Breadth-first signature of trees and rational languages

Victor Marsault,
joint work with Jacques Sakarovitch

CNRS / Telecom-ParisTech, Paris, France

Journées Montoises 2014, Nancy,
2014–09–23

NS = Numeration system

Prefix-closed Abstract Rational NS (Lecomte–Rigo 2001)

Built from an arbitrary prefix-closed rational language.

Dumont-Thomas NS (Dumont-Thomas, 1989)

Built from an arbitrary substitution.

NS = Numeration system

Prefix-closed Abstract Rational NS (Lecomte–Rigo 2001)

Built from an arbitrary prefix-closed rational language.

Dumont-Thomas NS (Dumont-Thomas, 1989)

Built from an arbitrary substitution.

Definition: Signature

Tree or language $\longmapsto$ infinite word

NS = Numeration system

Prefix-closed Abstract Rational NS (Lecomte–Rigo 2001)

Built from an arbitrary prefix-closed rational language.

Dumont-Thomas NS (Dumont-Thomas, 1989)

Built from an arbitrary substitution.

Definition: Signature

Tree or language $\longmapsto$ infinite word

In particular: Rational language $\longmapsto$ substitutive word

NS = Numeration system

Prefix-closed Abstract Rational NS (Lecomte–Rigo 2001)

Built from an arbitrary prefix-closed rational language.

Dumont-Thomas NS (Dumont-Thomas, 1989)

Built from an arbitrary substitution.

Definition: Signature

Tree or language $\longmapsto$ infinite word
In particular: Rational language $\longmapsto$ substitutive word

Theorem

Every DTNS is a prefix-closed ARNS.

Every prefix-closed ARNS is easily† convertible to a DTNS.

† Through a finite, letter-to-letter and pure sequential transducer.

Outline

We call *tree* a...

Directed graph which is

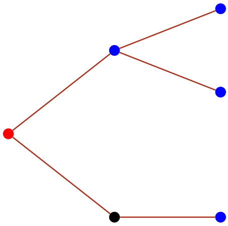
- **Rooted:** a node is called *the root* (leftmost in the figures)
- **Directed outward from the root:** there is a unique path from the root to every other node.
- **Ordered:** the children of every node are ordered

  (In the figures, lower children are smaller.)

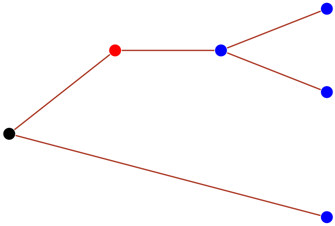
We call *tree* a...

Directed graph which is

- **Rooted:** a node is called *the root* (leftmost in the figures)
- **Directed outward from the root:** there is a unique path from the root to every other node.
- **Ordered:** the children of every node are ordered
  (In the figures, lower children are smaller.)



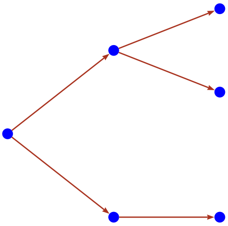
$\neq$



We call *tree* a...

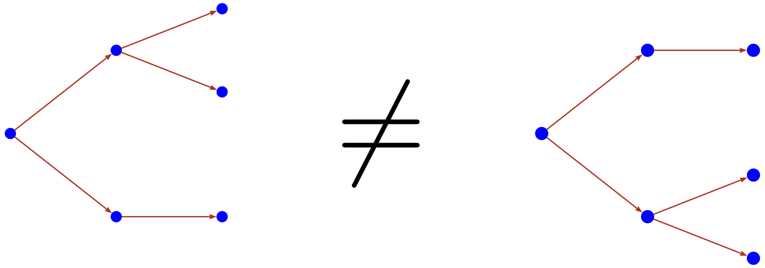
Directed graph which is

- **Rooted:** a node is called *the root* (leftmost in the figures)
- **Directed outward from the root:** there is a unique path from the root to every other node.
- **Ordered:** the children of every node are ordered
  (In the figures, lower children are smaller.)

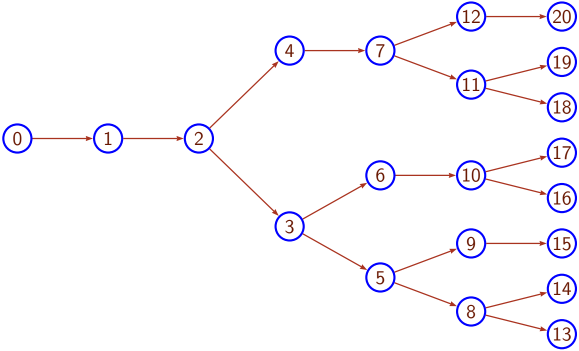


Directed graph which is

- **Rooted:** a node is called *the root* (leftmost in the figures)
- **Directed outward from the root:** there is a unique path from the root to every other node.
- **Ordered:** the children of every node are ordered (In the figures, lower children are smaller.)

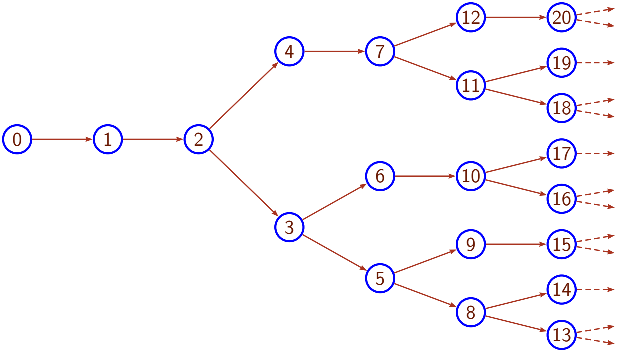


Every tree has a canonical breadth-first traversal



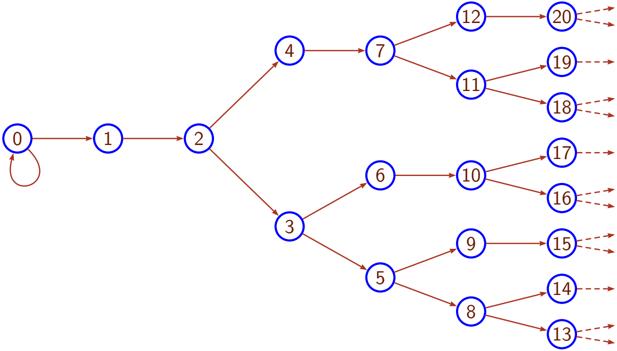
Two more features

- We consider infinite trees only.



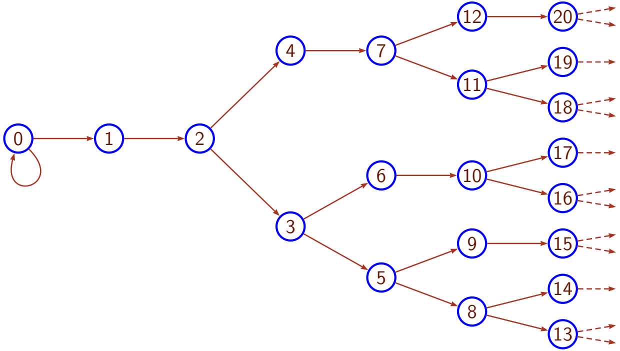
Two more features

- We consider infinite trees only.
- For convenience, there is loop on the root.



Definition

The **signature** of a tree is the sequence of the degree of the nodes taken in breadth-first order.

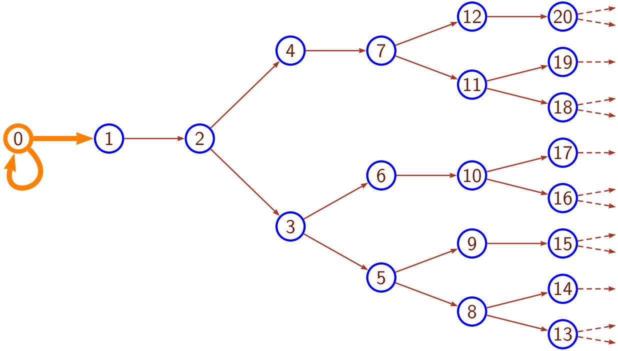


$\mathbf{s} =$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degree of the nodes taken in breadth-first order.

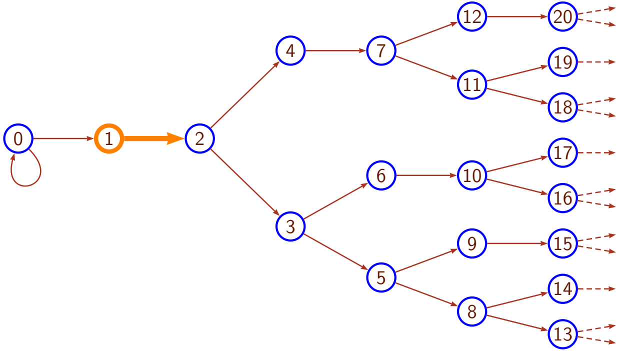


$\mathbf{s} = 2$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degree of the nodes taken in breadth-first order.

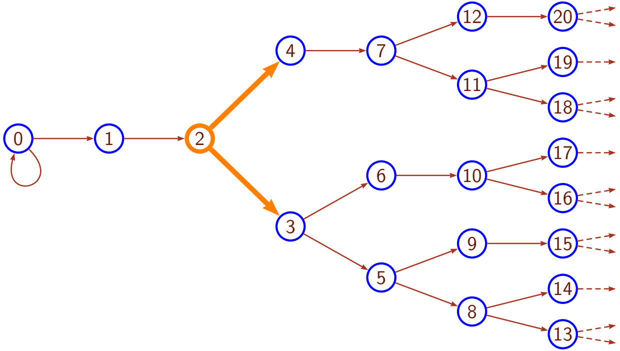


$$\mathbf{s} = 2\ 1$$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degree of the nodes taken in breadth-first order.

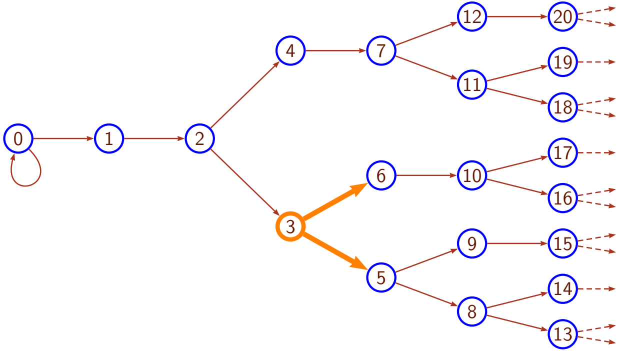


$\mathbf{s} = 2\ 1\ 2$

Signature of a tree

Definition

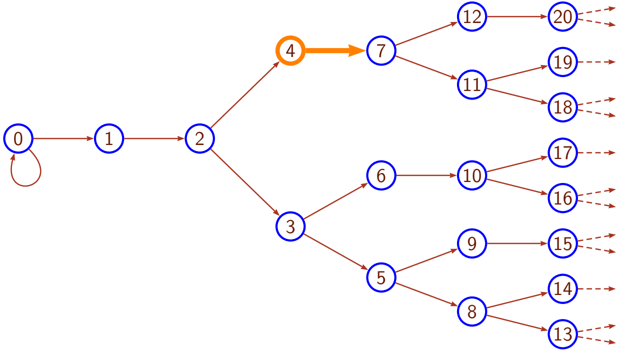
The **signature** of a tree is the sequence of the degree of the nodes taken in breadth-first order.



$\mathbf{s} = 2\ 1\ 2\ 2$

Definition

The **signature** of a tree is the sequence of the degree of the nodes taken in breadth-first order.

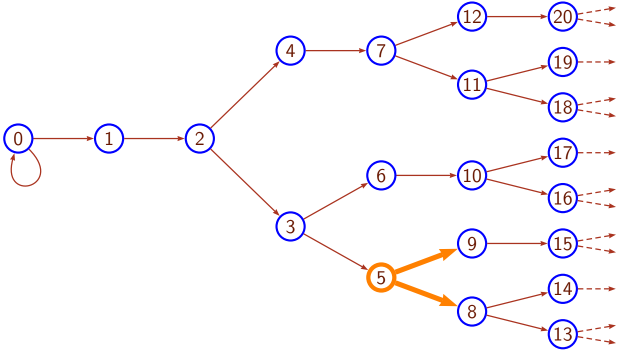


$$\mathbf{s} = 2\ 1\ 2\ 2\ 1$$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degree of the nodes taken in breadth-first order.

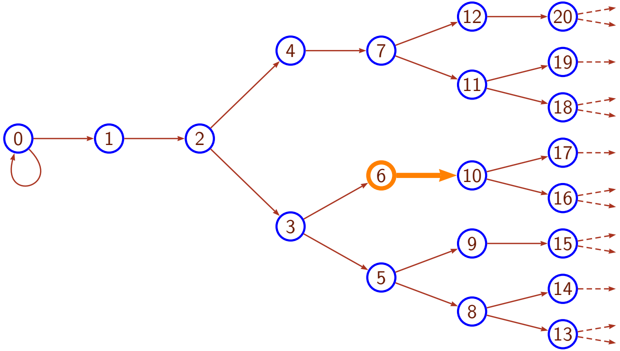


$$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2$$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degree of the nodes taken in breadth-first order.

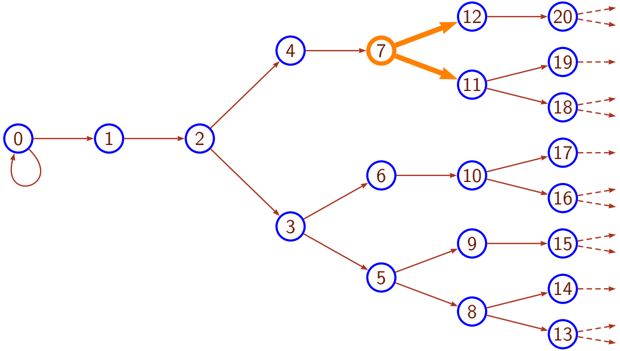


$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degree of the nodes taken in breadth-first order.

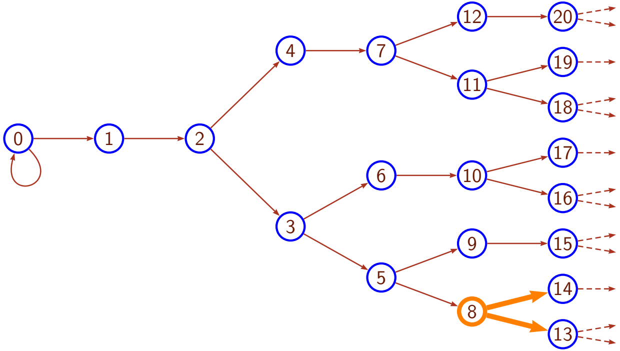


$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degree of the nodes taken in breadth-first order.

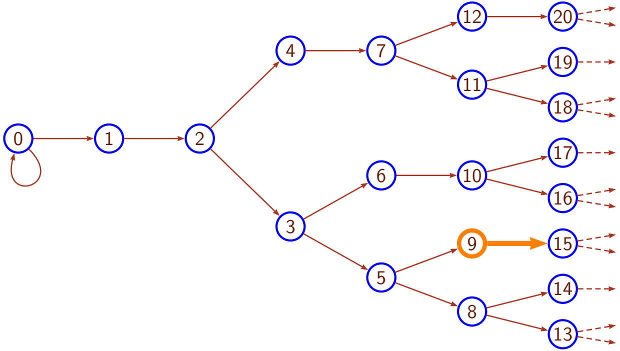


$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degree of the nodes taken in breadth-first order.

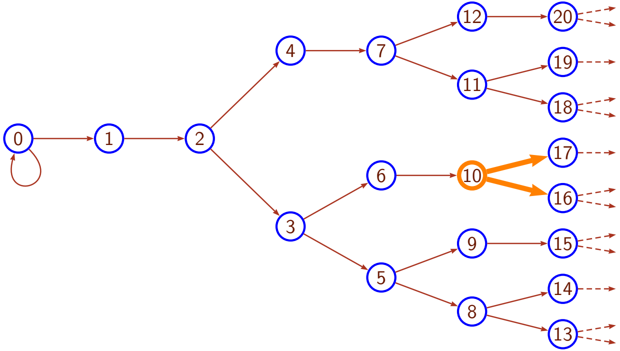


$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 1$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degree of the nodes taken in breadth-first order.

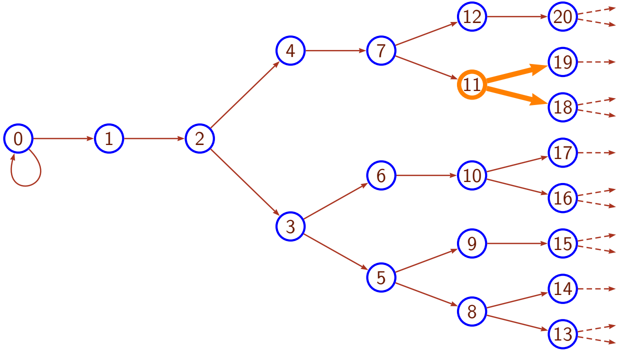


$$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 1\ 2$$

Signature of a tree

Definition

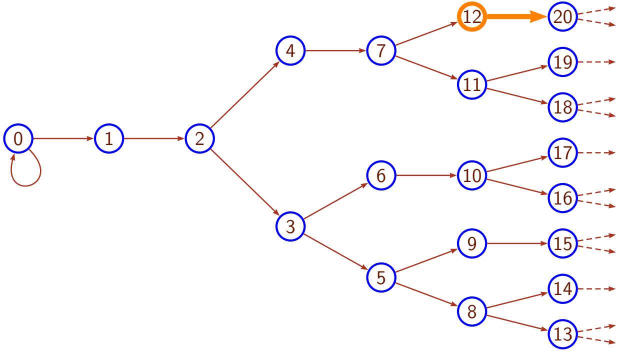
The **signature** of a tree is the sequence of the degree of the nodes taken in breadth-first order.



$$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 1\ 2\ 2$$

Definition

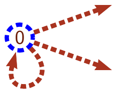
The **signature** of a tree is the sequence of the degree of the nodes taken in breadth-first order.



$$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 1\ 2\ 2\ 1\ \cdots$$

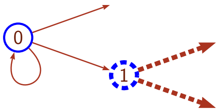
Signature is characteristic of tree

$$\mathbf{s} = (\; 3 \;\; 2 \;\; 1 \;)^{\omega}$$



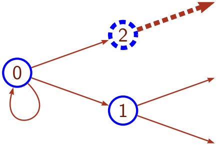
Signature is characteristic of tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



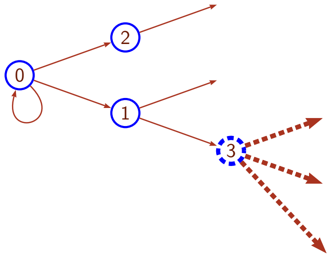
Signature is characteristic of tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



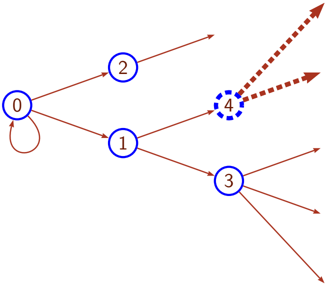
Signature is characteristic of tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



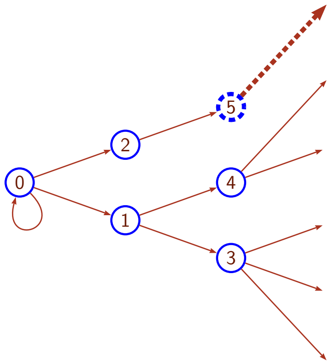
Signature is characteristic of tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



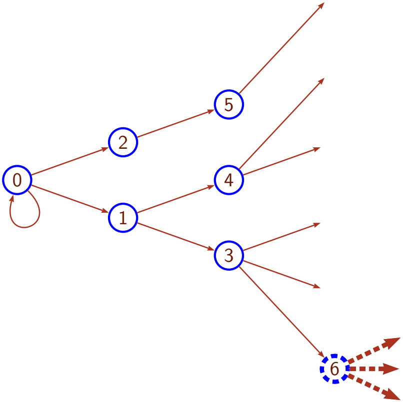
Signature is characteristic of tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



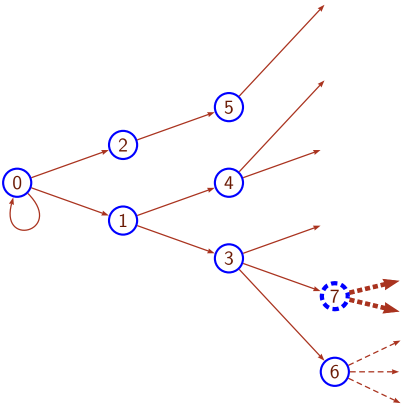
Signature is characteristic of tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



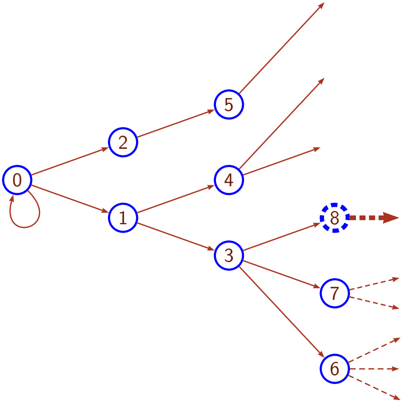
Signature is characteristic of tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



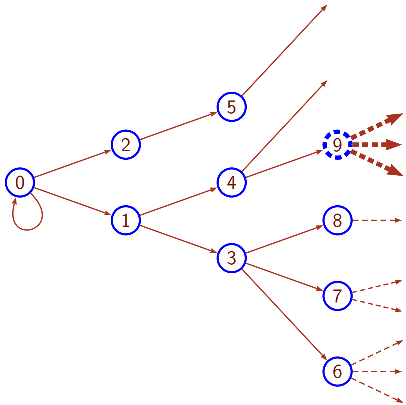
Signature is characteristic of tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



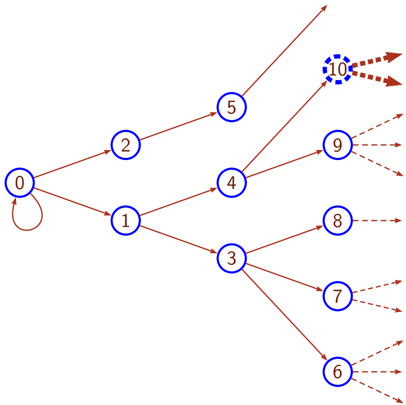
Signature is characteristic of tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



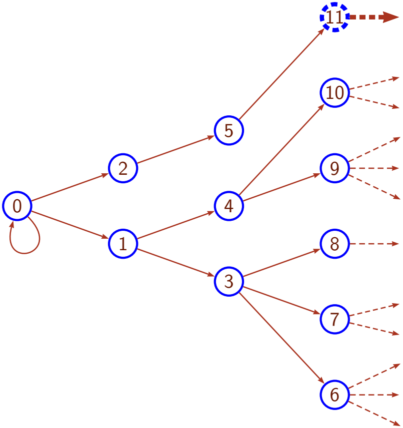
Signature is characteristic of tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



Signature is characteristic of tree

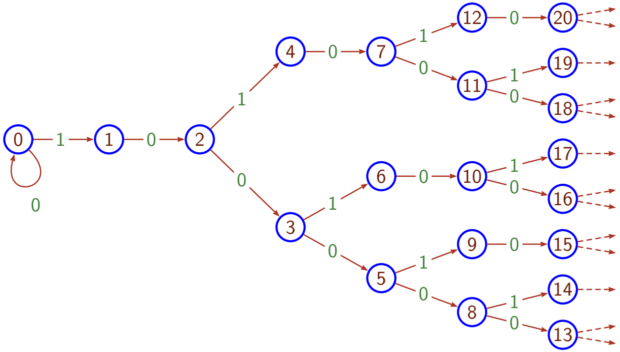
$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



Prefix-closed languages and labelled trees

Alphabets are ordered hence

prefix-closed languages = labelled trees.



Alphabets are ordered hence

prefix-closed languages = labelled trees.

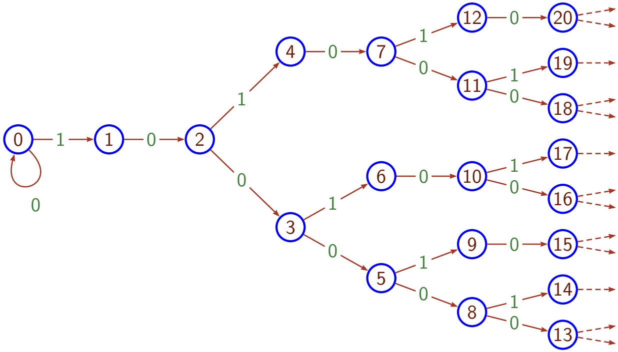


Figure : Integer representations in the Fibonacci numeration system.

Alphabets are ordered hence

prefix-closed languages $=$ labelled trees.

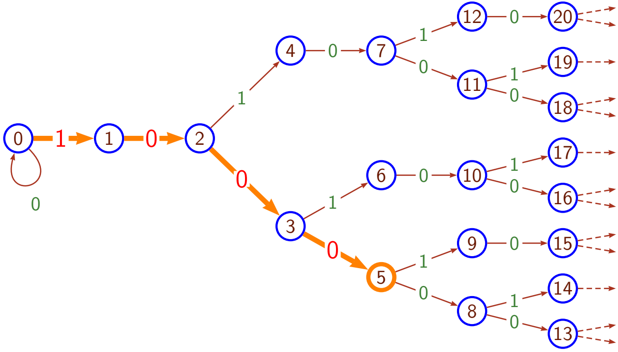


Figure : Integer representations in the Fibonacci numeration system.

Alphabets are ordered hence

prefix-closed languages = labelled trees.

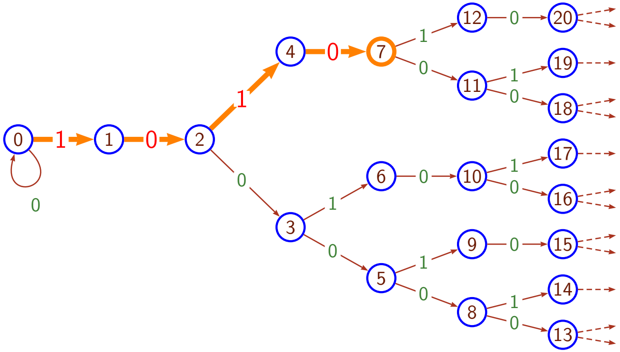
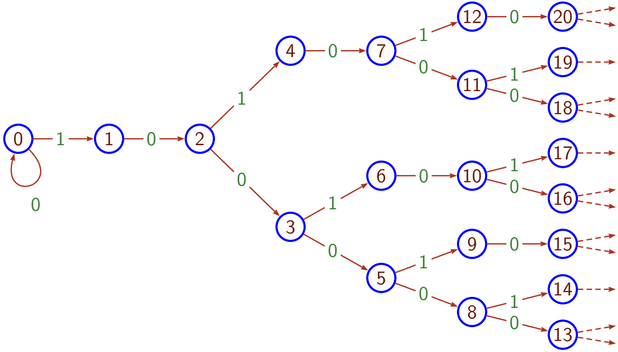


Figure : Integer representations in the Fibonacci numeration system.

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



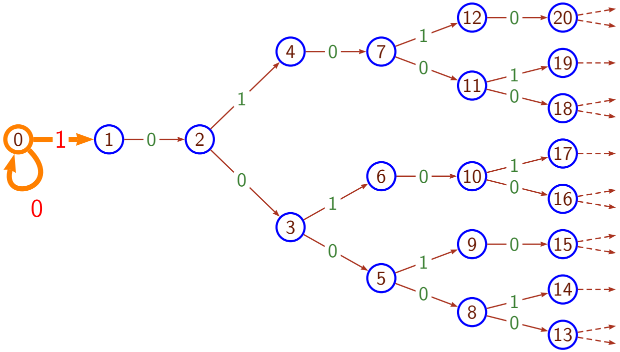
$\mathbf{s} =$

$\lambda =$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.

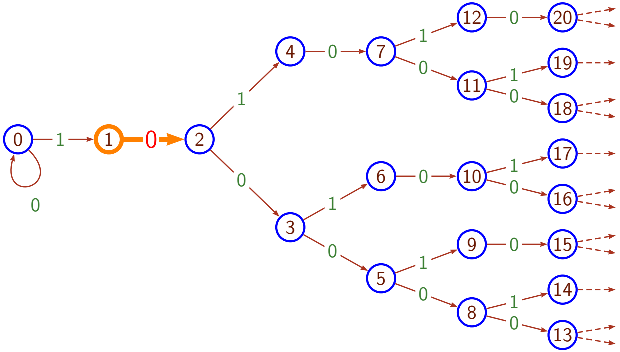


$\mathbf{s} = 2$

$\lambda = 01$

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.

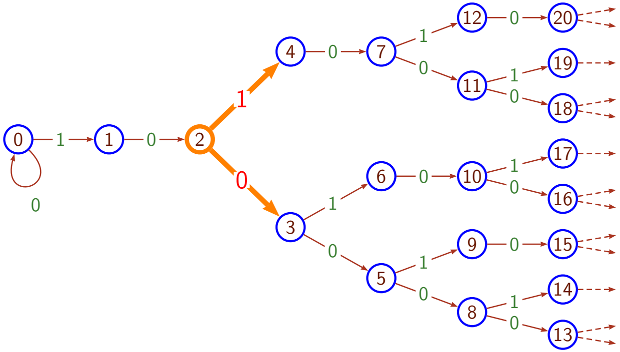


$$\mathbf{s} = 2\ 1$$

$$\lambda = 01\ 0$$

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.

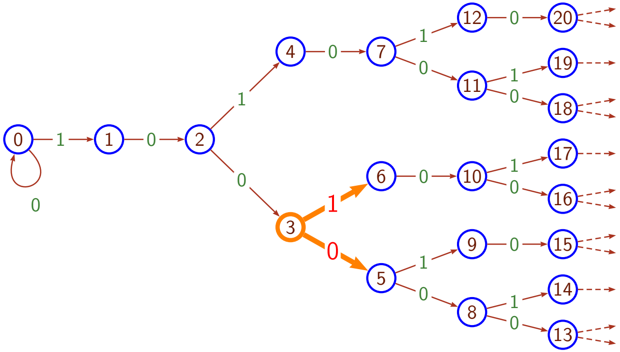


$\mathbf{s} = 2\ \ 1\ \ 2$

$\lambda = 01\ 0\ 01$

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.

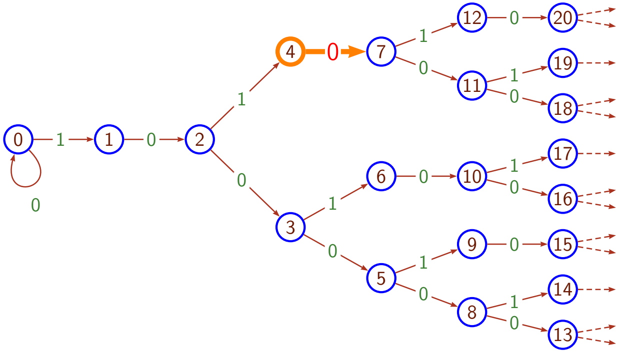


$$\mathbf{s} = 2 \ 1 \ 2 \ \ 2$$

$$\lambda = 01 \ 0 \ 01 \ 01$$

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



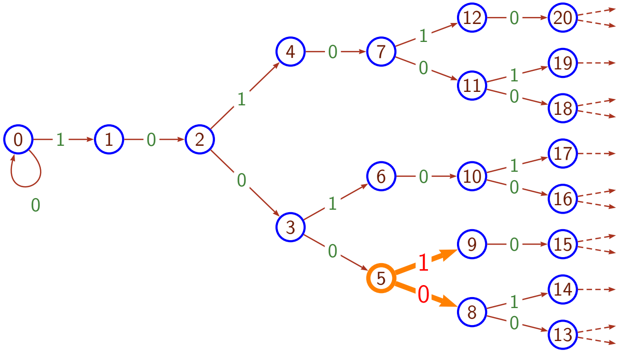
$$\mathbf{s} = 2\ 1\ 2\ \ 2\ \ 1$$

$$\lambda = 01\ 0\ 01\ 01\ 0$$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.

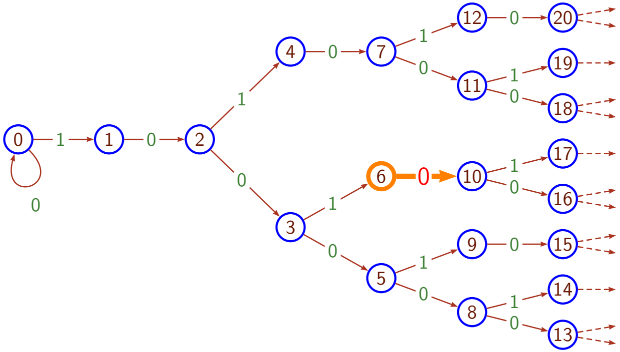


$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2$$

$$\lambda = 01\ 0\ 01\ 01\ 0\ 01$$

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



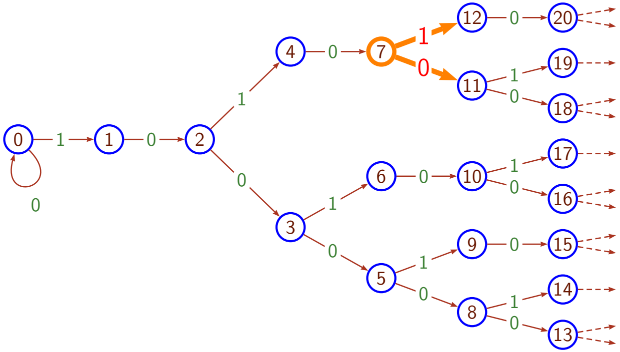
$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1$

$\lambda = 01\ 0\ 01\ 01\ 0\ 01\ 0$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



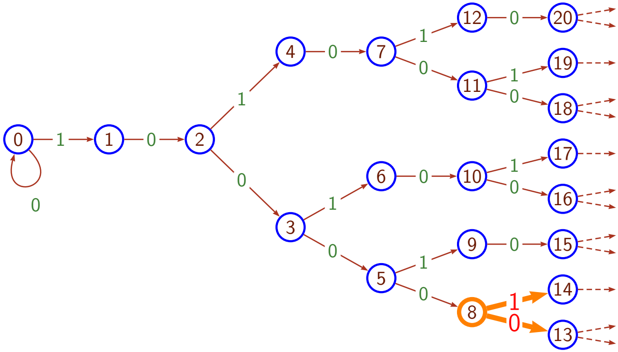
$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2$

$\lambda = 01\ 0\ 01\ 01\ 0\ 01\ 0\ 01$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



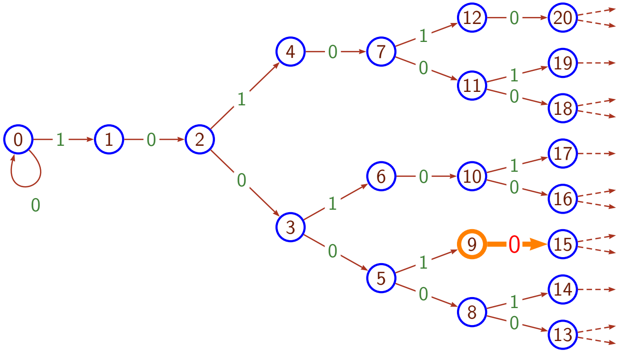
$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2\ \ 2$$

$$\lambda = 01\ 0\ 01\ 01\ 0\ 01\ 0\ 01\ \ 01$$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



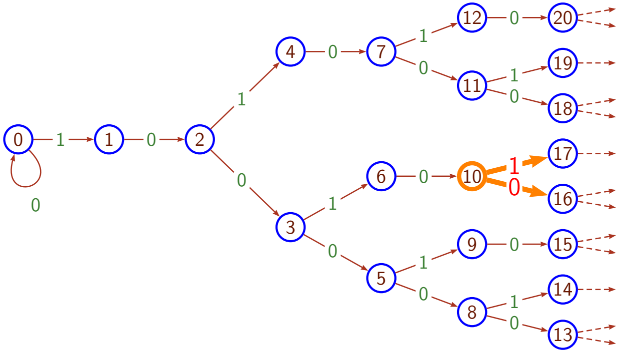
$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2\ \ 2\ 1$$

$$\lambda = 01\ 0\ 01\ 01\ 0\ 01\ 0\ 01\ 01\ 0$$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



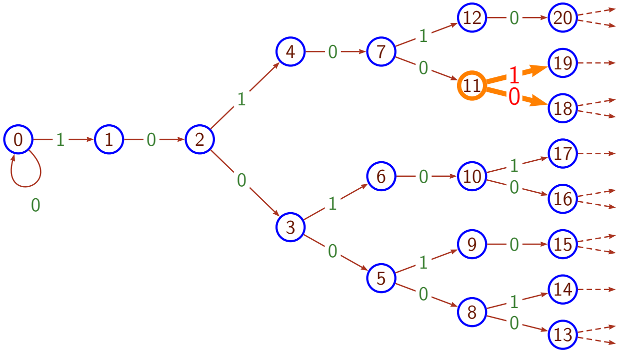
$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2\ \ 2\ 1\ 2$$

$$\lambda = 01\ 0\ 01\ 01\ 0\ 01\ 0\ 01\ 01\ 0\ 01$$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



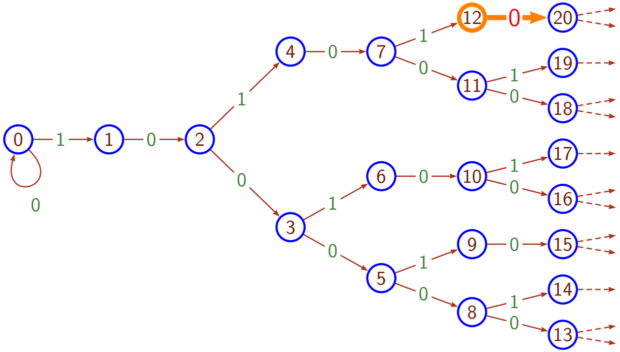
$$\mathbf{s} = 2\ 1\ 2\quad 2\ 1\ 2\ 1\ 2\quad 2\ 1\ 2\quad 2$$

$$\lambda = 01\ 0\ 01\ 01\ 0\ 01\ 0\ 01\ 01\ 0\ 01\ 01$$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.

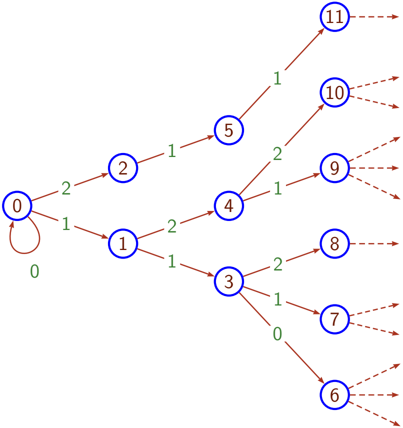


$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2\ \ 2\ 1\ 2\ \ 2\ 1\ \cdots$$

$$\lambda = 01\ 0\ 01\ 01\ 0\ 01\ 0\ 01\ 01\ 0\ 01\ 01\ 0\ \cdots$$

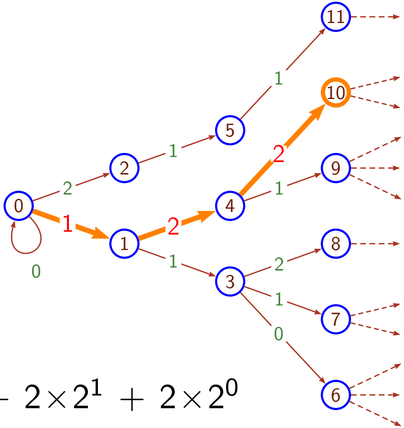
The pair signature/labelling is characteristic

$$\mathbf{s} = (3\ 2\ 1)^\omega$$
$$\lambda = (012\ 12\ 1)^\omega$$



The pair signature/labelling is characteristic

$$\mathbf{s} = (3\ 2\ 1)^\omega$$
$$\lambda = (012\ 12\ 1)^\omega$$



$$10 = 1\times 2^2 + 2\times 2^1 + 2\times 2^0$$

Figure : Non-canonical integer representations in base 2.

Theorem

L: a prefix-closed language.
Signature(L) is substitutive $\Leftrightarrow$ L is accepted by a finite automaton.

A word on substitution

A substitution σ is a morphism $A^* \to A^*$.

Running examples

Fibonacci substitution: $\{a, b\} \to \{a, b\}^*$
$a \mapsto ab$
$b \mapsto a$

A word on substitution

A substitution σ is a morphism $A^* \to A^*$.

Running examples

Fibonacci substitution: $\{a, b\} \to \{a, b\}^*$

$a \mapsto ab$

$b \mapsto a$

Periodic substitution: $\{a, b, c\} \to \{a, b, c\}^*$

$a \mapsto abc$

$b \mapsto ab$

$c \mapsto c$

A word on substitution

A substitution σ is a morphism $A^* \to A^*$.

σ is prolongable on a if $\sigma(a)$ starts with the letter a.

Running examples

Fibonacci substitution: $\{a, b\} \to \{a, b\}^*$

$a \mapsto ab$

$b \mapsto a$

Periodic substitution: $\{a, b, c\} \to \{a, b, c\}^*$

$a \mapsto abc$

$b \mapsto ab$

$c \mapsto c$

A word on substitution

A substitution σ is a morphism $A^* \to A^*$.

σ is prolongable on a if $\sigma(a)$ starts with the letter a.
In this case, $\sigma^\omega(a)$ exists and is called a purely substitutive word.

Running examples

Fibonacci substitution: $\{a, b\} \to \{a, b\}^*$
$a \mapsto ab$
$b \mapsto a$

Periodic substitution: $\{a, b, c\} \to \{a, b, c\}^*$
$a \mapsto abc$
$b \mapsto ab$
$c \mapsto c$

Substitutive signature

σ: a substitution prolongable on a.

f : a letter-to-letter morphism.
$\rightarrow$ $f(\sigma^\omega(a))$ is called a subtitutive word.

Substitutive signature

σ: a substitution prolongable on a.

f : a letter-to-letter morphism.
$\rightarrow$ $f(\sigma^\omega(a))$ is called a subtitutive word.

Definitions

let $f_\sigma : A^* \rightarrow D^*$ be the (letter-to-letter) morphism defined by

- $D \subset N$
- $\forall b,\ f_\sigma(b) = |\sigma(b)|$

We call $f_\sigma(\sigma^\omega(a))$ a **subtitutive signature**.

Substitutive signature

σ: a substitution prolongable on a.

f : a letter-to-letter morphism.
$\rightarrow$ $f(\sigma^\omega(a))$ is called a subtitutive word.

Definitions

let $f_\sigma : A^* \rightarrow D^*$ be the (letter-to-letter) morphism defined by

- $D \subset N$
- $\forall b, \; f_\sigma(b) = |\sigma(b)|$

We call $f_\sigma(\sigma^\omega(a))$ a **subtitutive signature**.

If g is a morphism such that

- $\forall b, \; |g(b)| = |\sigma(b)|$
- if $g(b) = c_0 c_1 \cdots c_k$ then $c_0 < c_1 < \cdots < c_k$

We call $g(\sigma^\omega(a))$ a **substitutive labelling**.

Example 1 – the Fibonacci signature

$$\sigma(a) = \text{ab} \implies f_\sigma(a) = 2$$
$$\sigma(b) = \text{a} \implies f_\sigma(b) = 1$$
$$f_\sigma(\sigma^\omega(a)) = 2122121221221212212122\cdots$$

Example 1 – the Fibonacci signature

$$\sigma(a) = \text{ab} \implies f_\sigma(a) = 2$$
$$\sigma(b) = \text{a} \implies f_\sigma(b) = 1$$
$$f_\sigma(\sigma^\omega(a)) = 2122121221221212212122\cdots$$

if we choose g:

$$g(a) = 01$$
$$g(b) = 0$$
$$g(\sigma^\omega(a)) = 01\,0\,01\,01\,0\,01\,0\,01\,01\,0\,01\,01\,0\,\cdots$$

Example 1 – the Fibonacci signature

$$\sigma(a) = \text{ab} \quad \Longrightarrow \quad f_\sigma(a) = 2$$
$$\sigma(b) = \text{a} \quad \Longrightarrow \quad f_\sigma(b) = 1$$
$$f_\sigma(\sigma^\omega(a)) = 21221212212212122122\cdots$$

if we choose g:

$$g(a) = 01$$
$$g(b) = 0$$
$$g(\sigma^\omega(a)) = 01\,0\,01\,01\,0\,01\,0\,01\,01\,0\,01\,01\,0\,\cdots$$

Example 1 – the Fibonacci signature

$$\sigma(a) = \text{ab} \implies f_\sigma(a) = 2$$
$$\sigma(b) = \text{a} \implies f_\sigma(b) = 1$$
$$f_\sigma(\sigma^\omega(a)) = 2122121221221212212122\cdots$$

if we choose g:

$$g(a) = 01$$
$$g(b) = 0$$
$$g(\sigma^\omega(a)) = 01\,0\,01\,01\,0\,01\,0\,01\,01\,0\,01\,01\,0\,\cdots$$

Example 1 – the Fibonacci signature

$$\begin{array}{lcl} \sigma(a) = \text{ab} & \Longrightarrow & f_\sigma(a) = 2 \\ \sigma(b) = \text{a} & \Longrightarrow & f_\sigma(b) = 1 \end{array}$$

$$f_\sigma(\sigma^\omega(a)) \quad = \quad 2122121221221212212122\cdots$$

if we choose g:

$g(a) = 01$

$g(b) = 0$

$$g(\sigma^\omega(a)) \quad = \quad 01\,0\,01\,01\,0\,01\,0\,01\,01\,0\,01\,01\,0\cdots$$

This pair signature/labelling defines the language of integer representations in the Fibonacci numeration system.

Example 2 – a periodic signature

$$\sigma(a) = abc \implies f_\sigma(a) = 3$$
$$\sigma(b) = ab \implies f_\sigma(b) = 2$$
$$\sigma(c) = c \implies f_\sigma(c) = 1$$
$$\sigma(abc) = abc\,abc \quad \text{hence } f_\sigma(\sigma^\omega(a)) = (321)^\omega$$

Example 2 – a periodic signature

$$\sigma(a) = abc \implies f_\sigma(a) = 3$$
$$\sigma(b) = ab \implies f_\sigma(b) = 2$$
$$\sigma(c) = c \implies f_\sigma(c) = 1$$
$$\sigma(abc) = abc\,abc \quad \text{hence } f_\sigma(\sigma^\omega(a)) = (321)^\omega$$

If we choose g:

$g(a) = 012$

$g(b) = 12$

$g(c) = 1$

$$g(\sigma^\omega(a)) = (012\,12\,1)^\omega$$

Example 2 – a periodic signature

$\sigma(a) = abc \implies f_\sigma(a) = 3$
$\sigma(b) = ab \implies f_\sigma(b) = 2$
$\sigma(c) = c \implies f_\sigma(c) = 1$

$$\sigma(abc) = abc\,abc \quad \text{hence } f_\sigma(\sigma^\omega(a)) = (321)^\omega$$

If we choose g:
$g(a) = 012$
$g(b) = 12$
$g(c) = 1$

$$g(\sigma^\omega(a)) = (012\,12\,1)^\omega$$

This pair signature/labelling defines a *non-canonical* representation of integers in base 2.

Forward direction of the theorem

Theorem

L: a prefix-closed language.
Signature(L) is substitutive $\Leftrightarrow$ L is accepted by a finite automaton.

Forward direction of the theorem

Theorem

L: a prefix-closed language.
Signature(L) is substitutive $\Leftrightarrow$ L is accepted by a finite automaton.

(σ, g): a substitutive signature.
(σ, g) defines a finite automaton $\mathcal{A}_{(\sigma,g)}$.
It is analogous to

- the *prefix graph/automaton* in Dumont–Thomas '89,'91,'93
- or the correspondence used in Maes–Rigo '02.

Forward direction of the theorem

Theorem

L: a prefix-closed language.
Signature(L) is substitutive $\Leftrightarrow$ L is accepted by a finite automaton.

(σ, g): a substitutive signature.
(σ, g) defines a finite automaton $\mathcal{A}_{(\sigma,g)}$.
It is analogous to

- the *prefix graph/automaton* in Dumont–Thomas '89,'91,'93
- or the correspondence used in Maes–Rigo '02.

Proposition

The language accepted by $\mathcal{A}_{(\sigma,g)}$ has signature (σ, g).

Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle \, \mathsf{A} \, , B, \, \delta \, , \, \{\mathsf{a}\} \, , \, \mathsf{A} \, \rangle$$

$\sigma(\,\mathsf{a}\,) = \,\mathsf{a\,b}$
$\sigma(\,\mathsf{b}\,) = \,\mathsf{a}$

$g(\,\mathsf{a}\,) = \,0\,1$
$g(\,\mathsf{b}\,) = \,0$

Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathsf{A}\,, B,\, \delta\,,\, \{\mathsf{a}\}\,,\, \mathsf{A}\,\rangle$$

$\sigma(\,\mathsf{a}\,) = \mathsf{a\,b}$
$\sigma(\,\mathsf{b}\,) = \mathsf{a}$

$g(\,\mathsf{a}\,) = 0\,1$
$g(\,\mathsf{b}\,) = 0$

Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathsf{A}\,, B,\, \delta\,,\, \{\mathsf{a}\},\, \mathsf{A}\,\rangle$$

$\sigma(\mathsf{a}) = \mathsf{a}\,\mathsf{b}$
$\sigma(\mathsf{b}) = \mathsf{a}$

$g(\mathsf{a}) = 0\,1$
$g(\mathsf{b}) = 0$

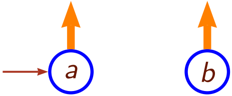
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathsf{A}\,, B,\, \delta\,,\, \{\mathsf{a}\}\,,\, \mathsf{A}\,\rangle$$

$\sigma(\,\mathsf{a}\,) = \mathsf{a\,b}$
$\sigma(\,\mathsf{b}\,) = \mathsf{a}$

$g(\,\mathsf{a}\,) = 0\,1$
$g(\,\mathsf{b}\,) = 0$



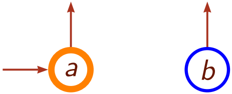
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathsf{A}\,, B, \delta, \{\mathsf{a}\}\,, \mathsf{A} \,\rangle$$

$\sigma(\mathsf{a}) = \mathsf{a}\,\mathsf{b}$
$\sigma(\mathsf{b}) = \mathsf{a}$

$g(\mathsf{a}) = 0\,1$
$g(\mathsf{b}) = 0$



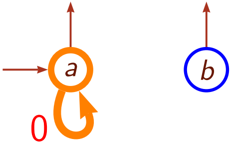
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathrm{A}\,, B, \delta, \{\mathrm{a}\}\,, \mathrm{A}\,\rangle$$

$\sigma(\mathrm{a}) = \mathrm{a}\,\mathrm{b}$
$\sigma(\mathrm{b}) = \mathrm{a}$

$g(\mathrm{a}) = 0\,1$
$g(\mathrm{b}) = 0$



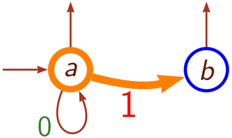
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathsf{A}\,, B, \delta, \{\mathsf{a}\}\,, \mathsf{A}\,\rangle$$

$\sigma(\,\mathsf{a}\,) = \mathsf{a}\,\mathsf{b}$
$\sigma(\,\mathsf{b}\,) = \mathsf{a}$

$g(\,\mathsf{a}\,) = 0\,1$
$g(\,\mathsf{b}\,) = 0$



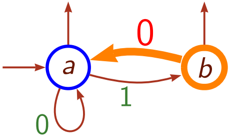
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathrm{A}\,, B, \delta, \{\mathrm{a}\}\,, \mathrm{A}\,\rangle$$

$\sigma(\mathrm{a}) = \mathrm{a\,b}$
$\sigma(\mathrm{b}) = \mathrm{a}$

$g(\mathrm{a}) = 0\,1$
$g(\mathrm{b}) = 0$



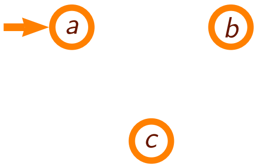
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle \mathsf{A}, B, \delta, \{\mathsf{a}\}, \mathsf{A} \rangle$$

$$\sigma(\mathsf{a}) = \mathsf{a\,b\,c}$$
$$\sigma(\mathsf{b}) = \mathsf{a\,b}$$
$$\sigma(\mathsf{c}) = \mathsf{c}$$

$$g(\mathsf{a}) = 0\,1\,2$$
$$g(\mathsf{b}) = 1\,2$$
$$g(\mathsf{c}) = 1$$



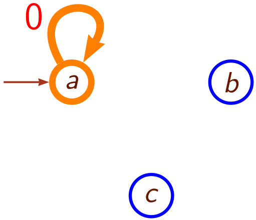
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathsf{A}\,, B, \delta, \{\mathsf{a}\}\,, \mathsf{A}\,\rangle$$

$\sigma(\mathsf{a}) = \mathsf{a\,b\,c}$
$\sigma(\mathsf{b}) = \mathsf{a\,b}$
$\sigma(\mathsf{c}) = \mathsf{c}$

$g(\mathsf{a}) = 0\,1\,2$
$g(\mathsf{b}) = 1\,2$
$g(\mathsf{c}) = 1$



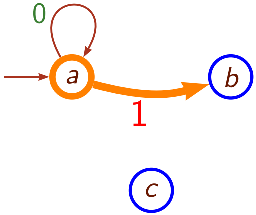
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, A\,, B, \delta, \{a\}\,, A \,\rangle$$

$\sigma(a) = abc$
$\sigma(b) = ab$
$\sigma(c) = c$

$g(a) = 012$
$g(b) = 12$
$g(c) = 1$



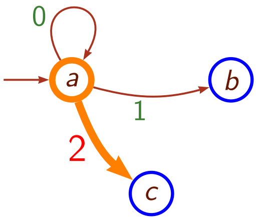
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathsf{A}\,, B, \delta, \{\mathsf{a}\}\,, \mathsf{A}\,\rangle$$

$\sigma(\mathsf{a}) = \mathsf{a\,b\,c}$
$\sigma(\mathsf{b}) = \mathsf{a\,b}$
$\sigma(\mathsf{c}) = \mathsf{c}$

$g(\mathsf{a}) = 0\,1\,2$
$g(\mathsf{b}) = 1\,2$
$g(\mathsf{c}) = 1$



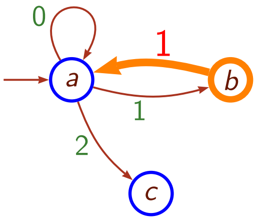
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle \mathsf{A}, B, \delta, \{\mathsf{a}\}, \mathsf{A} \rangle$$

$\sigma(\mathsf{a}) = \mathsf{a\,b\,c}$
$\sigma(\mathsf{b}) = \mathsf{a\,b}$
$\sigma(\mathsf{c}) = \mathsf{c}$

$g(\mathsf{a}) = 0\,1\,2$
$g(\mathsf{b}) = 1\,2$
$g(\mathsf{c}) = 1$



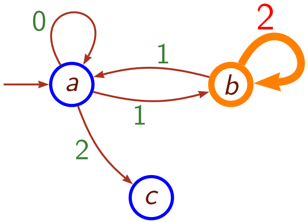
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathsf{A}\,, B, \delta, \{\mathsf{a}\}\,, \mathsf{A}\,\rangle$$

$\sigma(\mathsf{a}) = \mathsf{a\,b\,c}$
$\sigma(\mathsf{b}) = \mathsf{a\,b}$
$\sigma(\mathsf{c}) = \mathsf{c}$

$g(\mathsf{a}) = 0\,1\,2$
$g(\mathsf{b}) = 1\,2$
$g(\mathsf{c}) = 1$



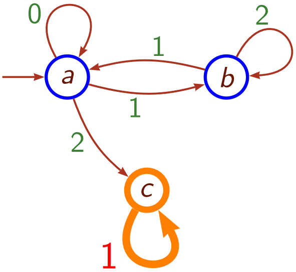
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle \mathsf{A}, B, \delta, \{\mathsf{a}\}, \mathsf{A} \rangle$$

$\sigma(\mathsf{a}) = \mathsf{a\,b\,c}$
$\sigma(\mathsf{b}) = \mathsf{a\,b}$
$\sigma(\mathsf{c}) = \mathsf{c}$

$g(\mathsf{a}) = 0\,1\,2$
$g(\mathsf{b}) = 1\,2$
$g(\mathsf{c}) = 1$



Forward direction of the theorem

Theorem

L: a prefix-closed language.
Signature(L) is substitutive $\Leftrightarrow$ L is accepted by a finite automaton.

(σ, g): a substitutive signature.
(σ, g) defines a finite automaton $\mathcal{A}_{(\sigma,g)}$.
It is analogous to

- the *prefix graph/automaton* in Dumont Thomas '89,'91,'93
- or the correspondence used in Maes Rigo '02.

Proposition

The language accepted by $\mathcal{A}_{(\sigma,g)}$ has signature (σ, g).

Forward direction of the theorem

Theorem

L: a prefix-closed language.
Signature(L) is substitutive $\Leftrightarrow$ L is accepted by a finite automaton.

(σ, g): a substitutive signature.
(σ, g) defines a finite automaton $\mathcal{A}_{(\sigma,g)}$.
It is analogous to

- the *prefix graph/automaton* in Dumont Thomas '89,'91,'93
- or the correspondence used in Maes Rigo '02.

Proposition

The language accepted by $\mathcal{A}_{(\sigma,g)}$ has signature (σ, g).

Proof: Unfold the automaton $\mathcal{A}_{(\sigma,g)}$.

Backward direction of the theorem

Theorem

L: a prefix-closed language.
Signature(L) is substitutive $\Leftrightarrow$ L is accepted by a finite automaton.

$\mathcal{B}$: a finite automaton.
We define (σ_B, g_B) such that

$$\mathcal{B} \;=\; \mathcal{A}_{(\sigma_B, g_B)}$$

Backward direction of the theorem

Theorem

L: a prefix-closed language.
Signature(L) is substitutive $\Leftrightarrow$ L is accepted by a finite automaton.

$\mathcal{B}$: a finite automaton.
We define (σ_B, g_B) such that

$$\mathcal{B} = \mathcal{A}_{(\sigma_B, g_B)}$$

Proposition

The language accepted by $\mathcal{B}$ has signature $(\sigma_{\mathcal{B}}, g_{\mathcal{B}})$.

Follows directly from the other direction.

Observation

In basically every NS, the representations of integers follows the *radix order*: $\forall n, p \quad \langle n \rangle \leq_{rad} \langle n + p \rangle$

Observation

In basically every NS, the representations of integers follows the *radix order*: $\forall n, p \quad \langle n \rangle \leq_{rad} \langle n+p \rangle$

$$u <_{\text{rad}} v \quad \text{if} \quad |u| < |v| \quad \text{or} \quad |u| = |v| \ \& \ |u| <_{\text{lex}} |v|$$

Example: $2 <_{\text{rad}} 12 \qquad 12 <_{\text{rad}} 21$.

Abstract Numeration System (ANS, Lecomte-Rigo)

Observation

In basically every NS, the representations of integers follows the *radix order*: $\forall n, p \quad \langle n \rangle \leq_{rad} \langle n+p \rangle$

$$u <_{\text{rad}} v \quad \text{if} \quad |u| < |v| \quad \text{or} \quad |u| = |v| \ \& \ |u| <_{\text{lex}} |v|$$

Example: $2 <_{\text{rad}} 12 \qquad 12 <_{\text{rad}} 21$.

Definition (ANS L)

L: language over an ordered alphabet A.
$\langle n \rangle_L$ is the $(n+1)$-th word of L in the radix order.

In our scheme, $\langle n \rangle_L$ is the word labelling the path $0 \longrightarrow n$.

Prefix-closed ARNS

L: prefix-closed language accepted by a finite automaton.

Prefix-closed ARNS

L: prefix-closed language accepted by a finite automaton.

Proposition

L: prefix-closed ARNS of signature (s, λ_1)
K: prefix-closed ARNS of signature (s, λ_2)

The conversion function $\langle n \rangle_L \mapsto \langle n \rangle_K$ is very simple[†].

[†] realised by a finite, pure sequential and letter-to-letter transducer.

Prefix-closed ARNS

L: prefix-closed language accepted by a finite automaton.

Proposition

L: prefix-closed ARNS of signature (s, λ_1)
K: prefix-closed ARNS of signature (s, λ_2)

The conversion function $\langle n \rangle_L \mapsto \langle n \rangle_K$ is very simple[†].

The proof relies on a modified automata product.

[†] realised by a finite, pure sequential and letter-to-letter transducer.

Dumont-Thomas Numeration System (DTNS)

$\sigma : A \to A^*$ prolongable on a.

Example : $\sigma(a) = abc$ $\sigma(b) = ab$ $\sigma(c) = c$

Dumont-Thomas Numeration System (DTNS)

$\sigma : A \to A^*$ prolongable on a.

Example : $\sigma(a) = abc$ $\sigma(b) = ab$ $\sigma(c) = c$

Definition

$$A_\sigma = \{[u] \mid u \text{ is a strict prefix of } \sigma(b) \text{ for some } b \in A\}$$

Example : $A_\sigma = \{ [\varepsilon], [a], [ab] \}$

Dumont-Thomas Numeration System (DTNS)

$\sigma : A \to A^*$ prolongable on a.

Example : $\sigma(a) = abc$ $\sigma(b) = ab$ $\sigma(c) = c$

Definition

$$A_\sigma = \{[u] \mid u \text{ is a strict prefix of } \sigma(b) \text{ for some } b \in A\}$$

Example : $A_\sigma = \{ [\varepsilon], [a], [ab] \}$

g_σ: morphism $A^* \to A_\sigma^*$

$g_\sigma(b) = [u_0]\,[u_1] \cdots [u_{k-1}]$

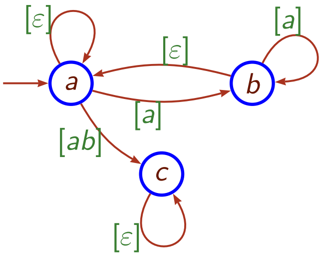
- $k = |\sigma(b)|$
- u_i is the prefix of length i of $\sigma(b)$

Example : $g_\sigma(a) = [\varepsilon]\,[a]\,[ab]$ $g_\sigma(b) = [\varepsilon]\,[a]$ $g_\sigma(c) = [\varepsilon]$

Dumont-Thomas automaton $\mathcal{A}_{(\sigma, g_\sigma)}$

$\sigma(a) = abc$
$\sigma(b) = ab$
$\sigma(c) = c$

$g_\sigma(a) = [\varepsilon]\,[a]\,[ab]$
$g_\sigma(b) = [\varepsilon]\,[a]$
$g_\sigma(c) = [\varepsilon]$



Dumont-Thomas Numeration System (DTNS)

Definition

ρ function ${A_\sigma}^* \to A^*$

$\rho([u_k] \ldots [u_2]\,[u_1]\,[u_0]) \;=\; \sigma^k(u_k)\sigma^{k-1}(u_{k-1})\cdots\sigma^2(u_2)\sigma(u_1)u_0$

Example: $\rho_\sigma([a]\,[\varepsilon]\,[ab]) \;=\; \sigma^2(a) \quad \sigma(\varepsilon) \quad ab$

$\qquad abc\,abc \quad \varepsilon \quad ab = \text{abc abc ab}$

Dumont-Thomas Numeration System (DTNS)

Definition

ρ function ${A_\sigma}^* \to A^*$

$\rho([u_k]\ \dots\ [u_2]\,[u_1]\,[u_0]) \ = \ \sigma^k(u_k)\sigma^{k-1}(u_{k-1})\cdots\sigma^2(u_2)\sigma(u_1)u_0$

Example: $\rho_\sigma([a]\,[\varepsilon]\,[ab]) = \begin{array}{ccc} \sigma^2(a) & \sigma(\varepsilon) & ab \\ abc\,abc & \varepsilon & ab \end{array} = \mathrm{abc\,abc\,ab}$

Theorem (Dumont Thomas '89)

$\forall n \in \mathbb{N}$

$\exists!$ word $[u_k]\ \dots\ [u_2]\,[u_1]\,[u_0]$ accepted by $\mathcal{A}_{(\sigma, g_\sigma)}$ such that

- $u_k \neq \varepsilon$
- $|\rho([u_k]\ \dots\ [u_2]\,[u_1]\,[u_0])| \ = \ n$

$[u_k]\ \dots\ [u_2]\,[u_1]\,[u_0]$ is the representation of n in the DTNS.

Dumont-Thomas Numeration System (DTNS)

Definition

ρ function $A_\sigma{}^* \to A^*$

$\rho([u_k]\,\ldots\,[u_2]\,[u_1]\,[u_0]) \;=\; \sigma^k(u_k)\sigma^{k-1}(u_{k-1})\cdots\sigma^2(u_2)\sigma(u_1)u_0$

Example: $\rho_\sigma([a]\,[\varepsilon]\,[ab]) = \underset{abc\,abc}{\sigma^2(a)}\;\underset{\varepsilon}{\sigma(\varepsilon)}\;\underset{ab}{ab} = \mathrm{abc\,abc\,ab}$

Theorem (Dumont Thomas '89)

$\forall n \in \mathbb{N}$

$\exists!$ word $[u_k]\,\ldots\,[u_2]\,[u_1]\,[u_0]$ accepted by $\mathcal{A}_{(\sigma, g_\sigma)}$ such that

- $u_k \neq \varepsilon$
- $|\rho([u_k]\,\ldots\,[u_2]\,[u_1]\,[u_0])| \;=\; n$

$[u_k]\,\ldots\,[u_2]\,[u_1]\,[u_0]$ is the representation of n in the DTNS.

Example: $[a]\,[\varepsilon]\,[ab]$ is the representation of 8 .

Dumont-Thomas Numeration System (DTNS)

Definition

ρ function ${A_\sigma}^* \to A^*$

$\rho([u_k] \ldots [u_2]\,[u_1]\,[u_0]) \;=\; \sigma^k(u_k)\sigma^{k-1}(u_{k-1})\cdots\sigma^2(u_2)\sigma(u_1)u_0$

Example: $\rho_\sigma([a]\,[\varepsilon]\,[ab]) = \begin{array}{ccc} \sigma^2(a) & \sigma(\varepsilon) & ab \\ abc\,abc & \varepsilon & ab \end{array} = \text{abc abc ab}$

Theorem (Dumont Thomas '89)

$\forall n \in \mathbb{N}$

$\exists!$ word $[u_k] \ldots [u_2]\,[u_1]\,[u_0]$ accepted by $\mathcal{A}_{(\sigma, g_\sigma)}$ such that

- $u_k \neq \varepsilon$
- $|\rho([u_k] \ldots [u_2]\,[u_1]\,[u_0])| \;=\; n$

$[u_k] \ldots [u_2]\,[u_1]\,[u_0]$ is the representation of n in the DTNS.

Example: $[a]\,[\varepsilon]\,[ab]$ is the representation of 8.

Theorem

1. Every DTNS is a prefix-closed ARNS.

2. Every prefix-closed ARNS is easily† convertible to a DTNS.

† Through a finite, letter-to-letter and pure sequential transducer.

Theorem

1. Every DTNS is a prefix-closed ARNS.
2. Every prefix-closed ARNS is easily† convertible to a DTNS.

† Through a finite, letter-to-letter and pure sequential transducer.

Sketch of proof of 2.

Prefix-Closed ARNS L

Theorem

1. Every DTNS is a prefix-closed ARNS.
2. Every prefix-closed ARNS is easily† convertible to a DTNS.

† Through a finite, letter-to-letter and pure sequential transducer.

Sketch of proof of 2.

Prefix-Closed ARNS $L \longrightarrow$ Automaton $\mathcal{A}$

Theorem

1. Every DTNS is a prefix-closed ARNS.

2. Every prefix-closed ARNS is easily† convertible to a DTNS.

† Through a finite, letter-to-letter and pure sequential transducer.

Sketch of proof of 2.

Prefix-Closed ARNS $L \longrightarrow$ Automaton $\mathcal{A} \longrightarrow$ Morphisms (σ, g)

Theorem

1. Every DTNS is a prefix-closed ARNS.
2. Every prefix-closed ARNS is easily† convertible to a DTNS.

† Through a finite, letter-to-letter and pure sequential transducer.

Sketch of proof of 2.

Prefix-Closed ARNS L of signature (s, λ_1) $\longrightarrow$ Automaton $\mathcal{A}$ $\longrightarrow$ Morphisms (σ, g)

where $s = f_\sigma(\sigma^\omega(a))$ $\lambda_1 = g(\sigma^\omega(a))$

Theorem

1. Every DTNS is a prefix-closed ARNS.

2. Every prefix-closed ARNS is easily† convertible to a DTNS.

† Through a finite, letter-to-letter and pure sequential transducer.

Sketch of proof of 2.

Prefix-Closed ARNS L of signature (s, λ_1) ⟶ Automaton $\mathcal{A}$ ⟶ Morphisms (σ, g)

↓

DTNS σ

where $s = f_\sigma(\sigma^\omega(a))$ $\lambda_1 = g(\sigma^\omega(a))$

Theorem

1. Every DTNS is a prefix-closed ARNS.
2. Every prefix-closed ARNS is easily† convertible to a DTNS.

† Through a finite, letter-to-letter and pure sequential transducer.

Sketch of proof of 2.

Prefix-Closed ARNS L of signature (s, λ_1) $\longrightarrow$ Automaton $\mathcal{A}$ $\longrightarrow$ Morphisms (σ, g)

$\downarrow$

DT Automaton $\mathcal{A}_{(\sigma, g_\sigma)}$ $\longleftarrow$ DTNS σ

where $s = f_\sigma(\sigma^\omega(a))$ $\qquad$ $\lambda_1 = g(\sigma^\omega(a))$

Theorem

1. Every DTNS is a prefix-closed ARNS.
2. Every prefix-closed ARNS is easily† convertible to a DTNS.

† Through a finite, letter-to-letter and pure sequential transducer.

Sketch of proof of 2.

Prefix-Closed ARNS L of signature (s, λ_1) $\longrightarrow$ Automaton $\mathcal{A}$ $\longrightarrow$ Morphisms (σ, g)

$\downarrow$

Prefix-closed ARNS K of signature (s, λ_2) $\longleftarrow$ DT Automaton $\mathcal{A}_{(\sigma, g_\sigma)}$ $\longleftarrow$ DTNS σ

where $s = f_\sigma(\sigma^\omega(a))$ $\quad \lambda_1 = g(\sigma^\omega(a))$ $\quad \lambda_2 = g_\sigma(\sigma^\omega(a))$

Theorem

1. Every DTNS is a prefix-closed ARNS.
2. Every prefix-closed ARNS is easily† convertible to a DTNS.

† Through a finite, letter-to-letter and pure sequential transducer.

Sketch of proof of 2.

Prefix-Closed ARNS L of signature (s, λ_1) $\longrightarrow$ Automaton $\mathcal{A}$ $\longrightarrow$ Morphisms (σ, g)

$\downarrow$

Prefix-closed ARNS K of signature (s, λ_2) $\longleftarrow$ DT Automaton $\mathcal{A}_{(\sigma, g_\sigma)}$ $\longleftarrow$ DTNS σ

where $s = f_\sigma(\sigma^\omega(a))$ $\quad \lambda_1 = g(\sigma^\omega(a))$ $\quad \lambda_2 = g_\sigma(\sigma^\omega(a))$

Proposition

L: prefix-closed ARNS of signature (s, λ_1)
K: prefix-closed ARNS of signature (s, λ_2)
The conversion function $\langle n \rangle_L \mapsto \langle n \rangle_K$ is very simple†.

$$\mathbf{s} \;=\; u\, r^{\omega} \qquad \text{with} \qquad r \;=\; r_0\, r_1\, r_2 \cdots r_{q-1}$$

Definition: growth ratio

$$\mathtt{gr}(\mathbf{s}) \;=\; \frac{r_0 + r_1 + \cdots + r_{q-1}}{q}$$

Other works: Ultimately periodic signatures

$$\mathbf{s} \;=\; u\, r^{\omega} \qquad \text{with} \qquad r \;=\; r_0\, r_1\, r_2 \cdots r_{q-1}$$

Definition: growth ratio

$$\mathtt{gr}(\mathbf{s}) \;=\; \frac{r_0+r_1+\cdots+r_{q-1}}{q}$$

Theorem (MS, to appear)

If $\mathtt{gr}(\mathbf{s}) \in \mathbb{N}$, then **s** generates the language of a finite automaton.
It is linked[‡] to the integer base $b = \mathtt{gr}(\mathbf{s})$.

If $\mathtt{gr}(\mathbf{s}) \notin \mathbb{N}$, then **s** generates a non-context-free language.
It is linked[‡] to the *rational base* $\frac{p}{q} = \mathtt{gr}(\mathbf{s})$. (cf. Akiyama et al. '08)

‡ It is a non-canonical representation of the integers (using extra digits).