HAL
archives-ouvertes.fr

Automata and Logics for Unranked and Unordered Trees

Iovka Boneva, Jean-Marc Talbot

▶ **To cite this version:**

Iovka Boneva, Jean-Marc Talbot. Automata and Logics for Unranked and Unordered Trees. 20th International Conference on Rewriting Techniques and Applications, 2005, Nara, Japan. pp.500–515. inria-00536694

HAL Id: inria-00536694
https://hal.inria.fr/inria-00536694

Submitted on 16 Nov 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Automata and Logics for Unranked and Unordered Trees

Iovka Boneva and Jean-Marc Talbot

Laboratoire d'Informatique Fondamentale de Lille - UMR CNRS/USTL 8022
INRIA Futurs - MOSTRARE Project

Abstract. In this paper, we consider the monadic second order logic (MSO) and two of its extensions, namely Counting MSO (CMSO) and Presburger MSO (PMSO), interpreted over unranked and unordered trees. We survey classes of tree automata introduced for the logics PMSO and CMSO as well as other related formalisms; we gather results from the literature and sometimes clarify or fill the remaining gaps between those various formalisms. Finally, we complete our study by adapting these classes of automata for capturing precisely the expressiveness of the logic MSO.

1 Introduction

Relationship between logics and tree automata for ranked trees has been established by Thatcher and Wright in their seminal paper [19]: they proved that languages of finite and ranked trees that are accepted by tree automata coincide with the models of monadic second-order logic (MSO) sentences when interpreted over (ranked) tree structures.

Recently, due to the development of semi-structured databases and in particular, of XML, there has been some new interest in unranked and ordered trees; for those trees, the number of children of some node is not *a priori* bounded and for instance, does not depend on the symbol labeling this position in the tree. Moreover, those trees are said to be ordered in the sense that there exists a total ordering on children of each node. The relationship between logics and automata has been carried over unranked and ordered trees [13],[1]: once again, languages that are definable by means of tree automata are exactly models of MSO sentences.

In this paper we consider unranked and unordered trees, *ie* trees that are unranked but without any ordering relation between children of the same node. As noticed by Courcelle in [4], the fact that there is no order between siblings drastically reduces the expressiveness of MSO: hence, for ordered unranked trees, properties such as "the root has an even number of children labeled with b" or such as "the number of nodes in the tree is a multiple of 5" can be expressed in MSO (where the ordering relation on sibling nodes is represented as an ordering relation or as some successor relation). It goes differently for unranked and unordered trees where those two latter properties can no longer be expressed in MSO. Courcelle proposed in [4] to extend MSO with some constraints for counting modulo on cardinalities of sets. He showed that this logic, named Counting MSO (CMSO), can be related to tree automata by the notion of algebraic recognizability in the sense of [12]: a set of trees can be expressed by some CMSO sentence iff it is recognizable.

Recently, Seidl, Schwentick and Muscholl introduced Presburger monadic second order logic (PMSO) [18]: it extends MSO with a new kind of atomic formula x/ϕ; in such an atomic formula, x is a variable denoting a node of the tree and ϕ is a Presburger formula expressing arithmetical constraints on the cardinality of sets when restricted to the children of x. Seidl *et al.* also defined a notion of automata, called Presburger tree automata, and showed that tree languages accepted by Presburger tree automata are precisely models of PMSO sentences.

The objective of this paper is two folds: first, we gather results concerning formalisms that can express sets of unranked and unordered trees definable by PMSO and CMSO sentences. This survey permits to clarify or sometimes make explicit the relationship of different formalisms, in particular, various classes of tree automata (*eg* Presburger tree automata [18], ACU equational tree automata [15], [20] and equational tree languages [4] when considering the logic PMSO). Our second aim is to try to get a uniform view on tree languages that can be defined by the logics CMSO and PMSO, but also by MSO: in particular, for PMSO and CMSO, we try to adapt systematically (when possible) a formalism associated to some specific logic to the other one. Finally, we investigate the expressiveness of the logic MSO: considering formalisms used for describing CMSO and PMSO definable sets, we propose subclasses capturing precisely MSO over unranked and unordered trees.

This paper is organized as follows: Section 2 presents definitions for trees as graphs, an algebraic view of trees and recall Presburger formulas. In Section 3 we define the three logic formalisms MSO, CMSO and PMSO. Sections 4 and 5 survey PMSO- and CMSO-complete formalisms respectively, and in Sections 6 and 7 we present new characterizations of PMSO- and CMSO-definable sets of trees. Finally, in Section 8 we give characterizations of MSO-definable sets of trees.

2 Preliminaries

2.1 The Tree Model

We consider here edge-labeled[1] unranked and unordered trees (called simply trees in the rest of the paper).

Trees will be finite non-empty directed graphs with a distinguished node, the root of the tree, such that for any node, there exists exactly one path from the root to this node. Additionally, we suppose a mapping associating with each edge of the graph a label from a finite set Λ. Formally, a tree is given by a triple (V, E, λ) such that V is a finite non-empty set of nodes, $E \subseteq V \times V$ is a finite set of edges and λ is a mapping from E to Λ. Moreover, it satisfies that any node is reached by a unique path from the root: for any nodes $v_n, v'_{n'}$, for any two sequences $v_0, v_1, \ldots, v_n$ and $v'_0, v'_1, \ldots, v'_{n'}$ such that v_0, v'_0 both denote the root of the tree, v_n,$v'_{n'}$ are equal and (v_i, v_{i+1}), (v'_j, v'_{j+1}) belong to E for all $0 \leq i \leq n-1$, $0 \leq j \leq n'-1$, the two sequences are identical.

[1] For simplicity, we assume that nodes are unlabeled. However, the results presented here could be extended to trees where both edges and nodes are labeled.

As usual, we consider two isomorphic trees as being equal. We denote Tree the set of all trees. We denote $\mathsf{root}(\tau)$ the root of the tree τ and for any node v, $\mathsf{children}(v)$ the set of nodes $\{v' \mid (v, v') \in E\}$.

2.2 An Algebraic View of Trees

We adopt the algebraic view of trees proposed in [4]. We consider the signature Σ given by the constant $\mathbf{0}$, the unary function symbols a for each a in Λ and the binary (infix) symbol $|$.

Let $\mathcal{T}$ be the Σ-algebra whose domain is the set of all finite edge-labeled trees. The constant $\mathbf{0}$ is interpreted in $\mathcal{T}$ as $\mathbf{0}^{\mathcal{T}}$ the tree having one single node and no edge (we consider only non-empty graphs). For any tree τ defined as (V, E, λ), the tree $a^{\mathcal{T}}(\tau)$ is given by $(V \cup \{r\}, E \cup \{r, \mathsf{root}(\tau)\}, \lambda')$ where r is a new node (not belonging to V) and λ' extends λ by letting $\lambda'((r, \mathsf{root}(\tau))) = a$. For trees τ, τ' defined as (V, E, λ), (V', E', λ') respectively, $\tau \mid^{\mathcal{T}} \tau'$ is the tree given by (V'', E'', λ'') where (assuming $V \cap V' = \varnothing$):

- $V'' = (V \cup V' \cup \{r\}) \setminus \{\mathsf{root}(\tau), \mathsf{root}(\tau')\}$ (where $r \notin V \cup V'$)
- $E'' = \{(r, v) \mid v \in \mathsf{children}(\mathsf{root}(\tau)) \cup \mathsf{children}(\mathsf{root}(\tau'))\} \cup (E \setminus \{(\mathsf{root}(\tau), v) \mid v \in V\}) \cup (E' \setminus \{(\mathsf{root}(\tau'), v') \mid v' \in V\})$.
- λ'' is defined as λ and λ' for edges in E'' coming from E and E' respectively and by $\lambda''((r, v)) = \lambda((\mathsf{root}(\tau), v))$ if $v \in E$ and $\lambda''((r, v)) = \lambda((\mathsf{root}(\tau'), v))$ if $v \in E'$.

Informally, $a^{\mathcal{T}}(t)$ adds a new edge labeled by a from a new node (the new root) to the ancient root of t whereas $t \mid^{\mathcal{T}} t'$ is obtained from t and t' by merging their roots. Figure 1 illustrates algebraic operations on trees.

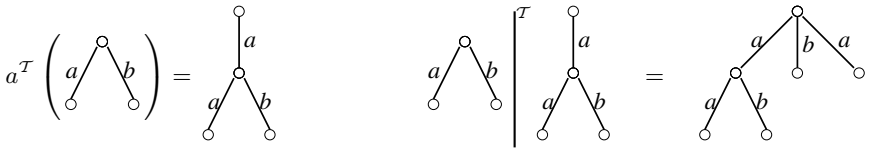

Fig. 1. Algebraic operations over trees.

One can remark that the set of trees Tree is finitely generated by Σ, that is each tree in Tree can be obtained by combining the operators from the Σ-algebra $\mathcal{T}$.

It should also be noticed that the operation $\mid^{\mathcal{T}}$ is associative and commutative over trees and that $\mathbf{0}^{\mathcal{T}}$ is its neutral element. Therefore, $(\mathsf{Tree}, \mid^{\mathcal{T}}, \mathbf{0}^{\mathcal{T}})$ is a commutative monoid.

We will also consider $\mathcal{C}$ the algebra of terms built over the signature Σ (ie the initial algebra over Σ). We will denote $h_{\mathcal{C}}$ the unique homomorphism from $\mathcal{C}$, the Σ-algebra of terms to $\mathcal{T}$, the algebra of trees.

2.3 Arithmetical Formulas

In this paper, we will have to consider different kinds of arithmetical formulas interpreted over $\mathbb{N}$ the set of natural numbers. Different logics will be defined depending on the atomic predicates that are allowed.

Let $\mathcal{U}$ be a set of natural variables and B be a set of atomic formulas whose free variables belong to $\mathcal{U}$. We define $\mathcal{F}_{\mathcal{U}}(B)$ as the least set of formulas such that (i) B is included in $\mathcal{F}_{\mathcal{U}}(B)$ and (ii) if ϕ,ϕ' are in $\mathcal{F}_{\mathcal{U}}(B)$ then $\phi \wedge \phi'$, $\neg\phi$ are in $\mathcal{F}_{\mathcal{U}}(B)$ as well.

For our purpose, we are going to consider only two kinds of atomic formulas: $p \leq p'$ and $Div_k(p)$, where k is some fixed natural number different from zero and p is an arithmetical term defined as:

$$p ::= n \mid u \mid p + p \qquad (u \in \mathcal{U}, n \in \mathbb{N})$$

Formulas in $\mathcal{F}_{\mathcal{U}}(B)$ are interpreted over $(\mathbb{N}, \{+\}, \{\leq, Div_k\})$ the structure of naturals where $+$ is interpreted as the addition function, $\leq$ as the usual ordering over $\mathbb{N}$ and finally, Div_k is the unary predicate such that $Div_k(n)$ holds if n is divisible by k. The semantics for Boolean connectives is the usual one.

Let ϕ be a formula from $\mathcal{F}_{\mathcal{U}}(B)$. We say that a valuation μ mapping free variables of ϕ to naturals is a solution of ϕ if the structure $(\mathbb{N}, \{+\}, \{\leq, Div_k\})$ is a model of ϕ under the valuation μ.

Formulas from $\mathcal{F}_{\mathcal{U}}(\{p \leq p', Div_k(p)\})$ are called *Presburger formulas* and the ones from $\mathcal{F}_{\mathcal{U}}(\{p \leq p'\})$ are called *ordering formulas*.

Strictly speaking, Presburger formulas usually allow also existential quantification $\exists u.\phi$. However, it is well-known that for any Presburger formula ϕ with quantification, there exists an equivalent (quantifier-free) formula ϕ' from $\mathcal{F}_{\mathcal{U}}(\{p \leq p', Div_k(p)\})$.[2] Note that this is not the case for ordering formulas for which adding existential quantification strictly increases their expressiveness.[3]

An atomic formula from $\{p \leq p', Div_k(p)\}$ is said to be *unary* if this formula contains only one variable (but possibly several occurrences of it). By extension, a formula ϕ from $\mathcal{F}_{\mathcal{U}}(\{p \leq p', Div_k(p)\})$ is *unary* if it is built over unary atomic formulas. Note that a unary formula may contain several different variables but any of its atoms contains only one variable.

We will denote $\mathcal{F}^1_{\mathcal{U}}(\{p \leq p', Div_k(p)\})$ (resp. $\mathcal{F}^1_{\mathcal{U}}(\{p \leq p'\})$) the set of unary Presburger formulas (resp. of unary ordering formulas).

2.4 Presburger-Definable Sets and Multiset Languages

Let $\mathbb{N}^l$ be the set of tuples of length l of naturals. A subset N of $\mathbb{N}^l$ is said to be *Presburger-definable* (resp. *ordering-definable*) if there exists a Presburger formula

[2] The first-order theory of formulas built over $\{p \leq p', Div_k(p)\}$ interpreted over natural numbers admits quantifier elimination.

[3] In presence of existential quantifications, Presburger and ordering formulas are equally expressive as $Div_k(p)$ can be written as $\exists y.p = \underbrace{y + \ldots + y}_{k}$.

(resp. an ordering formula) ϕ whose free variables are $(x_1, \ldots, x_l)$ considered as totally ordered and such that for any tuple $(n_1, \ldots, n_l)$ from N, the valuation $\{x_1 \mapsto n_1, \ldots, x_l \mapsto n_l\}$ is a solution of ϕ.

Let $A = (a_1, \ldots, a_l)$ be a sequence of symbols. We denote $\mathbb{M}(A)$ the set of all multisets whose elements are in A. The Parikh mapping [17] is a mapping from $\mathbb{M}(A)$ to $\mathbb{N}^l$ defined as $\pi_A(m) = (|m|_{a_1}, \ldots, |m|_{a_l})$, where $|m|_{a_i}$ is the number of occurrences of a_i in the multiset m. Parikh mappings are extended as mappings from multiset languages to subsets of $\mathbb{N}^l$ as follows: for $M \subseteq \mathbb{M}(A)$, $\pi_A(M) = \{\pi_A(m) \mid m \in M\}$.

Denoting $\varnothing$ the empty multiset and $\uplus$ the multiset union,

Definition 1. *The family $Rat(\mathbb{M}(A))$ of rational multiset languages is the least subset of $\mathbb{M}(A)$ which contains any finite subset of $\mathbb{M}(A)$ and such that if L, L' belong to $Rat(\mathbb{M}(A))$ then $L \cup L'$, $L \uplus L' = \{m \uplus m' \mid m \in L \text{ and } m' \in L'\}$, $L^* = \bigcup_{n \in \mathbb{N}} L^n$ (where $L^0 = \varnothing$ and $L^{i+1} = L^i \uplus L$ for $i > 0$) belong to $Rat(\mathbb{M}(A))$.*

It is well-known that

Note 1. Let N be a subset of $\mathbb{N}^l$ and $A = (a_1, \ldots, a_l)$ be some alphabet. Then N is Presburger-definable iff $\pi_A^{-1}(N) \in Rat(\mathbb{M}(A))$.

Definition 2. *A multiset language $L \in \mathbb{M}(A)$ is recognizable if there exists a monoid morphism h from $(L, \uplus, \{\varnothing\})$ to a finite monoid $(D, +, \iota)$ and a finite subset D' of D such that $L = h^{-1}(D')$.*

We denote $Rec(\mathbb{M}(A))$ the set of recognizable multiset languages. It is well-known that the set of recognizable multisets is strictly included into the set of rational multisets, *ie* $Rec(\mathbb{M}(A)) \subsetneq Rat(\mathbb{M}(A))$.

3 MSO-Based Logics for Trees

We consider in this section monadic second-order logic (MSO) as well as two extensions of it. First, let us recall how trees can be viewed as logical structures over which logical formulas are interpreted.

Let σ be the signature $\{\mathsf{label}_a \mid a \in \Lambda\}$ where the label_a's are binary predicates. With a tree $\tau = (V, E, \lambda)$, we associate a finite σ-structure $\mathcal{S}^\tau = \langle V, \{\mathsf{label}_a^\tau \mid a \in \Lambda\}\rangle$, such that $\mathsf{label}_a^\tau(v, v')$ holds in $\mathcal{S}^\tau$ if $(v, v') \in E$ and $\lambda((v, v')) = a$.

We assume a countable set of first-order variables ranging over by $x, y, z, \ldots$ and a countable set of second-order variables ranging over by $X, Y, Z, \ldots$.

Definition 3. *The formulas of the logic MSO are defined by the following syntax:*

$$\psi ::= \mathsf{label}_a(x, y) \mid x \in X \mid \psi \vee \psi \mid \neg\psi \mid \exists x.\psi \mid \exists X.\psi$$

Let $\mathcal{S}$ be a σ-structure whose domain is V. Let ρ be a valuation mapping first-order variables to elements of V and second-order variables to subsets of V. The structure $\mathcal{S}$ is a model of a MSO formula ψ under the valuation ρ (defined for free variables of ψ) denoted $\mathcal{S} \models_\rho \psi$, if:

– ψ is $\mathsf{label}_a(x, y)$ and $\mathsf{label}_a(\rho(x), \rho(y))$ holds in $\mathcal{S}$;
– ψ is $x \in X$ and $\rho(x)$ belongs to $\rho(X)$;
– ψ is $\psi_1 \vee \psi_2$ (resp. $\neg\psi'$) and $\mathcal{S} \models_\rho \psi_1$ or $\mathcal{S} \models_\rho \psi_2$ (resp. $\mathcal{S} \not\models_\rho \psi'$) holds;
– ψ is $\exists x.\psi'$ and there exists an element v from V s.t. $\mathcal{S} \models_{\rho[x \to v]} \psi'$ holds.
– ψ is $\exists X.\psi'$ and there exists a subset V' of V s.t. $\mathcal{S} \models_{\rho[X \to V']} \psi'$ holds.

Overloading the notation, for a closed MSO formula ψ and a tree τ, we write $\tau \models \psi$ whenever $\mathcal{S}^\tau \models \psi$ for the σ-structure $\mathcal{S}^\tau$ associated with τ; moreover, we write $[\![\psi]\!]$ to denote the set of all trees τ such that $\tau \models \psi$. We say that a set of trees T is *MSO-definable* if there exists some closed MSO formula ψ such that $[\![\psi]\!] = T$.

The logic CMSO Courcelle defined in [4] the counting MSO logic (CMSO) as an extension of MSO. The syntax of CMSO[4] augments the one from MSO with an atomic formula $Mod^i_j(X)$ where X is a second-order variable and i, j are naturals such that $i \neq 0$ and $j < i$. The formula $Mod^i_j(X)$ holds for a σ-structure $\mathcal{S}$ and a mapping ρ associating with X a subset of the domain of $\mathcal{S}$ if the cardinality of $\rho(X)$ modulo i is equal to j.

The logic PMSO Seidl *et al.* introduced in [18] an extension of MSO called Presburger MSO (PMSO). This extension is defined by a new kind of atomic formulas of the form x/ϕ, ϕ being a Presburger formula from $\mathcal{F}_\mathcal{V}(\{p \leq p', Div_k(p)\})$, where $\mathcal{V}$ is the set of integer variables $\{\#X \mid X \text{ is a second-order variable}\}$.

The formula x/ϕ holds in some σ-structure $\mathcal{S}$ under a valuation ρ if the valuation μ mapping each variable $\#X$ from ϕ to the cardinality of the set $\rho(X) \cap \mathsf{children}(\rho(x))$ is a solution for ϕ.[5]

CMSO-definable and *PMSO-definable* set of trees are defined on the same way that MSO-definable set of trees.

4 A Survey on PMSO-Complete Formalisms

In this section, we present various formalisms which are able to express precisely PMSO definable sets of trees.

4.1 Presburger Tree Automata

In [18], Seidl *et al.* introduced Presburger tree automata which correspond to the logic PMSO. We define here an adaptation of these automata for edge-labeled trees. Later on we identify precisely subclasses of these automata for the logics MSO and CMSO. These automata are also very close to sheaves automata from [8],[7].

[4] Actually, the syntax of CMSO from [4] is richer than the one we consider here; there, the logic has two sorts for both individual and set variables, respectively a sort for nodes and a sort for edges. However, Courcelle showed in [5] that this two-sorted extension does not add expressive power when trees are considered.

[5] PMSO allows to express quite complex relationships between cardinalities of sets; however, those sets are always relative to some precise node. For arbitrary sets, the associated monadic second order logic would be undecidable [11].

Definition 4. *A* Presburger tree automaton *(PTA) is given by a tuple* (Λ, Q, F, δ) *where* Λ *is a finite set of labels,* Q *is a finite set of states,* δ *is a transition mapping from* $Q \times \Lambda$ *to* $\mathcal{F}_{\mathcal{U}}(\{p \leq p', Div_k(p)\})$ *where* $\mathcal{U}$ *is* $\{x_q \mid q \in Q\}$ *and finally,* $F \in \mathcal{F}_{\mathcal{U}}(\{p \leq p', Div_k(p)\})$ *is the acceptance condition.*

A run r_A for a tree $\tau = (V, E, \lambda)$ and a PTA $A = (\Lambda, Q, F, \delta)$ is a mapping from E to Q such that for all edges (v, v') in E, $\mu_v \models \delta(r_A((v, v')), \lambda((v, v')))$ where μ_v is the valuation associating with each variable x_q the cardinality of the set $\{(v', v'') \mid (v', v'') \in E$ and $r_A((v', v'')) = q\}$.

Informally, a run labels edges with states from Q: the state labeling some edge $e = (v, v')$ depends on the label of the edge as well as on the multiplicity of the states labeling the edges originating from the node v' (*ie* edges of the form (v', v'') for some node v'').

A tree $\tau = (V, E, \lambda)$ is accepted by a Presburger tree automaton $A = (\Lambda, Q, F, \delta)$ if there exists a run r for τ and A such that $\mu_F \models F$ where μ_F is the valuation associating with each variable x_q the cardinality of the set $\{(\mathsf{root}(\tau), v) \mid (\mathsf{root}(\tau), v) \in E$ and $r_A((\mathsf{root}(\tau), v)) = q\}$. For some PTA A, we denote $L(A)$ the set of all trees accepted by A.

Example 1. The Presburger tree automaton A_1 here after accepts precisely the set of trees of height 1 such that the root has as many a outgoing edges as b ones: $A_1 = (\{a, b\}, \{q_a, q_b\}, x_{q_a} = x_{q_b}, \delta)$ where δ is the transition mapping such that $\delta((q_a, a)) = \delta((q_b, b)) = x_{q_a} \leq 0 \wedge x_{q_b} \leq 0$ and $\delta((q_a, b)) = \delta((q_b, a)) = \mathit{false}$. The automaton A_2 accepts precisely the set of trees satisfying that each node has as many a outgoing edges as b ones: $A_2 = (\{a, b\}, \{q_a, q_b\}, x_{q_a} = x_{q_b}, \delta)$ where δ is the transition mapping such that $\delta((q_a, a)) = \delta((q_b, b)) = (x_{q_a} = x_{q_b})$ and $\delta((q_a, b)) = \delta((q_b, a)) = \mathit{false}$.

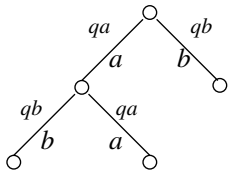

Fig. 2. Run of the automaton A_2.

Theorem 1. *[18] For any set of trees* T*,* T *is PMSO-definable iff there exists some Presburger tree automaton* A *that accepts* T*.*

4.2 Rational-Multiset Tree Automata

Colcombet proposed in [2] rational-multiset tree automata. We give here a slightly rephrased definition of those automata.

Definition 5. *A rational-multiset automaton (RatMA) is a tuple (Λ, Q, F, δ) where Λ is a finite set of labels, Q is a finite set of states, δ is a transition mapping from $Q \times \Lambda$ to $Rat(\mathbb{M}(Q))$ and $F \in Rat(\mathbb{M}(Q))$ is the acceptance condition.*

A run r_A for a tree $\tau = (V, E, \lambda)$ and a RatMA $A = (\Lambda, Q, F, \delta)$ is a mapping from E to Q such that for all edges (v, v') in E, the multiset $\{r_A((v', v_1)), \ldots, r_A((v', v_n))\}$ belongs to $\delta(r_A((v, v')), \lambda((v, v')))$, $v_1 \ldots v_n$ being exactly the children of v'.

A tree $\tau = (V, E, \lambda)$ is accepted by a RatMA $A = (\Lambda, Q, F, \delta)$ if there exists a run r for τ and A such that $\{r_A((\mathsf{root}(\tau), v_1)), \ldots, r_A((\mathsf{root}(\tau), v_n))\}$ belongs to F, $v_1 \ldots v_n$ being exactly the children of $\mathsf{root}(\tau)$. For some RatMA A, we denote $L(A)$ the set of all trees accepted by A.

Using Note 1, it is straightforward that

Proposition 1. *For any set of trees T, T is PMSO-definable iff there exists a rational-multiset automaton A that accepts T.*

4.3 ACU Equational Tree Automata

Let us consider the equational theory ACU stating that $|$ is associative and commutative and that $\mathbf{0}$ is its neutral element. Formally,

$$\mathsf{ACU} \begin{cases} x \mid \mathbf{0} = x \\ x \mid y = y \mid x \\ x \mid (y \mid z) = (x \mid y) \mid z \end{cases}$$

We write $t \simeq_{\mathsf{ACU}} t'$ whenever the two Σ-terms t and t' are equal modulo ACU. It is well-known that even when a term language L is regular, its ACU-closure, that is the set of terms $\{t \mid t \simeq_{\mathsf{ACU}} t' \text{ and } t' \in L\}$, may not be regular.

For dealing with languages obtained as closure of regular term languages by some equational theory, Ohsaki [15],[16] and Verma [20] have independently introduced so-called *equational tree automata*.[6]

An ACU equational tree automaton A over the signature Σ is given by a tuple (Σ, Q, F, Δ) where Q is a finite set of states, $F \subseteq Q$ is the set of final states and Δ is a finite set of transition rules of the form (q, q_1, q_2 being states from Q and a a unary symbol from Σ):

$$\mathbf{0} \to q \qquad a(q_1) \to q \qquad q_1 \mid q_2 \to q$$

A run for a Σ-term t in an ACU equational tree automaton $A = (\Sigma, Q, F, \Delta)$ is a sequence $t_1, \ldots, t_n$ of terms built over the signature $\Sigma \cup Q$ (where states from Q are considered as constants) such that $t_1 = t$, $t_n \in Q$ and for all $1 \leq i \leq n$, $t_i \simeq_{\mathsf{ACU}} t' \to_\Delta t'' \simeq_{\mathsf{ACU}} t_{i+1}$ for some terms t', t''. The relation $\to_\Delta$ is the ground rewriting relation induced by Δ. A run $t_1, \ldots, t_n$ is *accepting* if the state t_n belongs to F. A Σ-term t is accepted by some ACU equational tree automaton A if there exists

[6] For some equational theory, the classes of automata defined respectively in [15] and in [20] may differ. However, they do coincide for the ACU equational theory.

an accepting run for t in A. Finally, the language accepted by an ACU equational tree automaton A over the signature Σ is the set of all Σ-terms having an accepting run in A.

Definition 6. *A set of Σ-terms is* ACU-regular *if it is accepted by an* ACU *equational tree automaton.*

Ohsaki showed in [15] that a language E is ACU-regular iff there exists a regular set of Σ-terms E' such that $E = \{t' \mid t \simeq_{\mathsf{ACU}} t' \text{ and } t \in E'\}$.

Lemma 1. *For any two Σ-terms t, t', if $t \simeq_{\mathsf{ACU}} t'$ then $h_{\mathcal{C}}(t) = h_{\mathcal{C}}(t')$*

Proof. By definition, $t \simeq_{\mathsf{ACU}} t'$ holds iff there exists a sequence of terms $t_1, \ldots, t_n$ such that $t = t_1$, $t' = t_n$ and for all $i \in \{1, \ldots, n-1\}$, there exists an equation $l = r$ or $r = l$ in the ACU theory satisfying that $t_i = C[\theta(l)]$ and $t_{i+1} = C[\theta(r)]$ for some context C and some substitution θ mapping variables from l, r to Σ-terms. The proof goes by trivial induction over the context C.

From Colcombet's work [2], it follows easily that

Proposition 2. *For any* ACU*-closed set of Σ-terms E, E is* ACU*-regular iff $h_{\mathcal{C}}(E)$ is accepted by some rational-multiset automaton.*

As for any set of trees T, $h_{\mathcal{C}}^{-1}(T)$ is always ACU-closed (see Lemma 1),

Corollary 1. *For any set of trees T, T is accepted by some rational-multiset automaton iff $h_{\mathcal{C}}^{-1}(T)$ is* ACU*-regular.*

5 A Survey on CMSO-Complete Formalisms

We present here formalisms expressing precisely CMSO definable sets of trees.

5.1 Algebraic Recognizability

We focus first on the notion of *algebraic recognizability* in the sense of Mezei and Wright [12].

Definition 7 ([12]). *Let $\mathcal{M}$ be a Σ-algebra and B be a subset of the domain of $\mathcal{M}$. Then B is said to be* $\mathcal{M}$-recognizable *if there exists a finite Σ-algebra $\mathcal{A}$ with domain $dom(\mathcal{A})$, a homomorphism from $\mathcal{M}$ to $\mathcal{A}$ and a finite subset D of $dom(\mathcal{A})$ such that $B = h^{-1}(D)$.*

As a particular case, a tree language T is $\mathcal{T}$-recognizable if there exists a finite Σ-algebra $\mathcal{A}$ with domain $dom(\mathcal{A})$, a homomorphism from $\mathcal{T}$ to $\mathcal{A}$ and a finite subset D of $dom(\mathcal{A})$ such that $T = h^{-1}(D)$.

Theorem 2 ([4]). *For any set of trees T, T is CMSO-definable iff T is $\mathcal{T}$-recognizable.*

Starting from a sightly different algebra for trees, Niehren and Podelski defined in [14] a notion of (feature) tree automata for which accepted languages coincide with $\mathcal{T}$-recognizable sets of trees. Note that $\mathcal{T}$-recognizability can be defined alternatively as:

Proposition 3. *A tree language T is $\mathcal{T}$-recognizable iff there exists a finite Σ-algebra $\mathcal{A}$ with domain $dom(\mathcal{A})$ such that $(dom(\mathcal{A}), |^{\mathcal{A}}, \mathbf{0}^{\mathcal{A}})$ is a commutative monoid, h is a homomorphism from $\mathcal{T}$ to $\mathcal{A}$ and D is a finite subset of $dom(\mathcal{A})$ such that $T = h^{-1}(D)$.*

Proof. As T is $\mathcal{T}$-recognizable, there exists a finite Σ-algebra $\mathcal{A}$ with domain $dom(\mathcal{A})$, an homomorphism h from $\mathcal{T}$ to $\mathcal{A}$ and a finite subset D of $dom(\mathcal{A})$ such that $T = h^{-1}(D)$. Let us consider the sub-algebra $\mathcal{A}'$ of $\mathcal{A}$ whose domain is precisely $h(\mathsf{Tree})$. Obviously, T is $\mathcal{T}$-recognizable using the finite algebra $\mathcal{A}'$, the homomorphism h and the set $D \cap dom(\mathcal{A}')$. It is then easy to prove that $(dom(\mathcal{A}'), |^{\mathcal{A}'}, \mathbf{0}^{\mathcal{A}'})$ is a commutative monoid.

5.2 Recognizable-Multiset Tree Automata

In [6], Courcelle introduced a notion of tree automaton whose transitions are defined by means of recognizable sets of finite multisets. This notion can be rephrased in our settings as follows:

Definition 8. *A recognizable-multiset tree automaton is a rational-multiset tree automaton (Λ, Q, F, δ) such that $F \in Rec(\mathbb{M}(Q))$ and for all q in Q and a in Λ, $\delta(q, a) \in Rec(\mathbb{M}(Q))$.*

Theorem 3. *[6] For any set of trees T, T is CMSO-definable iff there exists some recognizable-multiset automaton A that accepts T.*

As recognizable sets of multisets are strictly included into rational sets of multisets, we have:

Corollary 2. *The PMSO logic is strictly more powerful than the CMSO logic over unranked and unordered trees.*

Courcelle proved in [4] that CMSO is strictly more expressive than MSO on unranked and unordered trees. So, this shows that MSO-CMSO-PMSO is a strict hierarchy for this kind of trees; this has to be contrasted with the case of ranked trees where it is known that MSO and CMSO have the same expressive power [4]. It is also not difficult to see that the extension to PMSO does not bring neither some new expressiveness for ranked trees. For unranked and ordered trees, it is quite simple to write an MSO formula for the atom $Mod_j^i(X)$, and thus, showing that MSO and CMSO have in that case the same expressiveness. But, PMSO is for unranked and ordered trees strictly more expressive than MSO [18].

6 New Characterizations for PMSO Definable Sets

We consider first sets of trees defined by means of a system of equations, namely, equational trees languages. Then as done in Section 5.1 for CMSO, we give a fully algebraic characterization of PMSO definable sets of trees.

6.1 Equational Tree Languages

Let $X_1, \ldots, X_n$ be a finite set of variables. We consider the signature $\Sigma \cup \{+\} \cup \{X_1, \ldots, X_n\}$ where $+$ is a binary symbol used in infix notation and $X_1, \ldots, X_n$ are considered as constants.

A system $\mathcal{S}$ of equations over the signature $\Sigma \cup \{+\}$ and the variables $X_1, \ldots, X_n$ is a set of equations of the form $X_i = s_i$ such that s_i is a term built over $\Sigma \cup \{+\} \cup \{X_1, \ldots, X_n\}$ and for each X_i, there exists precisely one equation in $\mathcal{S}$.

For a Σ-algebra $\mathcal{M}$ and a set of variables $\{X_1, \ldots, X_n\}$, a $\mathcal{M}$-valuation $\mathcal{I}$ is a mapping associating with each variable X_i a subset of the domain of $\mathcal{M}$. A $\mathcal{M}$-valuation $\mathcal{I}$ is extended to terms built over the signature $\Sigma \cup \{+\}$ as follows:

- $\mathcal{I}(\mathbf{0}) = \{\mathbf{0}^{\mathcal{M}}\}$
- $\mathcal{I}(a(s)) = \{a^{\mathcal{M}}(t) \mid t \in \mathcal{I}(s)\}$
- $\mathcal{I}(s_1 \mid s_2) = \{t_1 \mid^{\mathcal{M}} t_2 \mid t_1 \in \mathcal{I}(s_1), t_2 \in \mathcal{I}(s_2)\}$
- $\mathcal{I}(s_1 + s_2) = \mathcal{I}(s_1) \cup \mathcal{I}(s_2)$

A $\mathcal{M}$-valuation $\mathcal{I}$ is a solution of a system of equations $\mathcal{S}$ for the Σ-algebra $\mathcal{M}$ if for all equations $X_i = s_i$ in $\mathcal{S}$, it holds that $\mathcal{I}(X_i)$ is equal to $\mathcal{I}(s_i)$. Valuations (and thus, solutions) over the same set of variables are equipped with a natural partial ordering: $\mathcal{I}$ is smaller than $\mathcal{I}'$ if for all X_i, $\mathcal{I}(X_i) \subseteq \mathcal{I}(X_i')$. It is not difficult to prove that any system of equations $\mathcal{S}$ admits a least solution; we will denote $Least(\mathcal{S}, \mathcal{M})$ the least $\mathcal{M}$-valuation which is a solution of $\mathcal{S}$.

Definition 9 ([12]). *For a Σ-algebra $\mathcal{M}$, a subset L of the domain of $\mathcal{M}$ is equational if there exists a system of equations $\mathcal{S}$ (over the signature $\Sigma \cup \{+\}$) with some designated variable X such that $Least(\mathcal{S}, \mathcal{M})(X) = L$.*

As a particular case for the Σ-algebra $\mathcal{T}$, a set of trees T is equational if there exists a system of equations $\mathcal{S}$ with some designated variable X such that $Least(\mathcal{S}, \mathcal{T})(X) = T$. We denote $\mathsf{Equat}(\mathcal{T})$ the set of equational tree languages.

Courcelle proved in [4] that CMSO-definable languages are equational but that the converse is not true: some languages are equational but not CMSO-definable.

We recall in the next two propositions some useful properties of equational languages.

Proposition 4 ([12]). *Let $\mathcal{M}, \mathcal{M}'$ be two Σ-algebras and h a homomorphism from $\mathcal{M}$ to $\mathcal{M}'$. For any system of equations $\mathcal{S}$, for any variable X from $\mathcal{S}$, it holds that $Least(\mathcal{S}, \mathcal{M}')(X) = h(Least(\mathcal{S}, \mathcal{M})(X))$.*

Proposition 5 ([12]). *For the Σ-algebra of terms $\mathcal{C}$, a language L is equational (ie $L \in \mathsf{Equat}(\mathcal{C})$) iff L is regular (ie accepted by some "classical" tree automaton).*

Theorem 4. *For any set of trees T, T is PMSO-definable iff $T \in \mathsf{Equat}(\mathcal{T})$.*

Proof. By proposition 5, a set S of Σ-terms is regular iff it is equational over $\mathcal{C}$. By Proposition 4, we have that $h_{\mathcal{C}}(S)$ is equational over $\mathcal{T}$. Conversely, if T is equational over $\mathcal{T}$ then still by Proposition 4, there exists S equational over $\mathcal{C}$ such that $T = h_{\mathcal{C}}(S)$.

Then, by Proposition 1, it is sufficient to prove that $h_{\mathcal{C}}(S)$ is accepted by some rational-multiset automaton.

Let us denote $\mathsf{ACU}(S)$ the ACU-closure of S. By Proposition 2, $h_{\mathcal{C}}(\mathsf{ACU}(S))$ is accepted by some rational-multiset automaton. We conclude easily using that $h_{\mathcal{C}}(S) = h_{\mathcal{C}}(\mathsf{ACU}(S))$.

6.2 An Algebraic Characterization of PMSO Definability

We are going to define now an algebraic recognizability criteria for the logic PMSO. Recalling that $\mathcal{C}$ is the algebra of terms built over the signature Σ, it is obvious to see that the notion of $\mathcal{C}$-recognizability is the same as the one defined by "classical" tree automata [3] for ranked trees written over the signature Σ (*ie* for Σ-terms): the set of states is the domain of the finite Σ-algebra $\mathcal{A}$, the interpretation of the function symbols from Σ in $\mathcal{A}$ provides the transition rules (which are bottom-up deterministic) and D is the set of final states.

We define weak $\mathcal{T}$-recognizability for unranked and unordered trees as follows:

Definition 10. *A tree language T is weakly $\mathcal{T}$-recognizable iff there exists some $\mathcal{C}$-recognizable set of Σ-terms M such that $T = h_{\mathcal{C}}(M)$.*

Intuitively, we can consider Σ-terms as representatives for trees and $h_{\mathcal{C}}$ as the mapping associating with each Σ-term the tree it represents. However, $h_{\mathcal{C}}$ is not injective, *ie* a single tree may have several representatives (actually, countably many). The intuition of weak $\mathcal{T}$-recognizability is to consider recognizability for the representatives (*ie* the Σ-terms) instead of the trees themselves. This notion is therefore different from $\mathcal{T}$-recognizability as $\mathcal{T}$-recognizability requires all the representatives of some tree to be recognized (see Proposition 6).

Theorem 5. *A set of trees T is PMSO-definable iff T is weakly $\mathcal{T}$-recognizable.*

Sketch of proof. By definition, T is weakly $\mathcal{T}$-recognizable iff there exists some $\mathcal{C}$-recognizable set of Σ-terms M such that $T = h_{\mathcal{C}}(M)$. By Proposition 5, this is equivalent to the existence of some equational language M over the algebra $\mathcal{C}$ such that $T = h_{\mathcal{C}}(M)$. Using Proposition 4, this latter holds iff T is an equational language over the algebra $\mathcal{T}$. Finally, by Theorem 4, this amounts to have T PMSO-definable.

7 New Characterizations for CMSO Definable Sets

In this section we reformulate CMSO definability first in terms of $\mathcal{C}$-recognizability and then by a restricted subclass of Presburger tree automata.

7.1 CMSO-Definability and $\mathcal{C}$-recognizability

Proposition 6. *For any set of trees T, T is CMSO-definable iff the set of Σ-terms $h_{\mathcal{C}}^{-1}(T)$ is $\mathcal{C}$-recognizable.*

Proof. Immediate from Theorem 2 and Proposition 4.4 from [6] stating that T is $\mathcal{T}$-recognizable iff $h_{\mathcal{C}}^{-1}(T)$ is $\mathcal{C}$-recognizable.

7.2 CMSO-Definability and Presburger Tree Automata

Definition 11. *A* unary Presburger tree automaton *is a PTA* (Λ, Q, F, δ) *such that* $F \in \mathcal{F}^1_{\mathcal{U}}(\{p \leq p', Div_k(p)\})$ *and for all* $q \in Q$ *and all* $a \in \Lambda$, $\delta(q, a)$ *belongs to* $\mathcal{F}^1_{\mathcal{U}}(\{p \leq p', Div_k(p)\})$.

Lemma 2. *Let* N *be a subset of* $\mathbb{N}^l$ *and* $A = (a_1, \ldots, a_l)$ *be some alphabet. Then* N *is unary ordering-definable iff* $\pi_A^{-1}(N) \in Rec(\mathbb{M}(A))$.

Proof. Courcelle showed in [6] that $\pi_A^{-1}(N) \in Rec(\mathbb{M}(A))$ iff N is a finite union of Cartesian products of l ultimately periodic sets of naturals, *ie* N is a finite union of sets of the form $B_1 \times \ldots \times B_l$ where for each i, $B_i = \{b + \alpha p \mid \alpha \in \mathbb{N}\}$ for some $b, p \in \mathbb{N}$. We just prove then that N is unary ordering-definable iff N is a finite union of Cartesian products of l ultimately periodic sets of naturals

Then, as for Presburger tree automata and PMSO, we have

Proposition 7. *For any set of trees* T, T *is CMSO-definable iff there exists some unary Presburger tree automaton* A *that accepts* T.

Proof. Straightforward using Theorem 3 and Lemma 2.

8 Some Characterizations for MSO Definable Sets

In this section, we investigate sets of trees definable by means of MSO sentences; Mainly, we are going to study how restrictions over formalisms used to characterize CMSO or PMSO can be put.

8.1 MSO-Definability and Presburger Tree Automata

Definition 12. *A* unary ordering tree automaton *is a PTA* (Λ, Q, F, δ) *such that* $F \in \mathcal{F}^1_{\mathcal{U}}(\{p \leq p'\})$ *and for all* $q \in Q$ *and all* $a \in \Lambda$, $\delta(q, a)$ *belongs to* $\mathcal{F}^1_{\mathcal{U}}(\{p \leq p'\})$.

Proposition 8. *For any set of trees* T, T *is MSO-definable iff there exists some unary ordering tree automaton* A *that accepts* T.

Sketch of proof. The proof is rather standard. We show first that the existence of an accepting run for a tree can be expressed by some MSO sentence. For the converse, we show closure of the unary ordering tree automaton under union, complementation (by computing first a deterministic and complete automaton) and relabeling morphism. Then, we build such an automaton inductively over the structure of the MSO formula.

8.2 MSO-Definability and Aperiodic-Recognizable Tree Automata

Definition 13. *A multiset language $L \in \mathbb{M}(A)$ is aperiodically recognizable if there exists a monoid morphism h from $(L, \uplus, \{\varnothing\})$ to a finite aperiodic*[7] *monoid $(D, +, \iota)$ and a finite subset D' of D such that $L = h^{-1}(D')$.*

We denote $ApRec(\mathbb{M}(A))$ the set of aperiodically recognizable multiset languages.

Definition 14. *An aperiodic-recognizable multiset tree automaton is a rational-multiset tree automaton (Λ, Q, F, δ) such that $F \in ApRec(\mathbb{M}(Q))$ and for all q in Q and a in Λ, $\delta(q, a) \in ApRec(\mathbb{M}(Q))$.*

Lemma 3. *Let N be a subset of $\mathbb{N}^l$ and $A = (a_1, \ldots, a_l)$ be some alphabet, N is unary ordering-definable iff $\pi_A^{-1}(N) \in ApRec(\mathbb{M}(A))$.*

Sketch of proof. We prove first that N is unary ordering-definable iff N is a finite union of Cartesian products of l ultimately periodic sets of naturals with periods in $\{0, 1\}$, *ie* N is a finite union of sets of the form $B_1 \times \ldots \times B_l$ where for each i, $B_i = \{b + \alpha p \mid \alpha \in \mathbb{N}\}$ for some $b \in \mathbb{N}$ and $p \in \{0, 1\}$. Then, we use a result from [9] stating that N is a finite union of Cartesian products of l ultimately periodic sets of naturals with periods in $\{0, 1\}$ iff N is a star-free subset of $\mathbb{N}^l$, *ie* N can be obtained from finite subsets of $\mathbb{N}^l$ using sum $+$ and Boolean operations (union, intersection, complement). Finally, we can conclude using that $(\mathbb{N}^l, +)$ is isomorphic to $(\mathbb{M}(A), \uplus)$ and that over commutative monoids, star-free languages are precisely the recognizable and aperiodic ones [10].

Theorem 6. *For any set of trees T, T is MSO-definable iff there exists some aperiodic-recognizable multiset automaton A that accepts T.*

Proof. Straightforward from Proposition 8 and Lemma 3.

8.3 An Algebraic Characterization of MSO Definability

We relate here MSO definability and algebraic $\mathcal{T}$-recognizability.

Definition 15. *A tree language T is aperiodically $\mathcal{T}$-recognizable iff there exists a finite Σ-algebra $\mathcal{A}$ with domain $dom(\mathcal{A})$ such that $(dom(\mathcal{A}), |^{\mathcal{A}}, \mathbf{0}^{\mathcal{A}})$ is an aperiodic and commutative monoid, h is a homomorphism from $\mathcal{T}$ to $\mathcal{A}$ and D is a finite subset of $dom(\mathcal{A})$ such that $T = h^{-1}(D)$.*

Theorem 7. *For any set of trees T, T is MSO-definable iff T is aperiodically $\mathcal{T}$-recognizable.*

[7] We recall that a monoid $(S, .)$ is said to to be aperiodic if for all $s \in S$, there exists some natural n such that $s^n = s^{n+1}$ where $s^1 = s$ and $s^{k+1} = s^k.s$.

References

1. J. Carme, J. Niehren, and M. Tommasi. Querying Unranked Trees with Stepwise Tree Automata. In *International Conference on Rewriting Techniques and Applications*, volume 3091 of *LNCS*, pages 105–118. Springer, 2004.
2. T. Colcombet. Rewriting in the partial algebra of typed terms modulo AC. In *Electronic Notes in Theoretical Computer Science*, volume 68. Elsevier Science Publishers, 2002.
3. H. Comon, M. Dauchet, R. Gilleron, F. Jacquemard, D. Lugiez, S. Tison, and M. Tommasi. Tree Automata Techniques and Applications. Available on: http://www.grappa.univ-lille3.fr/tata, 1997. release October, 1rst 2002.
4. B. Courcelle. The Monadic Second-Order Logic of Graphs. I. Recognizable Sets of Finite Graphs. *IC*, 85(1):12–75, 1990.
5. B. Courcelle. The monadic second order logic of graphs VI: on several representations of graphs by relational structures. *Discrete Applied Mathematics*, 54(2-3):117–149, 1994.
6. B. Courcelle. Basic notions of universal algebra for language theory and graph grammars. *Theoretical Computer Science*, 163:1–54, 1996.
7. S. Dal-Zilio and D. Lugiez. XML Schema, Tree Logic and Sheaves Automata. In *Rewriting Techniques and Applications, 14th International Conference, RTA 2003*, volume 2706 of *LNCS*, pages 246–263. Springer, 2003.
8. S. Dal-Zilio, D. Lugiez, and C. Meyssonnier. A logic you can count on. In *31st Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, 2004.
9. S. Gaubert and A. Giua. Petri net languages and infinite subsets of $\mathbb{N}^m$. *Journal of Computer System Sciences*, 59(3):373–391, 1999.
10. G. Guaiana, A. Restivo, and S. Salemi. On Aperiodic Trace Languages. In *STACS 91, 8th Annual Symposium on Theoretical Aspects of Computer*, volume 480 of *LNCS*, pages 76–88. Springer, 1991.
11. F. Klaedtke and H. Rueß. Monadic Second-Order Logics with Cardinalities. In *30th International Colloquium on Automata, Languages and Programming, ICALP 2003*, volume 2719 of *Lecture Notes in Computer Science*. Springer Verlag, 2003.
12. J. Mezei and J.B. Wright. Algebraic automata and context-free sets. *Information and Control*, 11(2-3):3–29, 1967.
13. F. Neven and T. Schwentick. Query automata over finite trees. *Theoretical Computer Science*, 275(1–2):633–674, March 2002.
14. J. Niehren and A. Podelski. Feature Automata and Recognizable Sets of Feature Trees. In *Theory and Practice of Software Development, International Joint Conference CAAP/FASE/TOOLS*, volume 668 of *LNCS*, pages 356–375. Springer, 1993.
15. H. Ohsaki. Beyond Regularity: Equational Tree Automata for Associative and Commutative Theories. In *Proceedings of 15th International Conference of the European Association for Computer Science Logic - CSL 2001*, volume 2142 of *LNCS*, pages 539–553. Springer, 2001.
16. H. Ohsaki and T. Takai. Decidability and Closure Properties of Equational Tree Languages. In *Proceedings of 13th International Conference on Rewriting Techniques and Applications*, volume 2378 of *LNCS*, pages 114–128. Springer, 2002.
17. R. J. Parikh. On context-free languages. *Journal of the ACM*, 13(4):570–581, 1966.
18. H. Seidl, T. Schwentick, and A. Muscholl. Numerical Document Queries. In *Twenty-Second ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pages 155–166. ACM, 2003.
19. J. W. Thatcher and J. B. Wright. Generalized finite automata with an application to a decision problem of second-order logic. *Mathematical System Theory*, 2:57–82, 1968.
20. K. N. Verma. Two-Way Equational Tree Automata for AC-like Theories: Decidability and Closure Properties. In *Proceedings of 14th International Conference on Rewriting Techniques and Applications*, volume 2706 of *LNCS*, pages 180–197. Springer, 2003.