Breadth-first serialisation of trees and rational languages

Victor Marsault,
joint work with Jacques Sakarovitch

CNRS / Telecom-ParisTech, Paris, France

Numeration and Substitution 2014, Debrecen
2014–07–11

Outline

We call *tree* a...

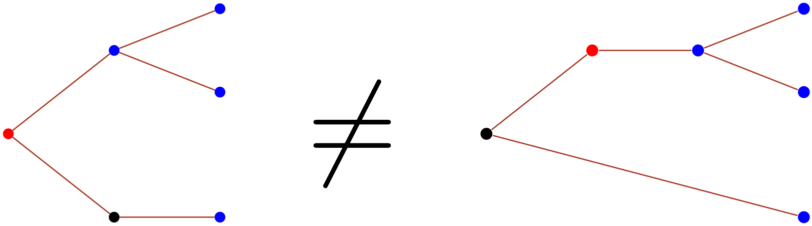
Directed graph which is

- **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 each 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 each other node.
- **Ordered:** the children of every node are ordered
  (In the figures, lower children are smaller.)



We call *tree* a...

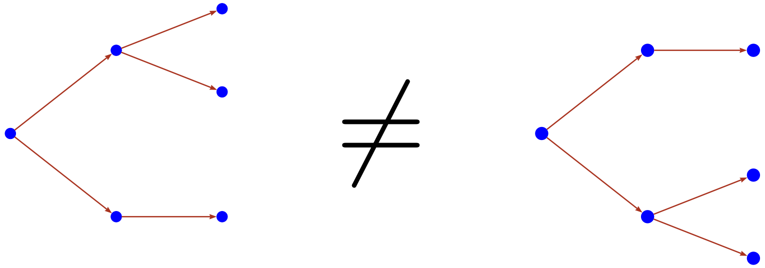
Directed graph which is

- **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 each other node.
- **Ordered:** the children of every node are ordered
  (In the figures, lower children are smaller.)

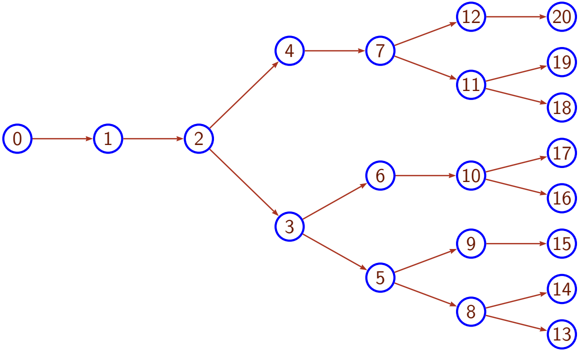
We call *tree* a...

Directed graph which is

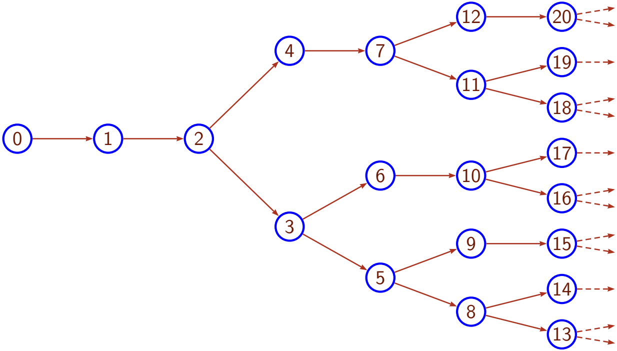
- **Rooted:** a node is called *the root* (leftmost in the figures)
- **Directed outward from the root:** there is a unique path from the root to each other node.
- **Ordered:** the children of every node are ordered
  (In the figures, lower children are smaller.)



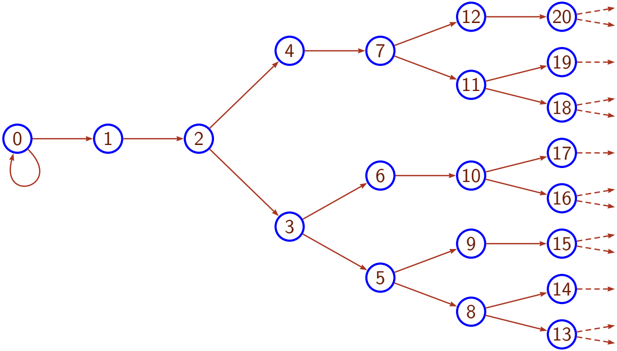
Trees have a canonical breadth-first traversal



Trees have a canonical breadth-first traversal



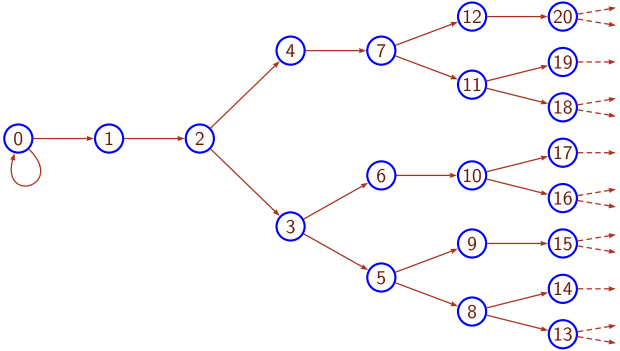
Trees have a canonical breadth-first traversal



Signature of a tree

Definition

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

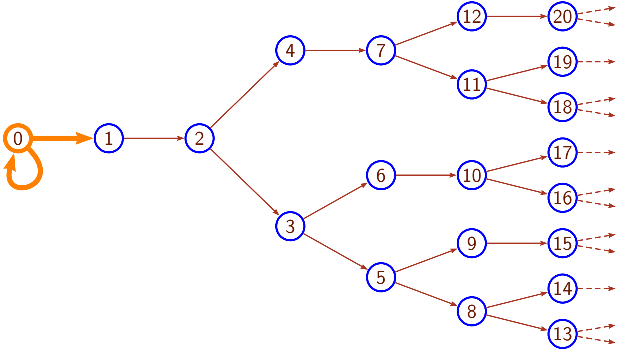


$\mathbf{s} =$

Signature of a tree

Definition

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

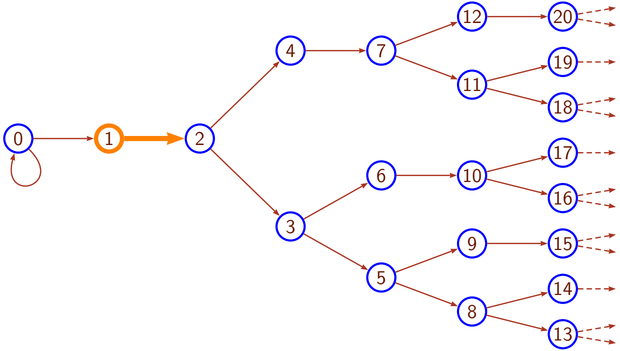


$\mathbf{s} = 2$

Signature of a tree

Definition

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

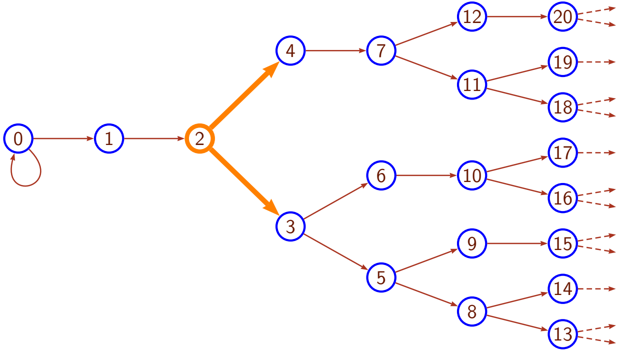


$$\mathbf{s} = 2\ 1$$

Signature of a tree

Definition

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

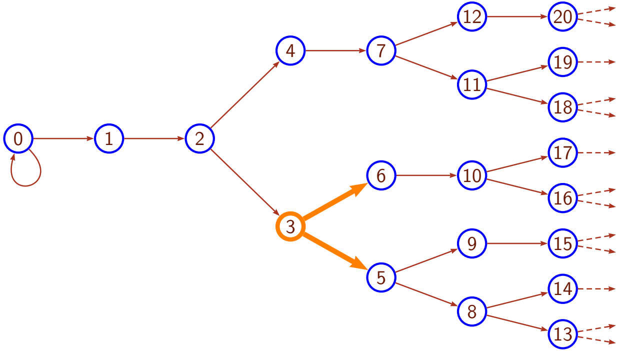


$\mathbf{s} = 2\ 1\ 2$

Signature of a tree

Definition

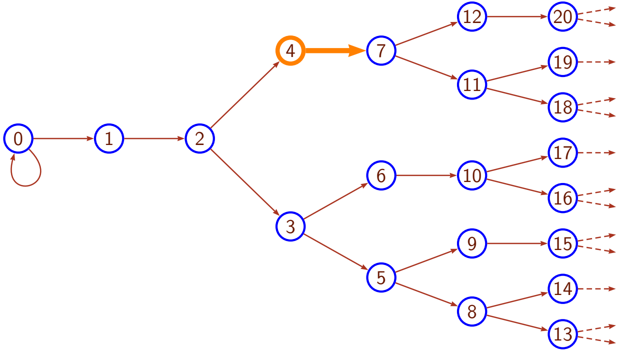
The **signature** of a tree is the sequence of the degree of its node taken in breadth-first order.



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

Definition

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

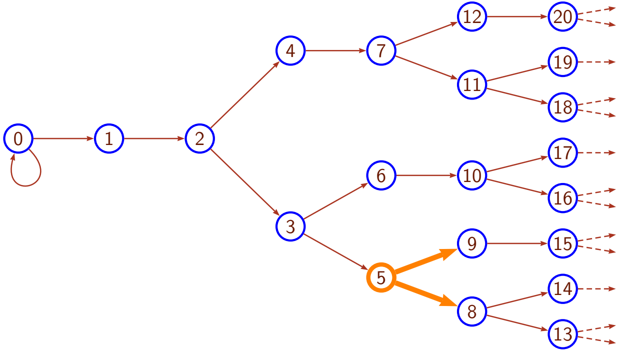


$\mathbf{s} = 2\ 1\ 2\ 2\ 1$

Signature of a tree

Definition

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

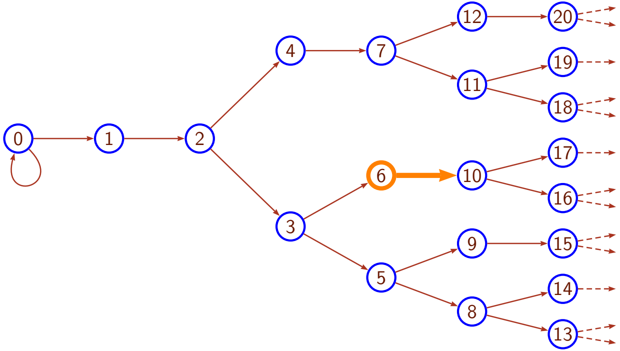


$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2$

Signature of a tree

Definition

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

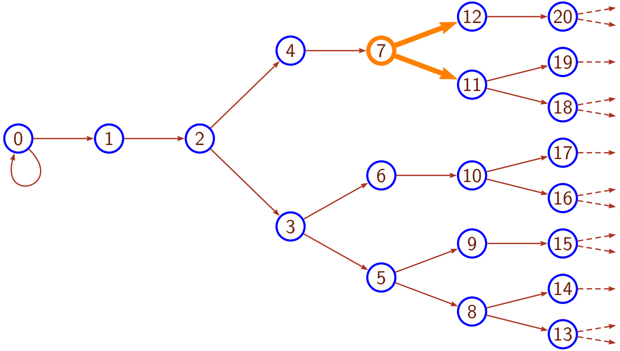


$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1$

Signature of a tree

Definition

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

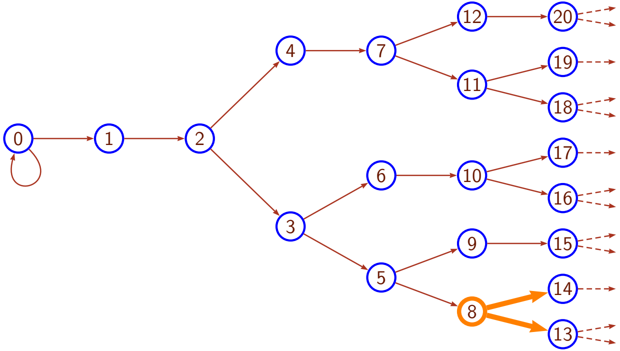


$$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2$$

Signature of a tree

Definition

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

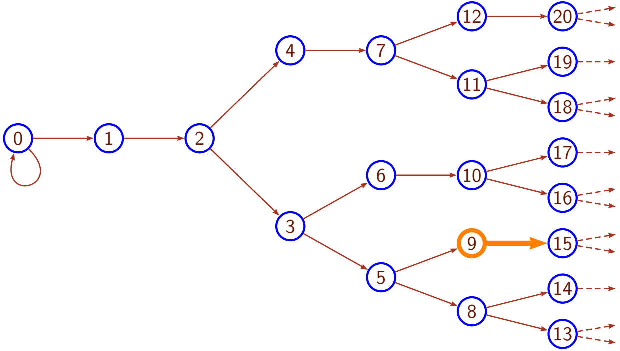


$$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2$$

Signature of a tree

Definition

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

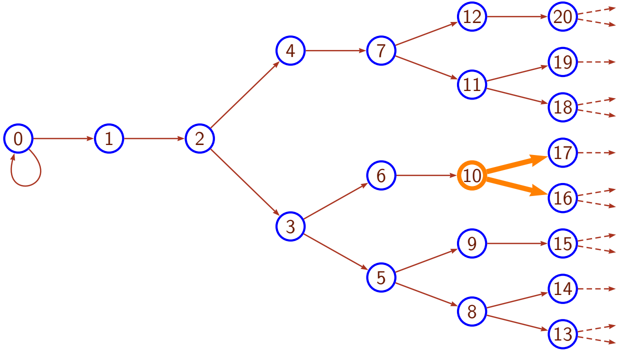


$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 1$

Signature of a tree

Definition

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

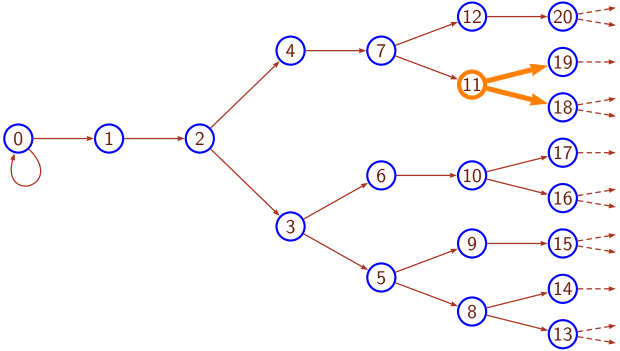


$$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 1\ 2$$

Signature of a tree

Definition

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

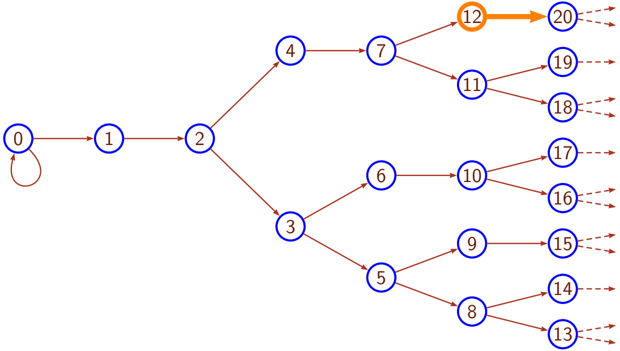


$$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 1\ 2\ 2$$

Signature of a tree

Definition

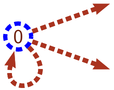
The **signature** of a tree is the sequence of the degree of its node taken in breadth-first order.



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

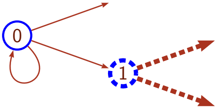
Tree generated by the signature $(3\,2\,1)^\omega$

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



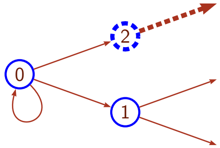
Tree generated by the signature $(3\,2\,1)^\omega$

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



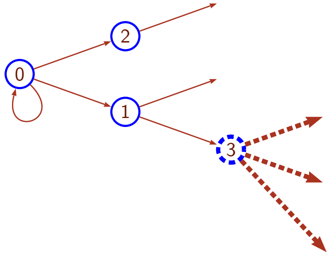
Tree generated by the signature $(3\,2\,1)^\omega$

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



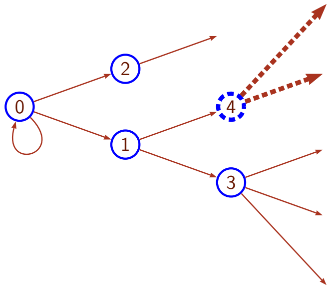
Tree generated by the signature $(3\,2\,1)^\omega$

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



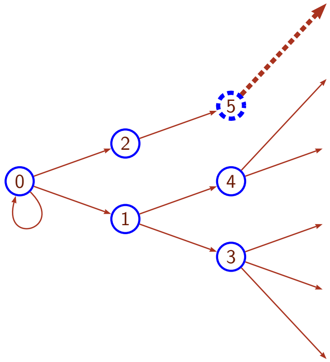
Tree generated by the signature $(3\,2\,1)^\omega$

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



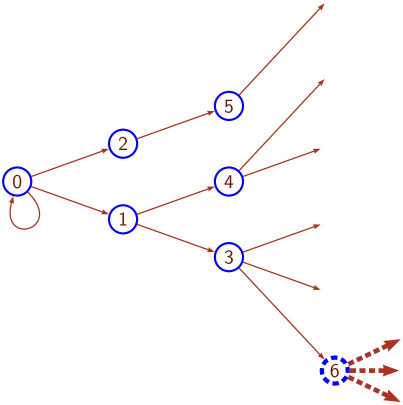
Tree generated by the signature $(3\,2\,1)^\omega$

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



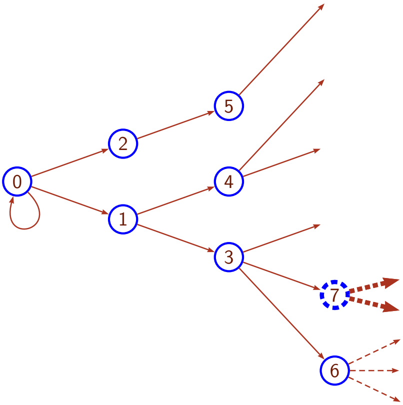
Tree generated by the signature $(3\,2\,1)^\omega$

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



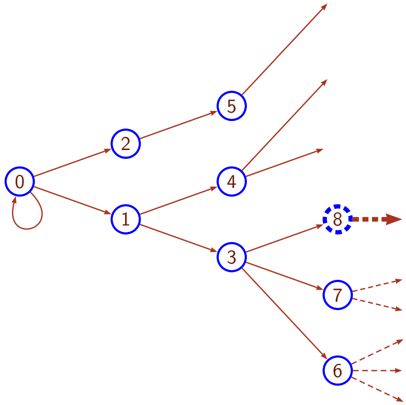
Tree generated by the signature $(3\,2\,1)^\omega$

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



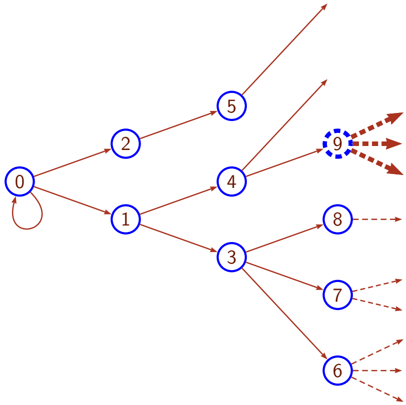
Tree generated by the signature $(3\,2\,1)^\omega$

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



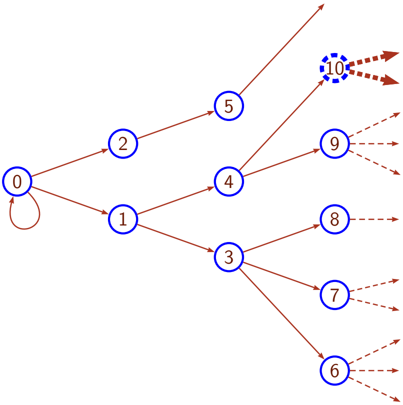
Tree generated by the signature $(3\,2\,1)^\omega$

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



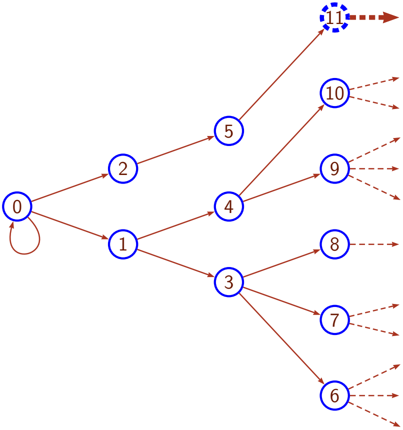
Tree generated by the signature $(3\,2\,1)^\omega$

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



Tree generated by the signature $(3\,2\,1)^\omega$

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



Prefix-closed languages and labelled trees

Alphabets are ordered hence

prefix-closed languages = labelled trees.

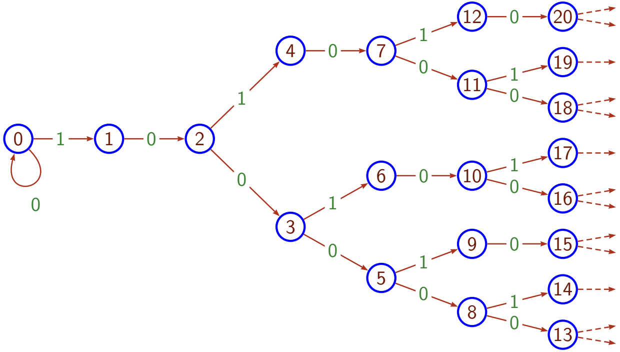
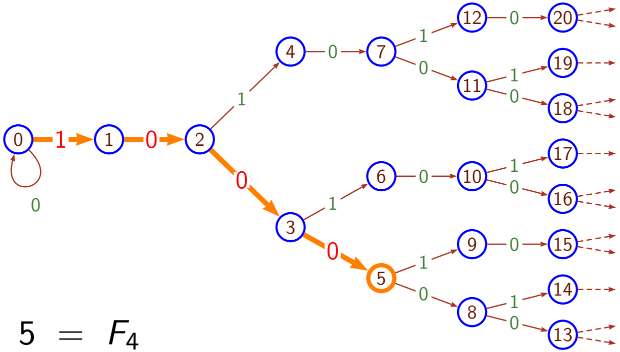


Figure : Integer representations in the Fibonacci numeration system.

Prefix-closed languages and labelled trees

Alphabets are ordered hence

prefix-closed languages $=$ labelled trees.

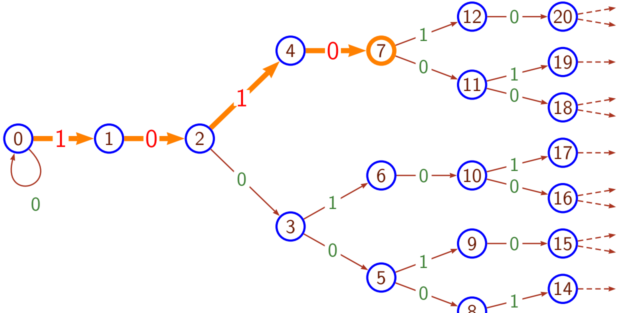


$5 = F_4$

Figure : Integer representations in the Fibonacci numeration system.

Alphabets are ordered hence

prefix-closed languages $=$ labelled trees.



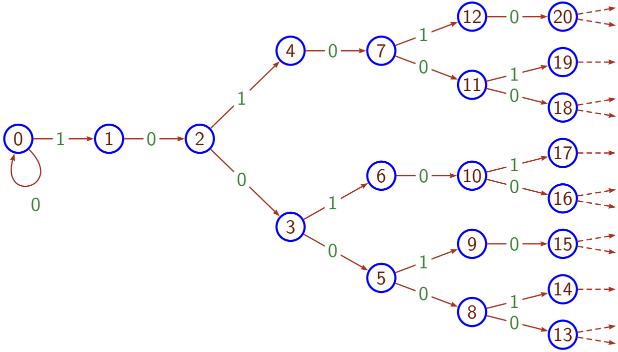
$$7 \;=\; 5 + 2 = F_4 + F_2$$

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** its transitions taken in breadth-first order.



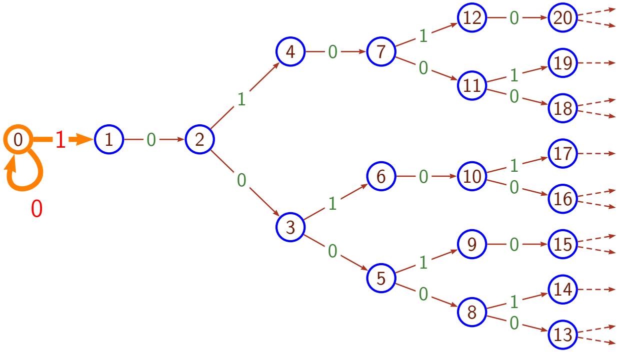
$\mathbf{s} =$

$\lambda =$

Serialisation of a prefix-closed language

Definition

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



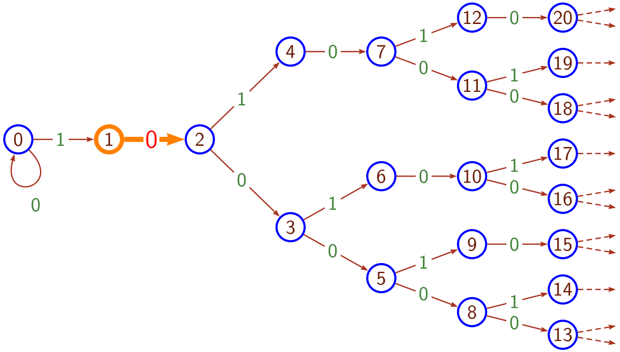
$\mathbf{s} = 2$

$\lambda = 01$

Serialisation of a prefix-closed language

Definition

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



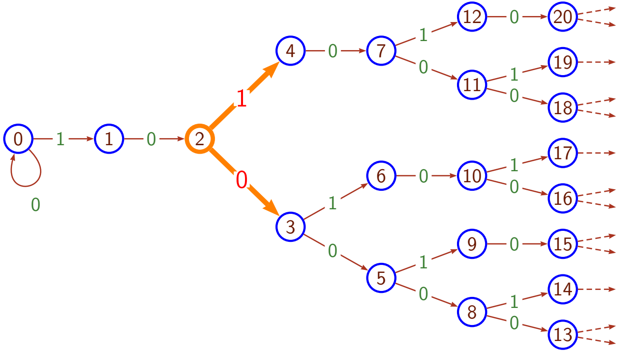
$\mathbf{s} = 2\ 1$

$\lambda = 01\ 0$

Serialisation of a prefix-closed language

Definition

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



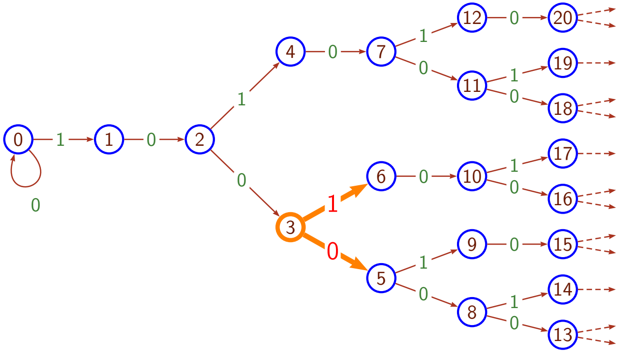
$$\mathbf{s} = 2 \ \ 1 \ \ 2$$

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

Serialisation of a prefix-closed language

Definition

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



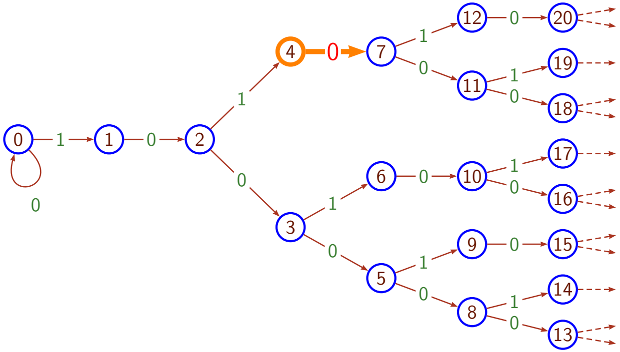
$\mathbf{s} = 2\ 1\ 2\ \ 2$

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

Serialisation of a prefix-closed language

Definition

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



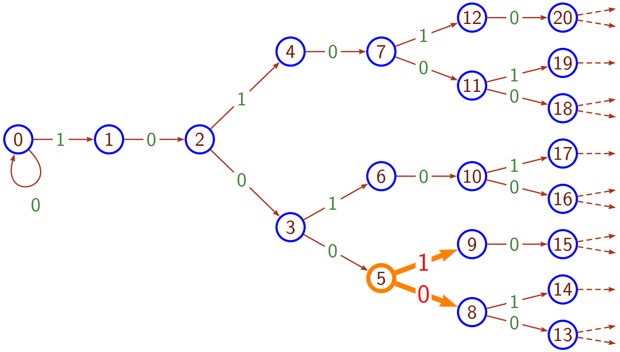
$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1$$

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

Serialisation of a prefix-closed language

Definition

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

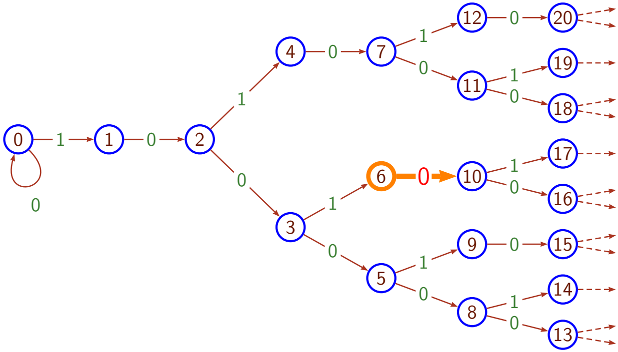


$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2$$

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

Definition

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

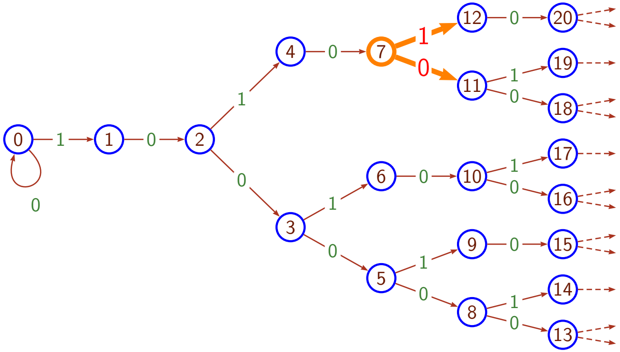


$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1$$
$$\lambda = 01\ 0\ 01\ 01\ 0\ 01\ 0$$

Serialisation of a prefix-closed language

Definition

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

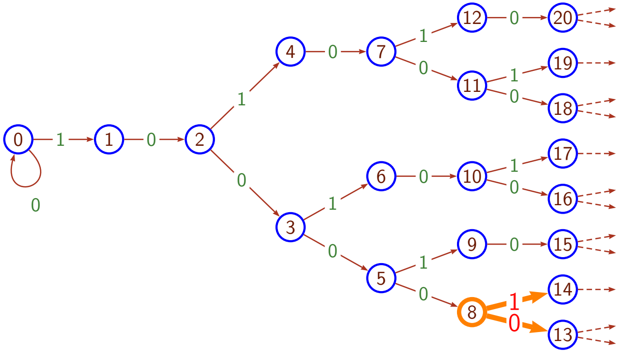


$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2$

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

Definition

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



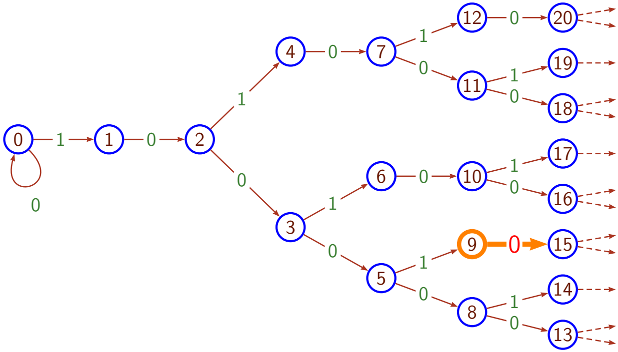
$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2\ \ 2$$

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

Serialisation of a prefix-closed language

Definition

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

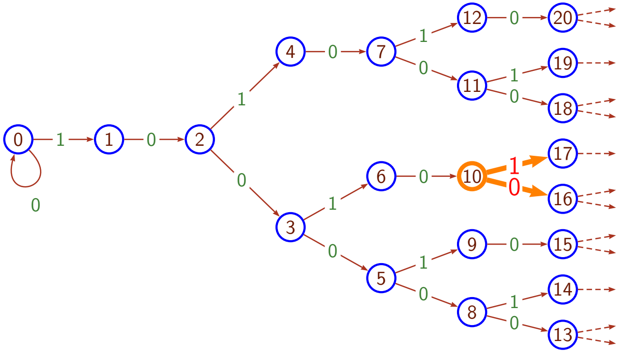


$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2\ \ 2\ 1$$

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

Definition

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



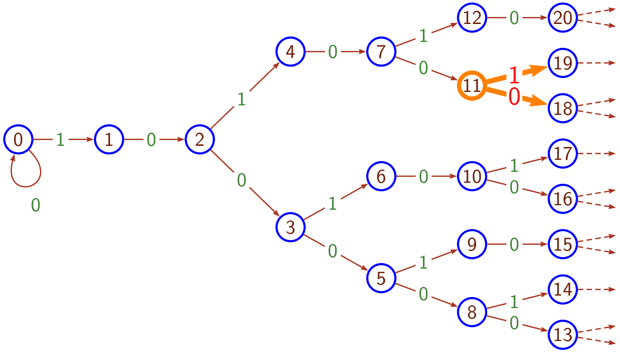
$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2\ \ 2\ 1\ 2$$

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

Serialisation of a prefix-closed language

Definition

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

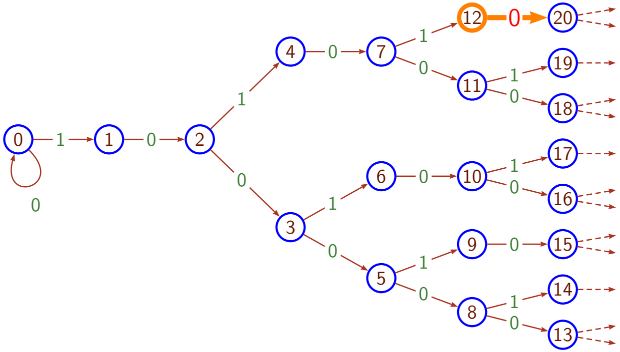


$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2\ \ 2\ 1\ 2\ \ 2$$

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

Definition

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



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

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

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

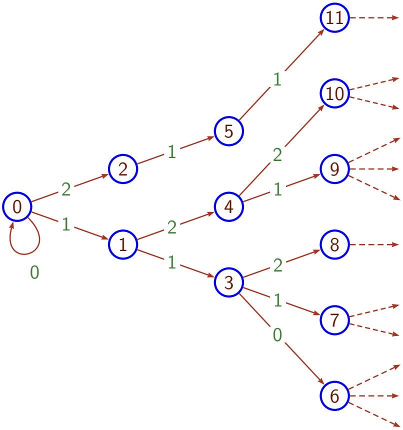
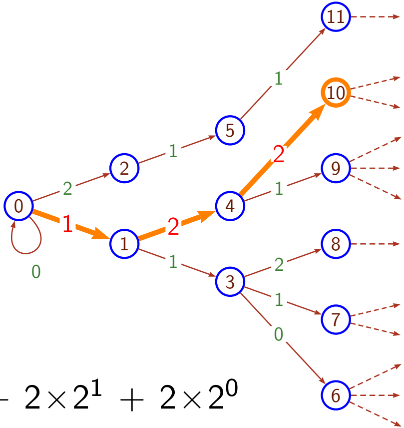


Figure : Non-canonical integer representations in base 2.

Generation of language by signature and labelling

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

$$\lambda = (012\ 12\ 1)^{\omega}$$



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

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.

Substitutive signature

A substitution σ is **prolongable on** a if $\sigma(a) = au$ for some u.

Definitions

Substitutive signature: $f_\sigma(\sigma^\omega(a))$ where $\forall b,\ f_\sigma(b) = |\sigma(b)|$

Substitutive signature

A substitution σ is **prolongable on** a if $\sigma(a) = au$ for some u.

Definitions

Substitutive signature: $f_\sigma(\sigma^\omega(a))$ where $\forall b,\ f_\sigma(b) = |\sigma(b)|$

Substitutive labelling: $g(\sigma^\omega(a))$

such that

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

Example 1 – the Fibonacci signature

$$\sigma(a) = ab \quad (f_\sigma(a) = 2)$$
$$\sigma(b) = a \quad (f_\sigma(b) = 1)$$
$$f_\sigma(\sigma^\omega(a)) = 21221212212212122\cdots$$

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

These signature/labelling define the language of integer representations in the Fibonacci numeration system.

Example 2 – a periodic signature

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

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

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

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

Example 2 – a periodic signature

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

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

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

Example 3 – the Thue-Morse morphism

$$\sigma(a) = ab \quad (f_\sigma(a) = 2)$$
$$\sigma(b) = ba \quad (f_\sigma(b) = 2)$$
$$f_\sigma(\sigma^\omega(a)) \quad = \quad 2^\omega$$

$\forall$ labelling g, the language is *essentially* $(0+1)^*$.

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.

Proposition

(σ, g): a substitutive signature defining a language L
L is accepted by the automaton $\mathcal{A}_{(\sigma,g)}$

Forward direction of the theorem

Theorem

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

Proposition

(σ, g): a substitutive signature defining a language L
L is accepted by the automaton $\mathcal{A}_{(\sigma,g)}$

This automaton is similar to

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

Automaton associated with a subst. signature

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

Definition: $\mathcal{A}_{(\sigma,g)}$

Set of states: A;
Alphabet: B;
Initial state: a;
Final states: A whole.

Automaton associated with a subst. signature

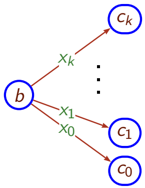
$\sigma : A^* \to A^*$ prolongable on a
$g : A^* \to B^*$

Definition: $\mathcal{A}_{(\sigma,g)}$

Set of states: A;
Alphabet: B;
Initial state: a;
Final states: A whole.

Transitions: $\sigma(b) = c_0 c_1 \cdots c_k$ and $g(b) = x_0 x_1 \cdots x_k$

$\forall i, 0 \leq i \leq k, \qquad b \xrightarrow{x_i} c_i$



Example 1 – the Fibonacci signature

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

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

Example 1 – the Fibonacci signature

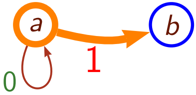
$\sigma(\text{a}) = \text{a}\,\text{b}$
$\sigma(\text{b}) = \text{a}$

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

Example 1 – the Fibonacci signature

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

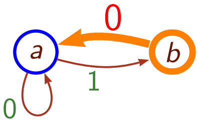
$g(\text{a}) = 0\,1$
$g(\text{b}) = 0$



Example 1 – the Fibonacci signature

$\sigma(\text{a}) = \text{a b}$
$\sigma(\text{b}) = \text{a}$

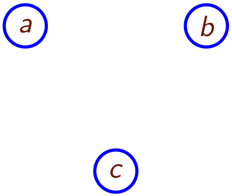
$g(\text{a}) = 0\,1$
$g(\text{b}) = 0$



Example 2 – a periodic signature

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

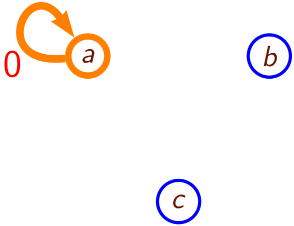
$g(\mathrm{a}) = 0\,1\,2$
$g(\mathrm{b}) = 1\,2$
$g(\mathrm{c}) = 1$



Example 2 – a periodic signature

$\sigma(\text{a}) = \text{a b c}$
$\sigma(\text{b}) = \text{a b}$
$\sigma(\text{c}) = \text{c}$

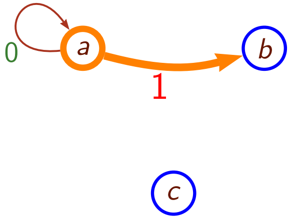
$g(\text{a}) = 0\,1\,2$
$g(\text{b}) = 1\,2$
$g(\text{c}) = 1$



Example 2 – a periodic signature

$$\sigma(a) = a\,b\,c$$
$$\sigma(b) = a\,b$$
$$\sigma(c) = c$$

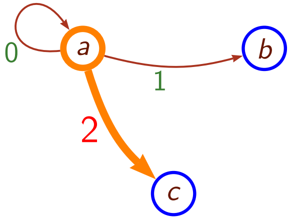
$$g(a) = 0\,1\,2$$
$$g(b) = 1\,2$$
$$g(c) = 1$$



Example 2 – a periodic signature

$\sigma(\text{a}) = \text{a b c}$
$\sigma(\text{b}) = \text{a b}$
$\sigma(\text{c}) = \text{c}$

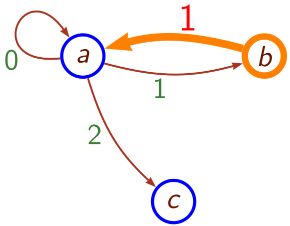
$g(\text{a}) = 0\,1\,2$
$g(\text{b}) = 1\,2$
$g(\text{c}) = 1$



Example 2 – a periodic signature

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

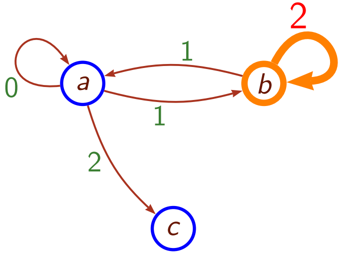
$g(\mathsf{a}) = 0\,1\,2$
$g(\mathsf{b}) = 1\,2$
$g(\mathsf{c}) = 1$



Example 2 – a periodic signature

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

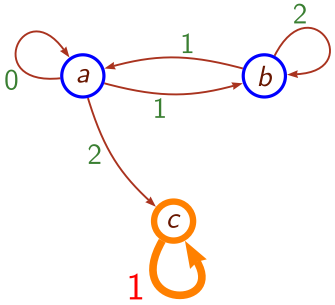
$g(\mathrm{a}) = 0\,1\,2$
$g(\mathrm{b}) = 1\,2$
$g(\mathrm{c}) = 1$



Example 2 – a periodic signature

$\sigma(\text{a}) = \text{a b c}$
$\sigma(\text{b}) = \text{a b}$
$\sigma(\text{c}) = \text{c}$

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



Abstract Numeration System (ANS, Lecomte-Rigo)

L: language over an ordered alphabet A.

Definition (Radix order $<_{\text{rad}}$)

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

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

Abstract Numeration System (ANS, Lecomte-Rigo)

L: language over an ordered alphabet A.

Definition (Radix order $<_{\text{rad}}$)

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

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

Definition (ANS L)

$\langle n \rangle_L$ is the $(n+1)$-th word of L.

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

Conversion function

Definition

L, K: two ANS's.

The conversion function $K \to L$: $\langle n \rangle_L \mapsto \langle n \rangle_K$.

Its complexity measures de relationship between K and L.

Example

- base p $\rightarrow$ base p^k is simple
  (ie. realised by a finite and sequential transducer).
- base $p \rightarrow$ base q is hard if $p \wedge q = 1$
  (ie. cannot be realised by a finite transducer).

Prefix-closed ARNS

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

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 $L \to K$ 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 $L \to K$ is very simple
(realised by a finite, *pure sequential* and *letter-to-letter* transducer).

The proof relies on a modified automata product.

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

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

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

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

Definition

$$A_\sigma \;=\; \{ \text{ strict prefixes of } \sigma(b) \mid b \in A\}$$

Example : $A_\sigma \;=\; \{\epsilon; a; ab\} \bigcup \{\epsilon; a\} \bigcup \{\epsilon\} \;=\; \{\epsilon; a; ab\}$

Dumont-Thomas Numeration System (DTNS)

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

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

Definition

$$A_\sigma \;=\; \{ \text{ strict prefixes of } \sigma(b) \mid b \in A\}$$

Example : $A_\sigma \;=\; \{\epsilon; a; ab\} \bigcup \{\epsilon; a\} \bigcup \{\epsilon\} \;=\; \{\epsilon; a; ab\}$

g_σ: morphism $A^* \to A_\sigma^*$

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

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

Example : $g_\sigma(a) = \epsilon, a, ab$ $g_\sigma(b) = \epsilon, a$ $g_\sigma(c) = \epsilon$

Dumont-Thomas Numeration System (DTNS)

Definition

ρ function ${A_\sigma}^* \to A^*$

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

Theorem (Dumont Thomas '89)

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

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

- $u_k \neq \epsilon$
- $|\rho(u_k, \ldots, u_1, u_0)| \;=\; n$

$(u_k, \ldots, u_1, u_0)$ is the representation of n in the DTNS.

Sum up

Prefix-closed ARNS	DTNS
Language accepted by a finite automaton	Substitution

Signature links an automaton with a substitution.

Sum up

Prefix-closed ARNS	DTNS
Language accepted by a finite automaton	Substitution

Signature links an automaton with a substitution.

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.

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

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

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

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

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

Future works : Directed signatures

Aperiodic signature: $\mathbf{s} = s_0\, s_1\, s_2 \cdots$

Definition: directed signature

$S_n = \Sigma_{k=0}^{n} s_k$: partial sums of $\mathbf{s}$.

$\mathbf{s}$ is directed by α if $S_n = \alpha\, n + o(1)$

Future works : Directed signatures

Aperiodic signature: $\mathbf{s} \; = \; s_0 \, s_1 \, s_2 \cdots$

Definition: directed signature

$S_n \; = \; \Sigma_{k=0}^{n} s_k$: partial sums of $\mathbf{s}$.

$\mathbf{s}$ is directed by α if $S_n = \alpha \, n + o(1)$

When α is Pisot, $\mathbf{s}$ generates a language linked to the base α.

When α is not Pisot... erratic behaviour.