MOTIVATION

- Understanding tree formalisms
- To construct reasonable fragments of XPATH with data
- To simplify formulas in verification.
Potthof’s example

Exists a maximal path of odd length

- “There is a maximal path of odd length” is definable in $\text{FO}[\leq, \text{left}, \text{right}]$ over ordered binary trees.
Potthof’s example: cont.

When the parity of depths of all leaves is the same

\[(\text{left} \cdot \text{right})^* \text{left path}.\]

When this is not the case

Find a maximal depth node where the property fails.

The formula expressing the property is not of the form “there is a path”.
**Boolean expressions**

**Positive boolean expressions**

Let $A = \{\lor, \land, 0, 1\}$, and consider the set of ordered binary trees that evaluate to 1 (and are well-formed).

![Binary tree diagram]

**Theorem (Heuter, Potthoff)**

*The above language is not definable in chain logic. (This language is aperiodic.)*
Words and their structure

- Concatenation is a privileged operation on $A^*$.
- The set of words $A^*$ with concatenation forms a monoid.

Recognizability: $\alpha : A^* \rightarrow S$ and $F \subseteq S$ such that $L = \alpha^{-1}(F)$.

Syntactic monoid of $L$: Smallest monoid recognizing $L$.

Aperiodicity

A semigroup $\langle S, \cdot \rangle$ is aperiodic iff there is $n$ such that $s^n = s^{n+1}$ for all $s \in S$.

Theorem (Schützenberger, McNaughton & Papert)

A regular language is FO definable iff its syntactic monoid is aperiodic.
A FOL definable language

Let $L = (ab)^* c\Sigma^*$. It is first-order definable.

The minimal automaton for $L$

![Automaton diagram](image)

Remarks

- The aperiodicity property is not that visible from the structure of the automaton.
- Every property of syntactic algebra is also a property of the minimal automaton. The converse is not true.
- This is good, as long as we know that properties of interest are properties of syntactic algebras.
What algebra gives us

- The structure gives new notions: idempotent, Green relations, . . .
- (pseudo) varieties. Limit the search space.
- Krohn-Rhodes decomposition theorem: A group free semigroup is a divisor of an iterated wreath product of $U_2$.
  - This implies LTL is equivalent to FO over words.
  - Decidability of Until-hierarchy for LTL [Therien, Wilke].
WHAT STRUCTURE FOR TREES

- $A$-algebras
- Pre-clones
- Forest algebras
- Seminearrings.
**A-algebras**

**Ranked alphabet**

\[ A_0, A_1, A_2, \ldots \]

**Free structure: ranked trees**

A letter in a node determines the number of children. These children are numbered.

Free \( A \)-algebra: trees with operations of adding a letter on a top.

**A-algebra**

\[ \langle S, a_0, b_2, c_3 \ldots \rangle \]

**Recognition**

\[ \varphi : \text{Free}(A) \rightarrow S, \text{ S-finite, } L = \varphi^{-1}(F) \]
**Recognition**

\( \varphi : \text{Free}(A) \rightarrow S, \ S\text{-finite}, \ L = \varphi^{-1}(F) \)

Syntactic algebra is the minimal leaves-to-root automaton.
**Preclones**

**Ranked alphabet**

\[ A_0, A_1, A_2, \ldots, id \in A_1. \]

**Free structure: multi-contexts**

Obtained from alphabet by substitution.

- \( AM_0 \) trees
- \( AM_1 \) contexts with one variable
- \( \ldots \)

Operation of substitution of multicontexts into multicontext.

**Pre-clone**

\[ S = \langle \{ S_i \}_{i=0,1,\ldots}, \{ \cdot \}_J, 1 \rangle \]

**Recognition**

\[ \varphi : MA \rightarrow S, S \text{ finitary}, F \subseteq S_0, L = \varphi^{-1}(F) \]
**Forest algebra**

**Unranked alphabet** $A$.

**Free structure**
- Forests over unranked alphabet
- Forest contexts with one hole.
- Operations:
  - concatenation of forests,
  - putting forest or a context into a hole of a context.

**Forest algebra** $\langle H, V, \text{act}, \text{in}_l, \text{in}_r \rangle$

- Two monoids: $H$ and $V$. We denote their operations by $+$ and $\cdot$, respectively.
- An action $\text{act} : H \times V \rightarrow H$. We write $vh$ for $\text{act}(h,v)$.
- Two operations $\text{in}_l, \text{in}_r : H \rightarrow V$.
- Axioms:
  - **ACTION** $(v \cdot w)h = v(wh)$;
  - **INSERTION** $\text{in}_l(g)h = g + h$ and $\text{in}_r(g)h = h + g$;
  - **FAITHFULNESS** for every two distinct $v, w \in V$ there is $h \in H$ with $vh \neq wh$;

**Recognizability**

$\varphi : A^\Delta \rightarrow \langle H, V, \text{act}, \text{in}_l, \text{in}_r \rangle$
**Seminearrings**

Unranked alphabet: $A$

**Free structure**
- Forest multicontexts
- Operations:
  - Horizontal concatenation
  - Vertical composition: simultaneous substitution.

**Seminearrings: $\langle N, 0, \Box, +, \cdot \rangle$**

Axioms:

\[(p + q) \cdot r = p \cdot r + q \cdot r\]

\[0 \cdot p = 0\]

...  

**Recognizability**

$\varphi: A^{\nu} \rightarrow \langle N, 0, \Box, +, \cdot \rangle$
Seminearrings forget more than forest algebras

“There is a $b$”

- $0$: “hopeless” multicontexts: arity 0 without an $b$.
- $\Box$: “possible” multicontexts: arity at least 1 without an $b$.
- $\infty$: “good” multicontexts: with $a$.

“Some node has label $b$ and no ancestor with label $c$”

- $0$: “hopeless” multicontexts.
- $\Box$: “possible” multicontexts: some hole with no $c$ ancestors.
- $\infty$: “good” multicontexts: in the language.

Only the first of the two languages is in EF.

Solution

Instead of properties of a seminearring, one looks at the properties of morphisms $\varphi : A^\nu \to \langle N, 0, \Box, +, \cdot \rangle$. 
VARIETIES OF FOREST ALGEBRAS
Varieties of forest algebras

Pseudovariety

A pseudovariety of finite forest algebras is a collection $\mathcal{V}$ of finite forest algebras with the following properties:

- $\mathcal{V}$ is closed under binary product.
  If $(H, V), (G, U) \in \mathcal{V}$ then $(H, V) \times (G, U) \in \mathcal{V}$.
- $\mathcal{V}$ is closed under faithful quotients of homomorphic images.
  If $(H, V) \in \mathcal{V}$ and $(H', V')$ is its homomorphic image, then $faith(H', V') \in \mathcal{V}$.
- $\mathcal{V}$ is closed under faithful quotients of subalgebras.
  If $(H, V) \in \mathcal{V}$ and $(H', V')$ is its subalgebra then $faith(H', V') \in \mathcal{V}$.

Theorem

If $\mathcal{V}$ is a pseudovariety iff it is ultimately defined by some set of equations.
First correspondence theorem

Varieties of forest languages

Let $\mathcal{V}$ be a variety of algebras. For every finite alphabet $A$ define

$$\mathcal{V}(A) = \{ L \subseteq H_A : (H^L, V^L) \in \mathcal{V} \}.$$ 

We call $\mathcal{V}$ the variety of forest languages associated to $\mathcal{V}$, and write

$$\mathcal{V} \mapsto \mathcal{V}.$$ 

Theorem

The mapping $\mathcal{V} \mapsto \mathcal{V}$ is one-to-one.
A prefix operator

If $L$ is a language of $A$-forests and $p$ is an $A$-context then we define

$$p^{-1}L = \{ t : pt \in L \}.$$ 

Theorem

Let $\mathcal{W}$ be an operator assigning to each $A$ a family $\mathcal{W}(A)$ of $A$-languages. $\mathcal{W}$ is a variety of languages if and only if the following hold:

1. every $\mathcal{W}(A)$ is closed under boolean operations.
2. if $L \in \mathcal{W}(A)$ and $p$ is an $A$-context then $p^{-1}L \in \mathcal{W}(A)$.
3. if $(\alpha, \beta) : A^\Delta \to B^\Delta$ is a homomorphism, and $L \in \mathcal{W}(B)$, then $\alpha^{-1}(L) \in \mathcal{W}(A)$. 

Igor Walukiewicz (LaBRI)
FREC 2010
October 2010 18 / 35
Consequences

**Fact**

$\text{FO}[\leq]$ is a variety of forest languages.

**Fact**

Chain logic is a variety of forest languages.

**Remark**

$\text{FO}[\text{succ}]$ is not a variety of forest languages. Closure under inverse homomorphic images fails.
WREATH PRODUCT
Wreath product

Wreath product: $(H_1, V_1) \circ (H_2, V_2)$

The product is $(H_1 \times H_2, V_1 \times V_2^{H_2})$:

- $(v_1, f)(h_1, h_2) = (v_1 h_1, f(h_1)v_2)$
- $(v, f) \cdot (v', f') = (vv', f'')$ where $f''(h) = (f(v'h)) \cdot (f'(h))$

- This is a standard definition of wreath product of transformation semigroups (but with actions on the left).
- All axioms are satisfied.
Theorem

For every pair
\[ \alpha : A^\Delta \to (G, W) \quad \beta : (A \times G)^\Delta \to (H, V) \]
there is a morphism into \((G, W) \circ (H, V)\) that is equal on forest to the sequential composition
\[ \alpha \otimes \beta : H_A \to G \times H \]
Every morphism \(A^\Delta \to (G, W) \circ (H, V)\) is realized this way.
Characterisations of language classes

$TL[\mathcal{A}]$

$TL[\mathcal{A}]$ is the class of languages recognized by iterated product of algebras from $\mathcal{A}$.

Theorem

We have the following:

- EF is $TL[\mathcal{U}_1]$
- CLT is $TL[\mathcal{U}_2]$
- CLT* is $TL[\text{distributive aperiodic algebras}]$
- PDL is $TL[\text{distributive algebras}]$
- FO[\leq] is $TL[\text{aperiodic path algebras}]$
- chain logic is $TL[\text{path algebras}]$
**Classes of algebras**

\[ \mathcal{U}_1 = (\{0, \infty\}, \{1, 0\}) \]

\[ 0 \cdot \infty = 0 \cdot 0 = \infty. \text{ Syntactic algebra of “some node with } b \text{”}. \]

\[ \mathcal{U}_2 = (\{0, \infty\}, \{1, c_0, c_\infty\}) \]

\[ C_h \cdot h' = h. \text{ Syntactic monoid of “some node with } b \text{ with ancestors only } c \text{”}. \]

**Distributive algebras**

- \( H \) commutative
- Action is distributive: \( v(h_1 + h_2) = vh_1 + vh_2 \)

**Path algebras**

- \( H \) aperiodic and commutative,
- \( vg + vh = v(g + h) + v0 \)
- \( u(g + h) = u(g + uh) \text{ if } u = u^2 \).
Infinite base

$A$ is a base for $L$ if $L = TL(A)$

If $L$ has a finite base then $L = TL((H, V))$.

**Theorem**

None of the language classes CTL*, FO[≤], PDL, or graded PDL has a finite base.

**Case of aperiodic $V$**

If a logic can define there is a path labelled $(a^n b)^* c$ then it does not have finite base.

**V has periods**

Languages there is a path $a^m b$ with $m$ a multiple of a fixed prime.
Decompositions

Theorem

A forest algebra divides and iterated wreath product of copies of $U_1$ iff $H$ is idempotent and commutative and $vh + h = vh$.

This is a characterisation of EF.
DEFINABILITY PROBLEM
The definability problem

Definition

Definability problem for a logic $\mathcal{L}$ is to decide if the language of a given automaton can be defined in $\mathcal{L}$.

What logics are known to have decidable definability problem

- Modal logic.
- EX, EF, EF+EX. [Bojańczyk & W.]
- FO[succ]. [Benedikt & Segoufin]
- EF+F^{-1}. [Bojańczyk]
- Boolean closure of the $\Sigma_1$ fragment of FO[≤_H, ≤_V] (an extension of Simon’s Theorem). [Bojańczyk & Segoufin & Straubing]
- $\Delta_2$ fragment of FO[≤_H, ≤_V]. [Bojańczyk & Segoufin]
**Confusions**

**Horizontal confusion**

$(H, V)$ has *horizontal confusion* if there is a multicontext $p$ and a set $G \subseteq H$ with $|G| > 1$ s.t. for every $g \in G$ and $x \in \text{holes}(p)$:

$$G \subseteq p[g/x][G].$$

**Example: Boolean expressions**

The diagram shows a Boolean expression tree with a root symbol $\land$, two subtrees $\lor$ and $\lor$, and leaves representing variables.

**Theorem**

Chain logic algebras do not have horizontal confusion
**Vertical confusion**

\((H, V)\) has *vertical confusion* if there is a multicontext \(p\) and a set \(\{g_0, \ldots, g_{k-1}\} \subseteq H\) with \(k > 1\) s.t. for every \(i = 0, \ldots, k - 1\):

\[
p[g_i] = g_j \text{ where } j = (i + 1) \pmod{k}.
\]

**Example:** Forests with a maximal path in \((ab)^*\)

**Example:** Binary trees with even path length

**Theorem**

- CTL algebras do not have vertical confusion.
- FO[\(\leq\)] algebras do not have vertical confusion for *uniform* multicontexts.
Comparing counting and non-counting logics

- CTL\(^*\) cannot count the number of successors. Neither can PDL.
- FO\([-\leq]\) is a counting correspondent of CTL\(^*\). Chain logics is that of PDL.

**Theorem**

Let \((H, V), (H_j, V_j), j = 1, \ldots, k\) be such that

- \(H\) is idempotent and commutative,
- each \((H_i, V_i)\) is a path algebra,
- \((H, V)\) divides \((H_1, V_1) \circ \cdots \circ (H_k, V_k)\).

then

- each \((H_i, V_i)\) has a **distributive** homomorphic image \((H'_i, V'_i)\)
- \((H, V)\) divides \((H'_1, V'_1) \circ \cdots \circ (H'_k, V'_k)\).
Reductions between different decision problems

UHV: Given a regular language \( L \) of unranked trees, decide if \( L \) can be defined in \( FO[\leq H, \leq V] \).

\( \uparrow \)

BHV: Given a regular language \( L \) of binary trees, decide if \( L \) can be defined in \( FO[\leq H, \leq V] \).

\( \downarrow \)

BV: Given a regular language \( L \) of binary trees, decide if \( L \) can be defined in \( FO[\leq V] \).

\( \downarrow \)

UV: Given a regular language \( L \) of unranked trees, decide if \( L \) can be defined in \( FO[\leq V] \).

\( \uparrow \)

CTL*: Given a regular language \( L \) of unranked trees, decide if \( L \) can be defined in CTL*.

Remark

If there is no vertical order than the problem is easy.
A tree decomposes into a “set of linear orders”.
ORDER INVARIANCE
We say that a property of trees is *order invariant* iff it does not depend on the horizontal order.
We say that a property of trees is **order invariant** iff it does not depend on the horizontal order.

A logic $\mathcal{L}$ is order invariant if every order invariant property definable in $\mathcal{L}$ is definable in $\mathcal{L}$ without the horizontal order.
We say that a property of trees is order invariant iff it does not depend on the horizontal order.

A logic $\mathcal{L}$ is order invariant if every order invariant property definable in $\mathcal{L}$ is definable in $\mathcal{L}$ without the horizontal order.

**Some facts**

- MSOL is order invariant over ranked trees.
- MSOL is not order invariant over unranked trees.
- MSOL order invariant $\equiv$ MSOL+counting modulo [Courcelle].
- FOL is not order invariant even over ranked trees.
- FOL[succ] is order invariant over ranked trees [Benedikt & Segoufin].
Objectives

- Decidable characterisations of logical fragments
  - Study confusions
  - CTL
- Understanding the power of order invariance
  - What is order invariant $\text{FO}[\leq_H, \leq_V]$?
- What is expressible with finite base
  - Beyond CTL.
- Profinite approach to varieties of trees
  - Delay theorem for trees.