}$ or $\xRightarrow{r_0 r_1 \ldots r_n}$ to denote the derivation that use the sequence of rules $m = r_0 r_1 \ldots r_n$, $r_i \in P, 1 \leq i \leq n$.

The language generated by G, denoted by $L(G)$, is defined as $L(G) = \{w \in T^* \mid S \Rightarrow^* w\}$. Two grammars G_1 and G_2 are called to be equivalent if and only if they generate the same language, i.e., $L(G_1) = L(G_2)$.

The Chomsky hierarchy classifies all grammars into four basic categories according to the form of production rules, i.e., a grammar $G = (N, T, S, P)$ is called

- unrestricted or recursively enumerable grammar (type-0) if its productions are in the form of $u \to v$ where $u \in (N \cup T)^+$, $v \in (N \cup T)^*$ and u contains at least one nonterminal symbol.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- context sensitive grammar (type-1) if its productions are in the form of $u \to v$ where $|u| \le |v|$, $u \in (N \cup T)^* N^+ (N \cup T)^*$ and $v \in (N \cup T)^+$.
- context free grammar (type-2) if its productions are in the form of $A \to w$ where $A \in N$ and $w \in (N \cup T)^*$.
- linear grammar if its productions are in the form of $A \to w$ where $A \in N$ and $w \in T^* \cup T^*NT^*$.
- regular grammar (type-3) if its productions are in the form of $A \to w$ where $w \in T^* \cup NT^*$ and $A \in N$.

The families of languages generated by arbitrary, unrestricted, context sensitive, context free, regular, linear and finite grammars are denoted by **RE**, **CS**, **CF**, **REG**, **LIN**, and **FIN**, respectively. For these language families, Chomsky hierarchy holds:

$$\mathbf{FIN} \subset \mathbf{REG} \subset \mathbf{LIN} \subset \mathbf{CF} \subset \mathbf{CS} \subset \mathbf{RE}.$$

2.3 Grammars with Regulated Rewriting

We recall some definitions of controlled grammars that will be used throughout the proposal.

A matrix grammar is a quadruple $G = (N, T, S, M)$ where N, T and S are defined as for context-free grammar and M is a set of matrices, that are finite sequences of context-free rules from $N \times (N \cup T)^*$. The language generated by G is defined by $L(G) = \{w \in T^* | S \overset{\pi}{\Rightarrow} w \text{ and } \pi \in M^*\}$.

An additive valence grammar is a 5-tuples $G = (N, T, S, P, v)$ where N, T, S, P are defined as for a context-free grammar and v is a mapping from P into $\mathbb{Z}$ ($\mathbb{Q}$). The language generated by the additive (multiplicative) grammar G consists of all string $w \in T^*$ such that there is a derivation $S \xRightarrow{r_1 r_2 \dots r_n} w$ where

$$\sum_{k=1}^{n} v(r_k) = 0 .$$

The families of languages generated by matrix and additive valence grammars (with erasing rues) are denoted by **MAT**, a**VAL**, (**MAT**$^\lambda$, a**VAL**$^\lambda$), respectively.

3. DEFINITIONS AND EXAMPLES

In this section, we give a meticulous definition of multiset controlled grammar, a derivation step and a successful derivation with some examples of language of multiset controlled grammar.

Definition 1. A multiset controlled grammar is a 6-tuples $G = (N, T, S, P, \oplus, F)$ where N, T and S are defined as for a context-free grammar, P is a finite subset of $N \times (N \cup T)^* \times T^{\oplus}$ and $F: T^{\oplus} \to \mathbb{Z}$ is a linear or nonlinear function. A triple $(A, w, \omega) \in P$ is written as $A \to w[\omega]$.

If $F(a_1, a_2, \dots, a_n), a_i \in T, 1 \le i \le n$ is a linear, then it is in the form of $F(a_1, a_2, \dots, a_n) = \sum_{i=1}^{n} c_i \mu(a_i) + c_0$ where $c_i \in \mathbb{Z}, 0 \le i \le n$. Then, as a nonlinear function F, we can consider logarithms, polynomials, rational, exponential, power and so on.

Definition 2. A string-weight pair $(u, \omega) \in (N \cup T)^* \times T^{\oplus}$ is called a sentential form, written as $u[\omega]$. Then, we say that $u \in (N \cup T)^* \times T^{\oplus}$ directly derives $v \in (N \cup T)^* \times T^{\oplus}$ in G if and only if $u = x_1 A x_2[\omega_1]$ and $v = x_1 w x_2[\omega_2]$ for some $x_1, x_2 \in (N \cup T)^*$ written as $u[\omega_1] \overset{r}{\Rightarrow} v[\omega_2]$ where $r : A \to w[\omega] \in P$ with $\omega_2 = \omega_1 \oplus \omega$.

Definition 3. A derivation in G such $S[\omega_0] \overset{r_1}{\Rightarrow} w_1[\omega_1] \overset{r_2}{\Rightarrow} w_2[\omega_2] \overset{r_3}{\Rightarrow} \cdots \overset{r_n}{\Rightarrow} w_n[\omega_n]$, $n \ge 1$ is called successful if and only if $\omega_n \in T^*$ and $T^{\oplus}$. For short, it can be written as $S \overset{\pi}{\Rightarrow} w_n[\omega_n]$ where $\pi = r_1 r_2 \cdots r_n$.

Definition 4. The language of the grammar G, denoted by $L(G)$, consists of all strings $w[\omega]$ obtained by successful derivations in G, i.e., $L(G) = \{w[\omega] \in T^* \times T^{\oplus} \mid S \overset{\pi}{\Rightarrow} w_n[\omega_n]$ where $\pi = r_1 r_2 \cdots r_n\}$.

Definition 5. The language $L(G, \alpha, *) = \{w \mid w[\omega] \in L(G),\ F(\omega) * \alpha\}$ where the relation $* \in \{=, <, >, \le, \ge\}$ and $\alpha \in W, W \subseteq \mathbb{Z}$ is a cut point set called a threshold language of the grammar under G and α.

The families of threshold of languages generated by multiset controlled regular, linear and context-free grammars with and without erasing rules are denoted by m**REG**, m**LIN**, m**CF** (m**REG**$^\lambda$, m**LIN**$^\lambda$, m**CF**$^\lambda$) respectively. We also use bracket notation $m\mathbf{X}^{[\lambda]}, \mathbf{X} \in \{\mathbf{REG}, \mathbf{LIN}, \mathbf{CF}\}$ to show that a statement holds both cases of with and without erasing rules.

Journal of Theoretical and Applied Information Technology
30th November 2017. Vol.95. No 22
© 2005 – ongoing JATIT & LLS
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Further, to demonstrate the derivation process of multiset controlled grammars, we provide two examples using linear and non-linear functions F.

Example 1 (Linear Function). Let $G_1 = (\{A, B, S\}, \{a, b, c\}, S, P, \oplus, F)$ be a multiset controlled grammar where P consists of the following productions

$r_0 : S \to AB[(0,0,0)]$,
$r_1 : A \to aAb[(1,1,0)]$,
$r_2 : A \to ab[(1,1,0)]$,
$r_3 : B \to bBc[(0,1,1)]$,
$r_4 : B \to bc[(0,1,1)]$ and
$F(a, b, c) = \mu(a) - \mu(c)$

where we consider the threshold language $L(G_1, 0, =)$ of grammar.

We start with the only applicable production rule r_0 which yields AB. Then we can either

- terminate the derivation by applying productions r_2r_4 to obtain $abbc$ or
- rewrite AB to $aAbbBc$ by applying productions r_1r_3.

Then, the derivation can be continued from A and B applying productions r_1 and r_3 any number of times. To terminate the derivation, productions r_2 and r_4 should be applied. In general, we can have the derivation as follows:

$$S \overset{r_0}{\Rightarrow} AB[(0,0,0)] \overset{r_1^*}{\Rightarrow} a^nAb^nB[(n,n,0)]$$
$$\overset{r_3^*}{\Rightarrow} a^nAb^nb^nBc^n[(n,2n,n)]$$
$$\overset{r_2}{\Rightarrow} a^{n+1}b^{n+1}b^nBc^n[(n+1,2n+1,n)]$$
$$\overset{r_4}{\Rightarrow} a^{n+1}b^{n+1}b^{n+1}c^{n+1}[(n+1,2n+2,n+1)].$$

Then, $(a,b,c) = (n+1) + (2n+2) + (-1)(2n+2) + (-1)(n+1) = 0$. For instance, the string $a^2b^4c^2$ can be obtained by the following derivation:

$$S[(0,0,0)] \overset{r_0}{\Rightarrow} AB[(0,0,0)] \overset{r_1}{\Rightarrow} aAbB[(1,1,0)]$$
$$\overset{r_2}{\Rightarrow} aabbB[(2,2,0)] \overset{r_3}{\Rightarrow} aabbbBc[(2,3,1)]$$
$$\overset{r_4}{\Rightarrow} aabbbbcc[(2,4,2)] = a^2b^4c^2[(2,4,2)]$$ and
$$F(a,b,c) = 2 + 4 + (-1)(4) + (-1)(2) = 0.$$

Obviously, this grammar generates a non-context free language:

$$L(G_1, 0, =) = \{a^nb^{2n}c^n \mid n \geq 1\} \in \mathbf{CS} - \mathbf{CF}.$$

Example 2 (Non-Linear Function). Let $G_2 = (\{A, S\}, \{a\}, S, P, \oplus, F)$ be a multiset controlled grammar where P consists of the following productions

$r_0 : S \to aA[(1)]$,
$r_1 : A \to aA[(1)]$,
$r_2 : A \to a[(1)]$ and
$F(a) = \log_2 \mu(a)$.

where we consider the threshold language $L(G_2, \alpha, =)$, $\alpha \in \mathbb{Z}^+$, of the grammar G_2.

We start with the only applicable production rule r_0 which yields aA. Then we can either

- terminate the derivation by applying productions r_2 to obtain a or
- rewrite A to aaA by applying productions r_1.

Then, the derivation can be continued from A production r_1 any number of times. To terminate the derivation, production r_2 should be applied. In general, we can see the derivation as

$$S \overset{r_0}{\Rightarrow} aA[(1)] \overset{r_1^*}{\Rightarrow} a^nA[(n)] \overset{r_2}{\Rightarrow} a^{n+1}[(n+1)].$$

Then, $F(a) = \log_2(n+1), n = 2^n - 1$, $n \geq 1$. For instance, the string $aaaa$ can be obtained by the following derivation:

$$S[(1)] \overset{r_0}{\Rightarrow} aA[(1)] \overset{r_1}{\Rightarrow} aaA[(2)] \overset{r_1}{\Rightarrow} aaaA[(3)]$$
$$\overset{r_2}{\Rightarrow} aaaa[(4)] = a^4[(4)]$$ and $F(a) = \log_2 4$.

Clearly, the grammar generates a non-context free language:

$$L(G_2, \alpha, =) = \{a^{2^n} : n \geq 1\} \in \mathbf{CS} - \mathbf{CF}.$$

4. GENERATIVE POWERS

In this section, we establish results concerning the lower and upper bounds of multiset controlled grammar defined above.

From the definitions, the next three lemmas follows immediately.

Lemma 1. For $L' = L(G, \alpha, =) \in m\mathbf{CF}$, $G = (N, T, S, P, \oplus, F)$, $\alpha \neq 0$, there exists $\exists L'' \in m\mathbf{CF}$ such that $L'' = (G', 0, =)$ with $G' = (N, T, S, P, \oplus, F')$.

Lemma 2. $m\mathbf{REG}^{[\lambda]} \subseteq m\mathbf{LIN}^{[\lambda]} \subseteq m\mathbf{CF}^{[\lambda]}$.

Lemma 3. For $X \in \{\mathbf{REG}, \mathbf{LIN}, \mathbf{CF}\}$, $\mathbf{X}^{[\lambda]} \subseteq m\mathbf{X}^{[\lambda]}$.
Proof: Let $G = (N, T, S, P)$ be a regular, linear and context-free grammar where $T = \{a_1, a_2, \dots, a_n\}$. We construct the multiset counterpart of G as $G' = (N, T, S, P', \oplus, F)$ where each $A \to w \in P$ is

replaced with $A \to w[\mathbf{0}], \mathbf{0} \in T^{\oplus}$, in P' and function F is taken as $F(a_1, a_2, \dots, a_n) = \sum_{i=0}^{n} \mu(a_i)$. Then, we define the threshold language as $L(G', 0, =)$. Thus, it is easy to see that $L(G) = L(G', 0, =)$.

Example 3 Let $G_3 = (\{A, B, S\}, \{a, b\}, S, P, \oplus, F)$ be a multiset controlled regular grammar where P consists of the following productions

$r_0 : S \to aA[(1,0)]$,
$r_1 : A \to aA[(1,0)]$,
$r_2 : A \to bB[(0,1)]$,
$r_3 : A \to b[(0,1)]$,
$r_4 : B \to bB[(0,1)]$,
$r_5 : B \to b[(0,1)]$ and
$F(a, b) = \mu(a) + (-1)\mu(b)$.

In general, we have the derivation:

$S \overset{r_0}{\Rightarrow} aA[(1,0)] \overset{r_1^*}{\Rightarrow} a^n A[(n, 0)]$
$\overset{r_2}{\Rightarrow} a^{n+1} A[(n + 1,0)] \overset{r_4}{\Rightarrow} a^{n+1} Ab[(n + 1,1)]$
$\overset{r_4^*}{\Rightarrow} a^{n+1} Ab^n[(n + 1, n)]$
$\overset{r_5}{\Rightarrow} a^{n+1} b^{n+1}[(n + 1, n + 1)]$, and
$F(a, b) = (n + 1) + (-1)(n + 1) = 0$.

Hence, G_3 generates the language
$L(G_3, 0, =) = \{ a^n b^n \mid n \geq 1 \} \in \mathbf{CF} \cap m\mathbf{REG} - \mathbf{REG}$.

From the Lemma 3 and Example 3, it follows:

Theorem 1. REG $\subset m$REG.

Example 4. Let $G_4 = (\{A, B, S\}, \{a, b, c, d\}, S, P, \oplus, F)$ be a multiset controlled context-free grammar where P consists of the following productions

$r_0 : S \to AB[(0,0,0,0)]$,
$r_1 : A \to aAb[(1,1,0,0)]$,
$r_2 : A \to ab[(1,1,0,0)]$,
$r_3 : B \to cBd[(0,0,1,1)]$,
$r_4 : B \to cd[(0,0,1,1)]$ and
$F(a, b, c) = \mu(a) + \mu(b) + (-1)\mu(c) + (-1)\mu(d)$.

Then, we have the following derivation:

$S \overset{r_0}{\Rightarrow} AB[(0,0,0,0)] \overset{r_1^*}{\Rightarrow} a^n Ab^n B[(n, n, 0,0)]$
$\overset{r_3^*}{\Rightarrow} a^n Ab^n c^n Bd^n[(n, n, n, n,)]$
$\overset{r_2}{\Rightarrow} a^{n+1} b^{n+1} c^n Bd^n[(n + 1, n + 1, n, n)]$
$\overset{r_4}{\Rightarrow} a^{n+1} b^{n+1} c^{n+1} d^{n+1}[(n + 1, n + 1, n + 1, n + 1)]$ and
$F(a, b, c, d) = (n + 1) + (n + 1) + (-1)(n + 1) + (-1)(n + 1) = 0$.

Thus, G_4 generates the language
$L(G_4, 0, =) = \{ a^n b^n c^n d^n \mid n \geq 1 \} \in \mathbf{CS} \cap m\mathbf{CF} - \mathbf{CF}$.

From Lemma 3 and Example 4, it follows

Theorem 2. CF $\subset m$CF.

Example 5 Let $G_5 = (\{A, B, S\}, \{a, b, c\}, S, P, \oplus, F)$ be a linear grammar where P consists of the following productions:

$r_0 : S \to aAc[(1,0,1)]$,
$r_1 : A \to aAc[(1,0,1)]$,
$r_2 : A \to bB[(0,1,0)]$,
$r_3 : A \to b[(0,1,0)]$,
$r_4 : B \to bB[(0,1,0)]$,
$r_5 : B \to b[(0,1,0)]$ and
$F(a, b, c) = \mu(a) + \mu(b) + (-1)\mu(b) + (-1)\mu(c)$.

It is not difficult to see that

$S \overset{r_0}{\Rightarrow} aAc[(1,0,1)]$
$\overset{r_1^*}{\Rightarrow} a^{n+1} Ac^{n+1}[(n + 1,0, n + 1)]$
$\overset{r_2}{\Rightarrow} a^{n+1} bBc^{n+1}[(n + 1, 1, n + 1)]$
$\overset{r_4^*}{\Rightarrow} a^{n+1} b^n Bc^{n+1}[(n + 1, n, n + 1)]$
$\overset{r_5}{\Rightarrow} a^{n+1} b^{n+1} c^{n+1}[(n + 1, n + 1, n + 1)]$ and
$F(a, b, c) = (n + 1) + (n + 1) + (-1)(n + 1) + (-1)(n + 1) = 0$.

Therefore, G_5 generates the language
$L(G_5, \{0\}, =) = \{ a^n b^n c^n \mid n \geq 1 \} \in \mathbf{CS} \cap m\mathbf{LIN} - \mathbf{LIN}$.

From Lemma 3 and Example 5, it follows

Theorem 3. LIN $\subset m$LIN.

Example 6. Let $G_6 = (\{A, B, C, S\}, \{a, b, c\}, S, P, \oplus, F)$ be a multiset controlled regular grammar where P consists of the following productions:

$r_0 : S \to aA[(1,0,0)]$,
$r_1 : A \to aA[(1,0,0)]$,
$r_2 : A \to bB[(0,1,0)]$,
$r_3 : B \to bB[(0,1,0)]$,
$r_4 : B \to c[(0,0,1)]$,
$r_5 : B \to cC[(0,0,1)]$,
$r_6 : C \to cC[(0,0,1)]$,
$r_7 : C \to c[(0,0,1)]$ and
$F(a, b, c) = \mu(a) + \mu(b) + (-1)\mu(b) + (-1)\mu(c)$.

Generally, we will have the derivation:

$S \overset{r_0}{\Rightarrow} aA[(1,0,0)] \overset{r_1^*}{\Rightarrow} a^{n+1} A[(n + 1,0,0)]$

$\stackrel{r_2}{\Rightarrow} a^{n+1}bB[(n+1,1,0)]$
$\stackrel{r_3^*}{\Rightarrow} a^{n+1}b^{n+1}B[(n+1,n+1,0)]$
$\stackrel{r_5}{\Rightarrow} a^{n+1}b^{n+1}cC[(n+1,n+1,1)]$
$\stackrel{r_6^*}{\Rightarrow} a^{n+1}b^{n+1}c^nC[(n+1,n+1,n)]$
$\stackrel{r_7}{\Rightarrow} a^{n+1}b^{n+1}c^{n+1}[(n+1,n+1,n+1)]$ and
$F(a,b,c) = (n+1) + (n+1) + (-1)(n+1) + (-1)(n+1) = \mathbf{0}$.

Hence, G_6 generates the language $L(G_6,\{0\},=) = \{a^nb^nc^n \mid n \geq 1\} \in \mathbf{CS} \cap m\mathbf{REG} - \mathbf{CF}$.

From the example above, it follows

Theorem 4. $m\mathbf{REG} - \mathbf{CF} \neq \emptyset$.

In Example 2.1.7 in (1), the language $L = \{a^nb^nc^n \mid n \geq 1\}^2$ has been proven cannot be generated by any additive valence grammar. However, this language can be generated by multiset controlled grammar as in example 7.

Example 7. $L(G_7,0,=) = \{a^nb^nc^n \mid n \geq 1\}^2 \in m\mathbf{CF} - a\mathbf{VAL}$.

The grammar for the language $L(G_7,0,=)$:
$r_0 : S \to ABCD[(0,0,0)]$,
$r_1 : A \to aaAb[(2,1,0)]$,
$r_2 : B \to bBcc[(0,1,2)]$,
$r_3 : C \to aaCb[(2,1,0)]$,
$r_4 : D \to bDcc[(0,1,2)]$,
$r_5 : A \to ab[(1,1,0)]$,
$r_6 : B \to c[(0,0,1)]$,
$r_7 : C \to ab[(1,1,0)]$,
$r_8 : D \to c[(0,0,1)]$,
$r_9 : A \to \lambda[(0,0,0)]$,
$r_{10} : B \to \lambda[(0,0,0)]$,
$r_{11} : C \to \lambda[(0,0,0)]$,
$r_{12} : D \to \lambda[(0,0,0)]$ and
$F(a,b,c) = \mu(a) + \mu(b) + (-1)\mu(b) + (-1)\mu(c)$.

Here, it is clearly that we can have the derivation such

$S \stackrel{r_0}{\Rightarrow} ABCD[(0,0,0)]$
$\stackrel{r_1}{\Rightarrow} aaAbBCD[(2,1,0)]$
$\stackrel{r_1^*}{\Rightarrow} a^nAb^{n/2}BCD\left[\left(n,\frac{n}{2},0\right)\right]$
$\stackrel{r_2}{\Rightarrow} a^nAb^{\frac{n}{2}}bBccCD\left[\left(n,\frac{n}{2}+1,2\right)\right]$
$\stackrel{r_2^*}{\Rightarrow} a^nAb^nBc^nCD[(n,n,n)]$
$\stackrel{r_3}{\Rightarrow} a^nAb^nBc^naaCbD[(n+2,n+1,n)]$
$\stackrel{r_3^*}{\Rightarrow} a^nAb^nBc^na^mCb^{m/2}D[(n+m,n+m/2,n)]$
$\stackrel{r_4}{\Rightarrow} a^nAb^nBc^na^mCb^{\frac{m}{2}}bDcc \left[\left(n+m,n+\frac{m}{2}+1,n+2\right)\right]$
$\stackrel{r_4^*}{\Rightarrow} a^nAb^nBc^na^mCb^mDc^m [(n+m,n+m,n+m)]$
$\stackrel{r_5r_6r_7r_8}{\Longrightarrow} a^{n+1}b^{n+1}c^{n+1}a^{m+1}b^{m+1}c^{m+1} [(n+m+1,n+m+1,n+m+1)]$ or
$\stackrel{r_9r_{10}r_{11}r_{12}}{\Longrightarrow} a^nb^nc^na^mb^mc^m [(n+m,n+m,n+m)]$ and
$F(a,b,c) = (n+m) + (n+m) + (-1)(n+m) + (-1)(n+m) = \mathbf{0}$.

Thus, G_7 generates the language $L(G_7,\mathbf{0},=) = \{a^nb^nc^n \mid n \geq 1\}^2$.

From Example 7, it follows

Theorem 5. $m\mathbf{CF} - a\mathbf{VAL} \neq \emptyset$.

Further, we discuss the upper bound for multiset controlled grammars.

Theorem 6. $m\mathbf{CF}^{[\lambda]} \subseteq \mathbf{MAT}^{[\lambda]}$

Proof: Let $G = (N,T,S,P,\bigoplus,F)$ be a multiset controlled context-free grammar where F is a linear function. Let $T = \{a_1,a_2,\dots,a_n\}$ and $F(a_1,a_2,\dots,a_n) = \sum_{i=1}^n c_i\,\mu(a_i)$. Since F is linear, without loss of generality, it can be represented as the difference of the sum of positive terms and the sum of negative terms, i.e.,

$$F(a_1,a_2,\dots,a_n) = \sum_{c_{i_j}>0} c_{i_j}\,\mu\left(a_{i_j}\right) + \sum_{c_{i_k}<0} c_{i_k}\,\mu\left(a_{i_k}\right).$$

Then, let $L' = L(G,0,=)$. We construct an equivalent matrix grammar $G' = (N',T,S',M')$ where $N' = N \cup \{S',X,Y,Z\}$ where S',X,Y,Z are new nonterminals. We introduce the start matrix

$$m_0 : (S' \to SZ) \quad (1)$$

and define the matrix m_r for each production $r = A \to w[\omega] \in P$ as

$$m_r : (A \longrightarrow w, [Z \to X^{\sum_{c_{i_j}>0} c_{i_j}\omega\left(a_{i_j}\right)}Y^{\sum_{c_{i_k}<0} c_{i_k}\omega\left(a_{i_k}\right)}Z]) \quad (2)$$

We also consider the erasing matrices

$$m_{\lambda,Z} : (Z \to \lambda), \quad (3)$$
$$m_\lambda : (X \to \lambda, Y \to \lambda) \quad (4)$$

The matrix set M' consists of the matrices (1) – (4) defined above. Further, we show that $L(G,0,=) = L(G')$.

(i) $L(G,0,=) \subseteq L(G')$

Let $D: S \overset{r_1}{\Rightarrow} w_1[\omega_1] \overset{r_2}{\Rightarrow} w_2[\omega_2] \Rightarrow \cdots \ldots \overset{r_t}{\Rightarrow} w_t[\omega_t],\ w_t \in T^*$ be a successful derivation, i.e.,

$F(a_1, a_2, \ldots, a_n) = \sum_{i=1}^{n} c_i\, \mu(a_i) = 0$ where $\mu(a_i) = \sum_{i=1}^{t} \omega(a_i)$.

We construct a derivation D' in G' simulating D. D' starts with matrix (1) and for each r_i in D, we choose m_{r_i} in D', i.e.,

$$S' \overset{m_0}{\Rightarrow} SZ \overset{m_{r_1}}{\Longrightarrow} w_1 X^{\sum_{c_{i_j}>0} c_{i_j}\omega(a_{i_j})} Y^{\sum_{c_{i_k}<0} c_{i_k}\omega(a_{i_k})} Z \overset{m_{r_2}}{\Longrightarrow} w_2 X^{\sum_{c_{i_j}>0} c_{i_j}(\omega_1(a_{i_j})+\omega_2(a_{i_j}))} Y^{\sum_{c_{i_k}<0} c_{i_k}(\omega_1(a_{i_k})+\omega_2(a_{i_k}))} Z \Rightarrow \ldots \ldots \overset{m_{r_t}}{\Longrightarrow} w_t X^A Y^B Z$$

where

$$A = \sum_{s=1}^{t}\left(\sum_{c_{i_j}>0} c_{i_j}\omega_s\left(a_{i_j}\right)\right),$$
$$B = \sum_{s=1}^{t}\left(\sum_{c_{i_k}<0} c_{i_k}\omega_s\left(a_{i_k}\right)\right).$$

Afterwards, we apply the erasing matrices (3) and (4) until Z, Xs and Ys are completely removed where

$$D' :\ S' \overset{*}{\Rightarrow} w_t X^A Y^B Z \overset{*}{\Rightarrow} w_t. \qquad (5)$$

Derivation (5) is possible since
$A + B = \sum_{i=1}^{n} c_i \mu(a_i) = F(a_1, a_2, \ldots, a_n) = 0.$

(ii) From the other hand, we show that $L(G') \subseteq L(G,0,=)$. We consider a successful derivation D' in G'. Any derivation in G' starts with applying m_0, then any matrix from (2) – (4) can be applied. Yet, as soon as matrix (3) is applied, matrices (2) further cannot be applied. Without loss of generality, we can assume that

$$D' :\ S' \overset{m_0}{\Rightarrow} SZ \overset{m_{r_1} m_{r_2} \ldots m_{r_t}}{\Longrightarrow} w_t X^A Y^B Z \overset{m_{\lambda,Z}}{\Longrightarrow} w X^A Y^B \overset{m_\lambda}{\Rightarrow} w_t$$

where $A = \sum_{S=1}^{t}\left(\sum_{c_{i_j}>0} c_{i_j}\omega_S\left(a_{i_j}\right)\right)$ and $B = \sum_{S=1}^{t}\left(\sum_{c_{i_k}<0} c_{i_k}\omega_S\left(a_{i_k}\right)\right)$.

Since m_λ matrix erases all Xs and Ys, $A + B = 0$. From the other hand,

$$\begin{aligned} A + B &= \sum_{S=1}^{t}\left(\sum_{c_{i_j}>0} c_{i_j}\omega_S\left(a_{i_j}\right)\right) + \sum_{S=1}^{t}\left(\sum_{c_{i_k}<0} c_{i_k}\omega_S\left(a_{i_k}\right)\right) \\ &= \sum_{S=1}^{t} c_{i_j}\left(\sum_{c_{i_j}>0} \omega_S\left(a_{i_j}\right)\right) + \sum_{S=1}^{t} c_{i_k}\left(\sum_{c_{i_k}<0} \omega_S\left(a_{i_k}\right)\right) \\ &= \sum_{c_{i_j}>0} c_{i_j}\mu\left(a_{i_j}\right) + \sum_{c_{i_k}<0} c_{i_k}\mu\left(a_{i_k}\right) \\ &= \sum_{i=1}^{n} c_i \mu(a_i) = F(a_1, a_2, \ldots, a_n). \end{aligned}$$

Then, the corresponding derivation in G is $D : S \overset{r_1 r_2 \ldots r_t}{\Longrightarrow} w$.

Next, we give an example to illustrate the idea of construction the matrix grammar for a multiset controlled grammar.

Example 8. Consider the language $L(G_8, \{0\}, =) = \{a^n b^n c^n d^n : n \geq 1\} \in \mathbf{CS} \cap m\mathbf{CF} - \mathbf{CF}$ generated by multiset controlled grammar $G_8 = (\{A, B, S\}, \{a, b, c, d\}, S, P, \oplus, F)$ with productions

$r_0 : S \to AB[(0,0,0,0)]$,
$r_1 : A \to aAb[(1,1,0,0)]$,
$r_2 : A \to ab[(1,1,0,0)]$,
$r_3 : B \to cBd[(0,0,1,1)]$,
$r_4 : B \to cd[(0,0,1,1)]$ and
$F(a,b,c,d) = \mu(a) + \mu(b) + (-1)\mu(c) + (-1)\mu(d)$.

We construct the matrix grammar $G_8{}' = (\{A, B, S, Z, X, Y\}, \{a, b, c, d\}, S, M)$ simulating G_8 with production such

$m_0 : (S' \to SZ)$,
$m_1 : (S \to AB[Z \to Z])$,
$m_2 : (A \to aAb[Z \to Z_1 Z_2 Z])$,
$m_3 : (A \to ab[Z \to Z_1 Z_2 Z])$,
$m_4 : (B \to cBd[Z \to Z_3 Z_4 Z])$,

© 2005 – ongoing JATIT & LLS
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$m_5 : (B \to cd[Z \to Z_3Z_4Z]),$
$m_6 : (Z_1 \to X, Z_2 \to X),$
$m_7 : (Z_3 \to Y, Z_4 \to Y),$
$m_8 : (Z \to \lambda),$
$m_9 : (X \to \lambda, Y \to \lambda).$

Now, we show the derivation of those two grammars using a string $a^2b^2c^2d^2$.

(1) By multiset controlled grammar

$$S \overset{r_0}{\Rightarrow} AB[(0,0,0,0)] \overset{r_1}{\Rightarrow} aAbB[(1,1,0,0)]$$
$$\overset{r_3}{\Rightarrow} aAbcBd[(1,1,1,1)] \overset{r_2}{\Rightarrow} aabbcBd[(2,2,1,1)]$$
$$\overset{r_4}{\Rightarrow} aabbccdd[(2,2,2,2)] = a^2b^2c^2d^2 \text{ and}$$
$$F(a,b,c,d) = \mu(2) + \mu(2) + (-1)\mu(2) + (-1)\mu(2) = \mathbf{0}.$$

(2) By matrix grammar

$$S' \overset{m_0}{\Rightarrow} SZ \overset{m_1}{\Rightarrow} ABZ \overset{m_2}{\Rightarrow} aAbBZ_1Z_2Z$$
$$\overset{m_4}{\Rightarrow} aAbcBdZ_1Z_2Z_3Z_4Z$$
$$\overset{m_3}{\Rightarrow} aabbcBdZ_1Z_1Z_2Z_2Z_3Z_4Z$$
$$\overset{m_5}{\Rightarrow} aabbccddZ_1Z_1Z_2Z_2Z_3Z_3Z_4Z_4Z$$
$$\overset{m_8}{\Rightarrow} aabbccddZ_1Z_1Z_2Z_2Z_3Z_3Z_4Z_4$$
$$\overset{m_6}{\Rightarrow} aabbccddXZ_1YZ_2Z_3Z_3Z_4Z_4$$
$$\overset{m_6}{\Rightarrow} aabbccddXXXXZ_3Z_3Z_4Z_4$$
$$\overset{m_7}{\Rightarrow} aabbccddXXXXYYZ_4Z_4$$
$$\overset{m_7}{\Rightarrow} aabbccddXXXXYYYY \overset{m_9}{\Rightarrow} aabbccddXXXYYY$$
$$\overset{m_9}{\Rightarrow} aabbccddXXYY \overset{m_9}{\Rightarrow} aabbccddXY$$
$$\overset{m_9}{\Rightarrow} aabbccdd = a^2b^2c^2d^2.$$

5. CONCLUSION

In this paper, we defined a new variant of regulated grammar called multiset controlled grammar as well as investigated its generative power. We showed that multiset controlled grammars are more powerful than the traditional Chomsky grammar and have at least the lower bound of computational power as additive valence grammars as well as they also can generate the languages that are included in family of languages of matrix. A comprehensive picture of the hierarchy of languages generated by multiset controlled grammars is portrayed in Figure 1.

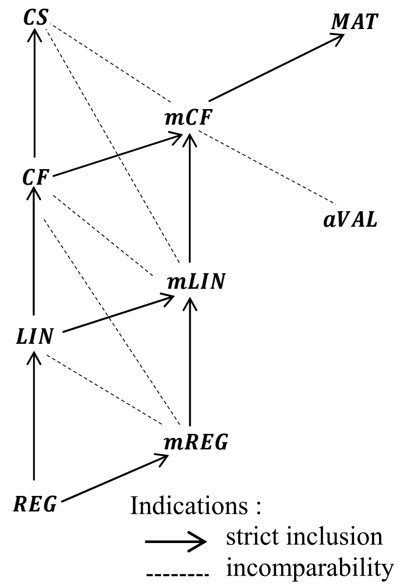

Figure 1. The hierarchy of families of language generated by multiset controlled grammar.

Yet, some of the interesting questions are still remaining unanswered such:

(1) Is $\mathbf{CF} - m\mathbf{REG} \neq \emptyset$? We conjecture that it is true. Perhaps, $L(G) = \{ww^R \mid ww \in \{a,b\}^+\}$ cannot be generated by any regular multiset controlled grammar.
(2) Is $\mathbf{CS} - m\mathbf{CF} \neq \emptyset$? We conjecture that it is also true. $L(G) = \{ww \mid ww \in \{a,b\}^+\}$ might not be generated by any context-free multiset controlled grammar.

For both questions, we do not have strong proves but we just knew that multiset can only do the counting but cannot remember the order of the strings. That way, both languages ww^R and ww are still cannot be constructed using multiset.

ACKNOWLEDGMENT

This work has been supported through International Islamic University Malaysia Research Initiative Grant Scheme **RIGS16-368-0532**.

REFERENCES:

[1] Dassow, J., Paun, G. Regulated rewriting in formal language theory. Springer Berlin Heidelberg: New York, 1989.
[2] Meduna, A., Zemek, P. Regulated grammars and automata. Springer Berlin Heidelberg: New York, 2014.
[3] Lobo, D., Vico, F.J., Dassow, J. Graph grammars with string-regulated rewriting. *Theoretical Computer Science*. 2011, 412(43), pp. 6101-11.
[4] Stiebe, R. On grammars controlled by Parikh vectors. In: Language Alive, LNCS 7300. Springer Berlin Heidelberg: New York. 2012, pp. 246–64.
[5] Dassow, J., Paun, G., Salomaa A. Grammars with controlled derivations. Handbook Formal Language, Vol 2 Linear Modelling: Background Application. Springer Berlin Heidelberg: New York. 1997, pp. 101–54.
[6] Dassow, J., Turaev, S. Petri net Controlled Grammars with a Bounded Number of Additional Places. Journal of Acta Cybernetica, 19, 2010, pp. 609-634.
[7] Dassow, J., Turaev, S. k-Petri net controlled grammars. In: Language and Automata Theory and Applications, LATA 2008, LNCS 5196, Springer Berlin Heidelberg: New York. 2008, pp. 209-20.
[8] Dassow, J., Turaev, S. Arbitrary Petri Net Controlled Grammars. In: Proceedings of 2nd International Workshop on Non-Classical Formal Languages in Linguistics, Tarragona, Spain, 2008, pp. 27-40.
[9] Stiebe, R., Turaev, S. Capacity Bounded Grammars and Petri Nets. In: J. Dassow, G. Pighizzini, B. Truthe (Eds.), 11th International Workshop on Descriptional Complexity of Formal Systems (DCFS 2009) EPTCS 3, 2009, pp. 193-203.
[10] Castano, J.M. Global index grammars and descriptive power. *Journal of Logic, Language and Information*. 2004, 13(4), pp. 403–419.
[11] Blizard, W,D. Multiset theory. *Notre Dame Journal of Formal Logic*. 1989, 30(1), pp. 36–66.
[12] Wildberger, N.J. A new look at multiset. School of Mathematics, UNSW, Sydney 2052, 2003.
[13] Singh, D., Ibrahim, A.M., Yohanna, T., Singh J.N. A Systematization of fundamentals of multisets. Lectures Matematicas 29. 2008, pp. 33–48.
[14] Kudlek, M., Martin, C., Paun, G. Toward a formal macroset theory. In: Calude, C.S., Paun, G., Rozenberg, G., Salomaa, A. (Eds.), Multiset Processing, WMC 2000, LNCS 2235, Springer Berlin Heidelberg: New York. 2001, pp. 123-33.
[15] Kudlek, M., Mitrana, V. Closure properties of multiset language families. *Fundamenta Informaticae*. 2002, 49(1-3), pp. 191–203.
[16] Csuhaj-Varju, E., Martin-Vide, C., Mitrana, V. Multiset automata. In: Calude CS, Paun G, Rozenberg G, Salomaa A (Eds.), Multiset Processing, WMC 2000, LNCS 2235, Springer Berlin Heidelberg: New York. 2001, pp. 69–83.
[17] Cavaliere, M., Freund, R., Oswald, M., Sburlan, D. Multiset random context grammars, checkers and transducers. *Theoretical Computer Science*. 2007, 372(2–3), pp. 136–51.
[18] Wang, J., Yin, M., Gu, W. Fuzzy multiset finite automata and their languages. *Soft Computing*. 2013, 17(3), pp. 381–90.
[19] Tiwari, S.P., Gautam, V., Dubey, M.K. On fuzzy multiset automata. *Journal of Applied Mathematics and Computing*. 2015, 51(1), pp. 643–57.
[20] Salomaa, A. Formal languages. Academic Press: New York, 1973.
[21] Sipser, M. Introduction to the theory of computation. 3rd edn. Cengage Learning: United States of America, 2013.
[22] Sharma, B.K., Syropoulos A., Tiwari, S.P. On fuzzy multiset regular grammars. *Annals of Fuzzy Mathematics and Informatics*. 2016, 12(5), pp. 617-639.