A Complexity Approach to Tree Algebras: the Bounded Case

Arthur Jaquard

joint work with Thomas Colcombet

Université de Paris, CNRS, IRIF, F-75006, Paris, France

ICALP 2021 (Track B)

INSTITUT
DE RECHERCHE
EN INFORMATIQUE
FONDAMENTALE

Université de Paris

Algebras are used to characterize classes of languages

Algebras are used to characterize classes of languages

Finite words
Monoids, semigroups

Algebras are used to characterize classes of languages

Schützenberger, 1965

A regular language L is star-free if and only if its syntactic monoid is aperiodic.

Finite words
Monoids, semigroups

Algebras are used to characterize classes of languages

Schützenberger, 1965

A regular language L is star-free if and only if its syntactic monoid is aperiodic.

Finite words

Monoids, semigroups

Infinite words

Wilke algebras, ω-semigroups, $\circ$-algebras…

Algebras are used to characterize classes of languages

Finite words
Monoids, semigroups
Infinite words
Wilke algebras, ω-semigroups, $\circ$-algebras...
Trees
Deterministic automata, Preclones, Hyperclones, Operads,...
Graphs
HR-algebras, VR-algebras

Schützenberger, 1965

A regular language L is star-free if and only if its syntactic monoid is aperiodic.

Algebras are used to characterize classes of languages

Schützenberger, 1965

A regular language L is star-free if and only if its syntactic monoid is aperiodic.

Finite words
Monoids, semigroups
Infinite words
Wilke algebras, ω-semigroups, $\circ$-algebras...
Trees
Deterministic automata, Preclones, Hyperclones, Operads,...
Graphs
HR-algebras, VR-algebras

Infinitely sorted algebras

$(A_n)_{n\in\mathbb{N}}$

$(A_X)_{X \text{ finite}}$

Algebras are used to characterize classes of languages

Schützenberger, 1965

A regular language L is star-free if and only if its syntactic monoid is aperiodic.

Finite words
Monoids, semigroups

Infinite words
Wilke algebras, ω-semigroups, $\circ$-algebras...

Trees
Deterministic automata, Preclones, Hyperclones, Operads,...

Graphs
HR-algebras, VR-algebras

Infinitely sorted algebras

$(A_n)_{n\in\mathbb{N}}$

$(A_X)_{X \text{ finite}}$

Problem: Hard to derive characterizations with infinitely sorted algebras

Algebras are used to characterize classes of languages

Schützenberger, 1965

A regular language L is star-free if and only if its syntactic monoid is aperiodic.

Finite words
Monoids, semigroups
Infinite words
Wilke algebras, ω-semigroups, $\circ$-algebras...
Trees
Deterministic automata, Preclones, Hyperclones, Operads,...
Graphs
HR-algebras, VR-algebras

Infinitely sorted algebras

$(A_n)_{n\in\mathbb{N}}$

$(A_X)_{X \text{ finite}}$

Problem: Hard to derive characterizations with infinitely sorted algebras

Objective: characterize classes that can be naturally defined using infinitely sorted algebras

Infinitely sorted tree algebras: FT_Σ-algebras

Let Σ be a ranked alphabet. The free FT_Σ-algebra has as carrier $(T_X)_{X \text{ finite}}$ where the X's are finite sets of variables.

$$T_X = \{\text{trees in which all the variables of } X \text{ appear on the leaves}\}$$

Infinitely sorted tree algebras: FT_Σ-algebras

Let Σ be a ranked alphabet. The free FT_Σ-algebra has as carrier $(T_X)_{X \text{ finite}}$ where the X's are finite sets of variables.

$$T_X = \{\text{trees in which all the variables of } X \text{ appear on the leaves}\}$$

Objects

$$\begin{array}{ccc} & a & \\ / & & \backslash \\ b & & c \end{array} \in T_\emptyset \qquad \begin{array}{ccc} & a & \\ / & & \backslash \\ x & & x \end{array} \in T_{\{x\}}$$

$$\begin{array}{ccc} & a & \\ / & & \backslash \\ x & & y \end{array} \in T_{\{x,y\}}$$

Infinitely sorted tree algebras: FT_Σ-algebras

Let Σ be a ranked alphabet. The free FT_Σ-algebra has as carrier $(T_X)_{X \text{ finite}}$ where the X's are finite sets of variables.

$$T_X = \{\text{trees in which all the variables of } X \text{ appear on the leaves}\}$$

Objects

```
  a                a
 / \ ∈ T_∅        / \ ∈ T_{x}
b   c            x   x

        a
       / \ ∈ T_{x,y}
      x   y
```

Substitution

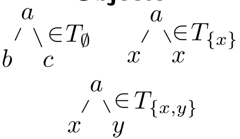
```
  a          a          a
 / \   ·x   / \   =    / \
x   y      b   c      a   y
                     / \
                    b   c
```

Infinitely sorted tree algebras: FT_Σ-algebras

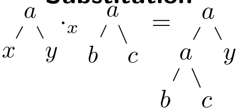
Let Σ be a ranked alphabet. The free FT_Σ-algebra has as carrier $(T_X)_{X \text{ finite}}$ where the X's are finite sets of variables.

$$T_X = \{\text{trees in which all the variables of } X \text{ appear on the leaves}\}$$

Objects

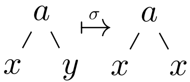


Substitution



Renaming

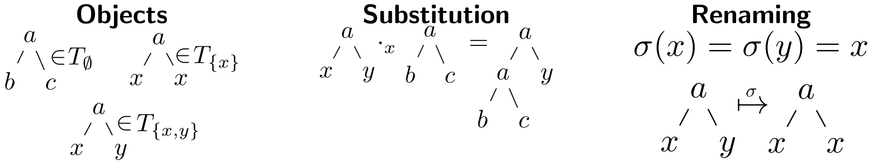
$$\sigma(x) = \sigma(y) = x$$



Infinitely sorted tree algebras: FT_Σ-algebras

Let Σ be a ranked alphabet. The free FT_Σ-algebra has as carrier $(T_X)_{X \text{ finite}}$ where the X's are finite sets of variables.

$$T_X = \{\text{trees in which all the variables of } X \text{ appear on the leaves}\}$$



Definition (Finite Tree algebras)

A finite FT_Σ-algebra $\mathcal{A}$ consists of an infinite series of finite carrier sets A_X indexed by finite sets of variables X, together with operations:

Constants. $a(x_0, \ldots, x_{n-1})^{\mathcal{A}} \in A_{\{x_0,\ldots,x_{n-1}\}}$ for all $a \in \Sigma_n$ and variables x_i,

Substitution. $\cdot_x^{\mathcal{A}} \colon A_X \times A_Y \to A_{X \setminus \{x\} \cup Y}$ for all finite X, Y and $x \in X$,

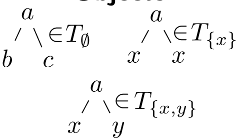
Renaming. $\mathrm{rename}^{\mathcal{A}}[\sigma] \colon A_X \to A_Y$ for all surjective maps $\sigma \colon X \to Y$.

Infinitely sorted tree algebras: FT_Σ-algebras

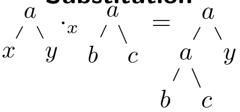
Let Σ be a ranked alphabet. The free FT_Σ-algebra has as carrier $(T_X)_{X \text{ finite}}$ where the X's are finite sets of variables.

$$T_X = \{\text{trees in which all the variables of } X \text{ appear on the leaves}\}$$

Objects

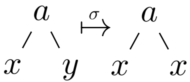


Substitution



Renaming

$$\sigma(x) = \sigma(y) = x$$



Definition (Finite Tree algebras)

A finite FT_Σ-algebra $\mathcal{A}$ consists of an infinite series of finite carrier sets A_X indexed by finite sets of variables X, together with operations:

Constants. $a(x_0, \ldots, x_{n-1})^{\mathcal{A}} \in A_{\{x_0,\ldots,x_{n-1}\}}$ for all $a \in \Sigma_n$ and variables x_i,

Substitution. $\cdot_x^{\mathcal{A}} \colon A_X \times A_Y \to A_{X \setminus \{x\} \cup Y}$ for all finite X, Y and $x \in X$,

Renaming. $\mathrm{rename}^{\mathcal{A}}[\sigma] \colon A_X \to A_Y$ for all surjective maps $\sigma \colon X \to Y$.

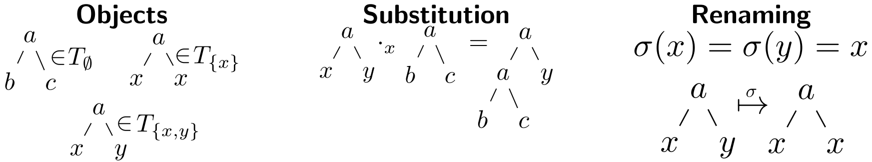
Identities? $\quad a(x, y) \cdot_y b \quad a(x, z) \cdot_z b$

We also define morphisms, congruences...

Infinitely sorted tree algebras: FT_Σ-algebras

Let Σ be a ranked alphabet. The free FT_Σ-algebra has as carrier $(T_X)_{X \text{ finite}}$ where the X's are finite sets of variables.

$$T_X = \{\text{trees in which all the variables of } X \text{ appear on the leaves}\}$$



Definition (Finite Tree algebras)

A finite FT_Σ-algebra $\mathcal{A}$ consists of an infinite series of finite carrier sets A_X indexed by finite sets of variables X, together with operations:

Constants. $a(x_0, \ldots, x_{n-1})^{\mathcal{A}} \in A_{\{x_0,\ldots,x_{n-1}\}}$ for all $a \in \Sigma_n$ and variables x_i,
Substitution. $\cdot_x^{\mathcal{A}} \colon A_X \times A_Y \to A_{X \setminus \{x\} \cup Y}$ for all finite X, Y and $x \in X$,
Renaming. $\mathrm{rename}^{\mathcal{A}}[\sigma] \colon A_X \to A_Y$ for all surjective maps $\sigma \colon X \to Y$.

Given a finite FT_Σ-algebra $\mathcal{A}$, there is a unique morphism from the free algebra to $\mathcal{A}$. It is called the evaluation morphism of $\mathcal{A}$.

Languages and the size of the algebra

Definition (Language recognized by an algebra)

A language L of finite trees over Σ is recognized by a finite FT_Σ-algebra $\mathcal{A}$ if there is a set $P \subseteq A_\emptyset$ such that $L = \alpha^{-1}(P)$ in which α is the evaluation morphism of $\mathcal{A}$.

Example $L =$ The language of all trees that only contain a's and b's

Languages and the size of the algebra

Definition (Language recognized by an algebra)

A language L of finite trees over Σ is recognized by a finite FT_Σ-algebra $\mathcal{A}$ if there is a set $P \subseteq A_\emptyset$ such that $L = \alpha^{-1}(P)$ in which α is the evaluation morphism of $\mathcal{A}$.

Example $L =$ The language of all trees that only contain a's and b's

$$\begin{array}{ccc} & a & \\ / & & \backslash \\ x & & x \end{array} \overset{\alpha}{\mapsto} \{a\} \in A_{\{x\}}$$

Languages and the size of the algebra

Definition (Language recognized by an algebra)

A language L of finite trees over Σ is recognized by a finite FT_Σ-algebra $\mathcal{A}$ if there is a set $P \subseteq A_\emptyset$ such that $L = \alpha^{-1}(P)$ in which α is the evaluation morphism of $\mathcal{A}$.

Example $L =$ The language of all trees that only contain a's and b's

```
   a
  / \   ↦α  {a} ∈ A_{x}
 x   x
```

$$A_X = 2^\Sigma \text{ for all } X$$

$$A \cdot_x B = A \cup B$$

Languages and the size of the algebra

Definition (Language recognized by an algebra)

A language L of finite trees over Σ is recognized by a finite FT_Σ-algebra $\mathcal{A}$ if there is a set $P \subseteq A_\emptyset$ such that $L = \alpha^{-1}(P)$ in which α is the evaluation morphism of $\mathcal{A}$.

Example $L =$ The language of all trees that only contain a's and b's

$$\begin{array}{c} a \\ / \quad \backslash \\ x \qquad x \end{array} \overset{\alpha}{\mapsto} \{a\} \in A_{\{x\}}$$

$$A_X = 2^\Sigma \text{ for all } X$$

$$A \cdot_x B = A \cup B$$

$|A_X| = 2^{|\Sigma|}$ is bounded (it does not depend on $|X|$).

Languages and the size of the algebra

Definition (Language recognized by an algebra)

A language L of finite trees over Σ is recognized by a finite FT_Σ-algebra $\mathcal{A}$ if there is a set $P \subseteq A_\emptyset$ such that $L = \alpha^{-1}(P)$ in which α is the evaluation morphism of $\mathcal{A}$.

Example $L =$ The language of all trees that only contain a's and b's

$$\begin{array}{c} a \\ / \quad \backslash \\ x \quad\; x \end{array} \overset{\alpha}{\mapsto} \{a\} \in A_{\{x\}}$$

$$A_X = 2^\Sigma \text{ for all } X$$

$$A \cdot_x B = A \cup B$$

$|A_X| = 2^{|\Sigma|}$ is bounded (it does not depend on $|X|$).

Example $L =$ trees without b's on the leftmost branch

Languages and the size of the algebra

Definition (Language recognized by an algebra)

A language L of finite trees over Σ is recognized by a finite FT_Σ-algebra $\mathcal{A}$ if there is a set $P \subseteq A_\emptyset$ such that $L = \alpha^{-1}(P)$ in which α is the evaluation morphism of $\mathcal{A}$.

Example $L =$ The language of all trees that only contain a's and b's

```
   a
  / \   --α-->  {a} ∈ A_{x}
 x   x
```

$A_X = 2^\Sigma$ for all X

$A \cdot_x B = A \cup B$

$|A_X| = 2^{|\Sigma|}$ is bounded (it does not depend on $|X|$).

Example $L =$ trees without b's on the leftmost branch

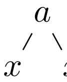
```
        a                                 a
      /   \                             /   \
     b     a     --α-->  {a, b, c}     a     a     --α-->  ({a}, x),
    / \   / \                         / \   / \
   c   c y   y                       x   c y   y
```

Languages and the size of the algebra

Definition (Language recognized by an algebra)

A language L of finite trees over Σ is recognized by a finite FT_Σ-algebra $\mathcal{A}$ if there is a set $P \subseteq A_\emptyset$ such that $L = \alpha^{-1}(P)$ in which α is the evaluation morphism of $\mathcal{A}$.

Example $L =$ The language of all trees that only contain a's and b's



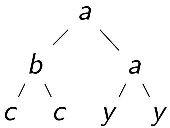
$\overset{\alpha}{\mapsto} \{a\} \in A_{\{x\}}$

$A_X = 2^\Sigma$ for all X

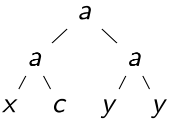
$A \cdot_x B = A \cup B$

$|A_X| = 2^{|\Sigma|}$ is bounded (it does not depend on $|X|$).

Example $L =$ trees without b's on the leftmost branch



$\overset{\alpha}{\mapsto} \{a, b, c\}$



$\overset{\alpha}{\mapsto} (\{a\}, x)$,

$A_X = 2^\Sigma \uplus (2^\Sigma \times X)$

$|A_X| = 2^{|\Sigma|} + 2^{|\Sigma|}|X|$ is linear in $|X|$.

Complexity

Definition (Complexity)

Given a finite FT_{Σ}-algebra $\mathcal{A}$ with carrier

$$(A_X)_{X \text{ finite}}, \text{ all } A_X \text{ finite}$$

its complexity map is $c_{\mathcal{A}}(|X|) = |A_X|$. $(|X| = |Y| \text{ implies } |A_X| = |A_Y|)$

Complexity

Definition (Complexity)

Given a finite FT_Σ-algebra $\mathcal{A}$ with carrier

$$(A_X)_{X \text{ finite}}, \text{ all } A_X \text{ finite}$$

its complexity map is $c_\mathcal{A}(|X|) = |A_X|$. $(|X| = |Y| \text{ implies } |A_X| = |A_Y|)$

$$A_X = 2^\Sigma$$

$$|A_X| = 2^{|\Sigma|}$$

Bounded complexity

Complexity

Definition (Complexity)

Given a finite FT_Σ-algebra $\mathcal{A}$ with carrier

$$(A_X)_{X \text{ finite}}, \text{ all } A_X \text{ finite}$$

its complexity map is $c_{\mathcal{A}}(|X|) = |A_X|$. $(|X| = |Y| \text{ implies } |A_X| = |A_Y|)$

$$A_X = 2^\Sigma$$

$$|A_X| = 2^{|\Sigma|}$$

Bounded complexity

$$A_X = 2^\Sigma \uplus (2^\Sigma \times X)$$

$$|A_X| = 2^{|\Sigma|} + 2^{|\Sigma|}|X|$$

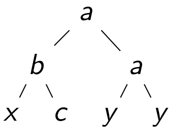
Linear complexity

Example: The language of all trees with at least a b on every branch

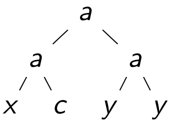
$L =$ trees with at least a b on every branch

Example: The language of all trees with at least a *b* on every branch

$L =$ trees with at least a b on every branch



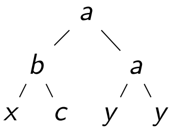
$\overset{\alpha}{\mapsto} (\{y\}, \top)$



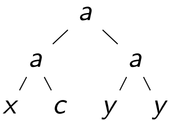
$\overset{\alpha}{\mapsto} (\{x, y\}, \bot)$

Example: The language of all trees with at least a b on every branch

$L =$ trees with at least a b on every branch



$\overset{\alpha}{\mapsto} (\{y\}, \top)$



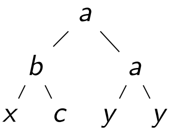
$\overset{\alpha}{\mapsto} (\{x, y\}, \bot)$

$$A_X = 2^X \times \{\top, \bot\}$$

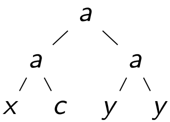
$$|A_X| = 2^{|X|} \times 2$$

Example: The language of all trees with at least a b on every branch

L = trees with at least a b on every branch



$\overset{\alpha}{\mapsto} (\{y\}, \top)$



$\overset{\alpha}{\mapsto} (\{x, y\}, \bot)$

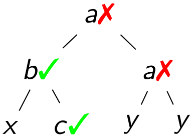
$$A_X = 2^X \times \{\top, \bot\}$$

$$|A_X| = 2^{|X|} \times 2$$

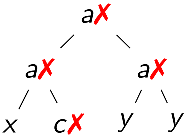
This algebra has exponential complexity.

Example: The language of all trees with at least a b on every branch

L = trees with at least a b on every branch



$\overset{\alpha}{\mapsto} (\{y\}, \top)$



$\overset{\alpha}{\mapsto} (\{x, y\}, \bot)$

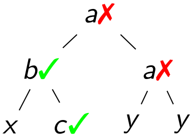
$$A_X = 2^X \times \{\top, \bot\}$$

$$|A_X| = 2^{|X|} \times 2$$

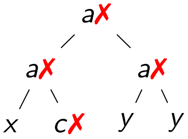
This algebra has exponential complexity.

Example: The language of all trees with at least a b on every branch

L = trees with at least a b on every branch



$\stackrel{\alpha}{\mapsto} (\{y\}, \top)$



$\stackrel{\alpha}{\mapsto} (\{x, y\}, \bot)$

$$A_X = 2^X \times \{\top, \bot\}$$

$$|A_X| = 2^{|X|} \times 2$$

This algebra has exponential complexity.

Languages recognized by top-down deterministic automata

All languages recognized by top-down deterministic automata are recognized by FT_Σ-algebras of exponential complexity.

What complexity means

Bounded complexity

The algebra does not remember anything about the variables.
$A_X \rightsquigarrow$ the variables that appear in the tree are exactly X.

What complexity means

Bounded complexity

The algebra does not remember anything about the variables.

$A_X \rightsquigarrow$ the variables that appear in the tree are exactly X.

Polynomial complexity

$A_X = X^k \rightsquigarrow k$ variables (e.g. k branches)

What complexity means

Bounded complexity
The algebra does not remember anything about the variables.
$A_X \rightsquigarrow$ the variables that appear in the tree are exactly X.
Polynomial complexity
$A_X = X^k \rightsquigarrow k$ variables (e.g. k branches)
Exponential complexity
$A_X = k^X \rightsquigarrow$ a function from X to k (e.g. a set of variables when $k = 2$, or modulo counting if $k = \mathbb{Z}/q\mathbb{Z}$)

What complexity means

Bounded complexity
The algebra does not remember anything about the variables.
$A_X \rightsquigarrow$ the variables that appear in the tree are exactly X.
Polynomial complexity
$A_X = X^k \rightsquigarrow k$ variables (e.g. k branches)
Exponential complexity
$A_X = k^X \rightsquigarrow$ a function from X to k (e.g. a set of variables when $k = 2$, or modulo counting if $k = \mathbb{Z}/q\mathbb{Z}$)
Doubly exponential complexity

Regular languages

A top-down nondeterministic automaton can be transformed into a FT_Σ-algebras of doubly-exponential complexity that recognizes the same language.
Conversely, any language recognized by a finite FT_Σ-algebra is regular.

Characterizing bounded complexity

Is complexity meaningful?

Characterizing bounded complexity

Is complexity meaningful?
What are the languages recognized by FT_Σ-algebras of bounded complexity?

Characterizing bounded complexity

Is complexity meaningful?
What are the languages recognized by FT_Σ-algebras of bounded complexity?

Consider for all X the group morphism induced by renaming

$$\begin{aligned}\varphi_X \colon \mathbf{Sym}(X) &\to \mathbf{Sym}(A_X)\\ \sigma &\mapsto \mathrm{rename}^{\mathcal{A}}[\sigma]\end{aligned}$$

Invariance under permutations

A finite syntactic FT_Σ-algebra is of bounded complexity if and only if for all sufficiently large finite set of variables X, $\mathrm{Ker}(\varphi_X) = \mathbf{Sym}(X)$.

Main result

Characterization theorem

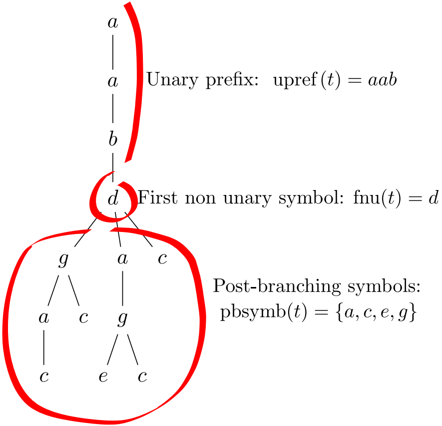
A language of finite trees is recognized by an FT_Σ-algebra of bounded complexity if and only if it is a Boolean combination of languages of the following kinds:

a. The language of finite trees with unary prefix in a given regular language of words $L \subseteq \Sigma_1^*$.

b. The language of finite trees with first non unary symbol b for a fixed non unary symbol b.

c. The language of finite trees with post-branching symbols B, for $B \subseteq \Sigma$.

d. A regular language K of bounded branching.



Bounded branching: $\exists k$ all trees in K have at most k branches

Conclusion

Complexity map: $c_{\mathcal{A}}(|X|) = |A_X|$

Bounded complexity ✓

Characterization theorem

A language of finite trees is recognized by an FT_Σ-algebra of bounded complexity if and only if it is a Boolean combination of languages of the following kinds:

a. The language of finite trees with unary prefix in a given regular language of words $L \subseteq \Sigma_1^*$.

b. The language of finite trees with first non unary symbol b for a fixed non unary symbol b.

c. The language of finite trees with post-branching symbols B, for $B \subseteq \Sigma$.

d. A regular language K of bounded branching.

Conclusion

Complexity map: $c_{\mathcal{A}}(|X|) = |A_X|$

Bounded complexity ✓

Polynomial complexity ?

Exponential complexity ?

Characterization theorem

A language of finite trees is recognized by an FT_{Σ}-algebra of bounded complexity if and only if it is a Boolean combination of languages of the following kinds:

a. The language of finite trees with unary prefix in a given regular language of words $L \subseteq \Sigma_1^*$.

b. The language of finite trees with first non unary symbol b for a fixed non unary symbol b.

c. The language of finite trees with post-branching symbols B, for $B \subseteq \Sigma$.

d. A regular language K of bounded branching.

Conclusion

Complexity map: $c_{\mathcal{A}}(|X|) = |A_X|$

Bounded complexity ✓

Polynomial complexity ?

Exponential complexity ?

Orbit complexity: renaming yields an action of $\mathbf{Sym}(X)$ over A_X.

$$c^{\circ}_{\mathcal{A}}(|X|) = |A_X/\mathbf{Sym}(X)|$$

Characterization theorem

A language of finite trees is recognized by an FT_{Σ}-algebra of bounded complexity if and only if it is a Boolean combination of languages of the following kinds:

a. The language of finite trees with unary prefix in a given regular language of words $L \subseteq \Sigma_1^*$.

b. The language of finite trees with first non unary symbol b for a fixed non unary symbol b.

c. The language of finite trees with post-branching symbols B, for $B \subseteq \Sigma$.

d. A regular language K of bounded branching.

Conclusion

Complexity map: $c_{\mathcal{A}}(|X|) = |A_X|$

Bounded complexity ✓

Polynomial complexity ?

Exponential complexity ?

Orbit complexity: renaming yields an action of $\mathbf{Sym}(X)$ over A_X.

$$c^{\circ}_{\mathcal{A}}(|X|) = |A_X/\mathbf{Sym}(X)|$$

Ongoing: polynomial complexity, bounded orbit complexity...

Characterization theorem

A language of finite trees is recognized by an FT_{Σ}-algebra of bounded complexity if and only if it is a Boolean combination of languages of the following kinds:

a. The language of finite trees with unary prefix in a given regular language of words $L \subseteq \Sigma_1^*$.

b. The language of finite trees with first non unary symbol b for a fixed non unary symbol b.

c. The language of finite trees with post-branching symbols B, for $B \subseteq \Sigma$.

d. A regular language K of bounded branching.

Conclusion

Complexity map: $c_{\mathcal{A}}(|X|) = |A_X|$

Bounded complexity ✓

Polynomial complexity ?

Exponential complexity ?

Orbit complexity: renaming yields an action of $\mathbf{Sym}(X)$ over A_X.

$$c^{\circ}_{\mathcal{A}}(|X|) = |A_X/\mathbf{Sym}(X)|$$

Ongoing: polynomial complexity, bounded orbit complexity...

A similar characterization of languages of infinite regular trees as Boolean combinations of a.-d. and other languages

Characterization theorem

A language of finite trees is recognized by an FT_Σ-algebra of bounded complexity if and only if it is a Boolean combination of languages of the following kinds:

a. The language of finite trees with unary prefix in a given regular language of words $L \subseteq \Sigma_1^*$.

b. The language of finite trees with first non unary symbol b for a fixed non unary symbol b.

c. The language of finite trees with post-branching symbols B, for $B \subseteq \Sigma$.

d. A regular language K of bounded branching.