A Class of Automata for Unranked Unordered Trees and Connection with TQL, MSO and PMSO Logics

Iovka Boneva, Jean-Marc Talbot and Sophie Tison

Laboratoire d'Informatique Fondamentale de Lille, France

MOSTRARE team of INRIA Futurs, France

Motivations and Related Work

- semistructured data can be represented as unranked unordered trees;
- in XML documents, attributes are unordered;
- different logics for unranked unordered trees:
  - monadic second-order logic (MSO),
  - Sheaves logic [DZL03], [DZLM04],
  - Presburger MSO [SSM03],
  - ambient logic [CG00] (TQL logic [CG01]),

Motivations and Related Work

- semistructured data can be represented as unranked unordered trees;
- in XML documents, attributes are unordered;
- different logics / automata for unranked unordered trees:
  - monadic second-order logic (MSO),
  - Sheaves logic / Sheaves automata [DZL03], [DZLM04],
  - Presburger MSO / Presburger automata [SSM03],
  - ambient logic / ? [CG00] (TQL logic [CG01]),

Open questions

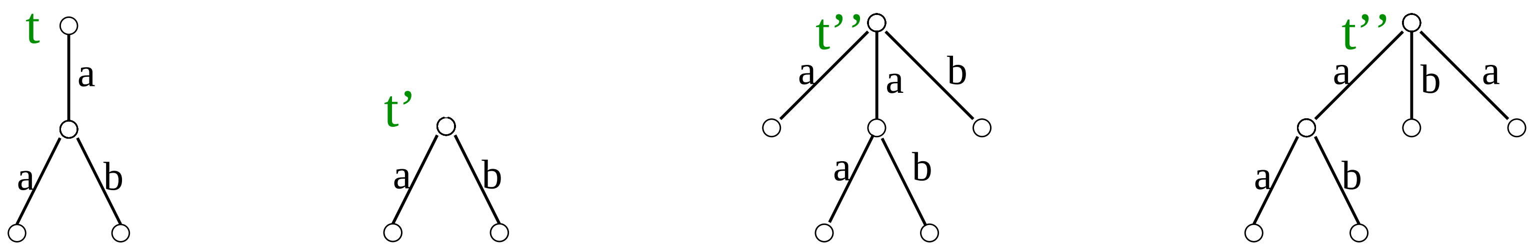
- automata for (fragments of) the TQL logic;
- exact relationship between the different logics;
- decidable fragments of TQL.

Outline

- Unranked and unordered trees
- Automata with numerical constraints
- Logics for unranked unordered trees
  - TQL logic
  - monadic second-order logic
  - Presburger monadic second-order logic
- Relationship between logics and automata

Unranked Unordered Trees

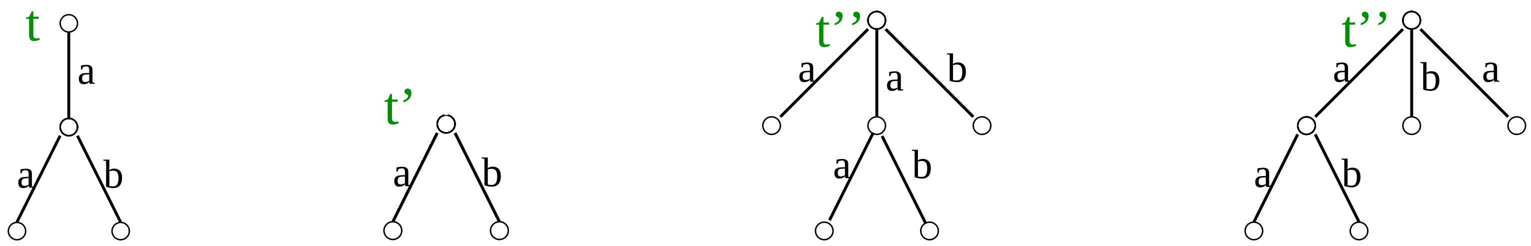
Unranked and unordered edge-labelled trees, finite set of edges Λ



t : element tree

Unranked Unordered Trees

Unranked and unordered edge-labelled trees, finite set of edges Λ



t : element tree

Operations on trees

- leaf $\circ$
- element tree $t = a[t']$
- composition $t'' = t' | t$

The set of trees is denoted Tree.

Automata with Numerical Constraints (1)

Close to Presburger automata [SSM03] , Sheaves automata [DZL03], [DZLM04] and Multitree automata [Lug03].

A run is a labelling of edges with states.

If the set of states is $\{q_1, q_2, \ldots, q_k\}$, then transitions are of the form

$$(\ (n_1, n_2, \ldots, n_k) : (q_1, q_2, \ldots, q_k),\ \ a\ ) \to q$$

$n_i \in \mathbb{N}, \quad a \in \Lambda$

Automata with Numerical Constraints (2)

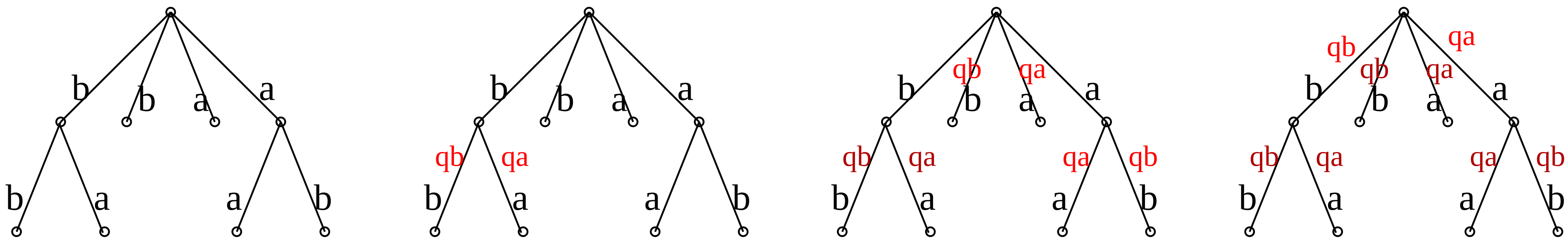
Automaton recognising trees having as many edges labelled by a as edges labelled by b under each node.

set of states	$\{qa, qb\}$
rules	$(\ (n, n) : (qa, qb)\ , a\ ) \rightarrow qa \quad \forall n \in \mathbb{N}$
rules	$(\ (n, n) : (qa, qb)\ , b\ ) \rightarrow qb \quad \forall n \in \mathbb{N}$

Automata with Numerical Constraints (2)

Automaton recognising trees having as many edges labelled by a as edges labelled by b under each node.

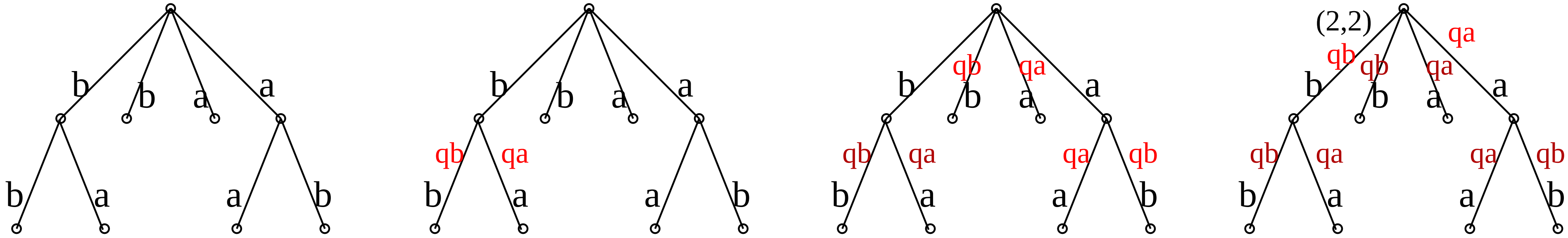
set of states	$\{qa, qb\}$
rules	$(\ (n,n) : (qa, qb)\ , a\ ) \rightarrow qa \quad \forall n \in \mathbb{N}$
rules	$(\ (n,n) : (qa, qb)\ , b\ ) \rightarrow qb \quad \forall n \in \mathbb{N}$



Automata with Numerical Constraints (2)

Automaton recognising trees having as many edges labelled by a as edges labelled by b under each node.

set of states	$\{qa, qb\}$
rules	$(\ (n,n):(qa,qb)\ ,a\ ) \to qa \quad \forall n \in \mathbb{N}$
rules	$(\ (n,n):(qa,qb)\ ,b\ ) \to qb \quad \forall n \in \mathbb{N}$
acceptance	$(n,n):(qa,qb) \quad \forall n \in \mathbb{N}$



Automata with Numerical Constraints (2)

Automaton recognising trees having as many edges labelled by a as edges labelled by b under each node.

set of states	$\{qa, qb\}$
rules	$(\ (n, n) : (qa, qb)\ , a\ ) \to qa \quad \forall n \in \mathbb{N}$
rules	$(\ (n, n) : (qa, qb)\ , b\ ) \to qb \quad \forall n \in \mathbb{N}$
acceptance	$(n, n) : (qa, qb) \quad \forall n \in \mathbb{N}$

$\mathcal{A} = \langle \Lambda, Q, \Delta, F \rangle$

- $Q = \{q_1, \ldots, q_k\}$ finite set of states
- $\Delta \subseteq \mathbb{N}^Q \times \Lambda \times Q$ transition relation
- $F \subseteq \mathbb{N}^Q$ acceptance condition

Representation of Numerical Constraints

- sets of vectors of intervals

  $\{\ ([2;4],[0;0]),\quad ([1;1],[0;\infty])\ \} = \{(2,0),(3,0),(4,0),(1,n)_{n\in\mathbb{N}}\}$

- semilinear sets

  finite union of linear sets
  $L \subseteq \mathbb{N}^k$ is linear iff $L = b + p_1\mathbb{N} + \cdots p_r\mathbb{N}$
  $b, p_1, \ldots, p_r$ in $\mathbb{N}^k$

  models of Presburger formulae [SSM03]

  $\{(n_1, \ldots, n_k) \mid f(n_1, \ldots, n_k)$ holds$\}$
  where $f(x_1, \ldots, x_k)$ is a Presburger formula

$$f ::= s_1 = s_2 \mid f_1 \wedge f_2 \mid \neg f \mid \forall x.f$$
$$p ::= x \mid n \mid s_1 + s_2$$

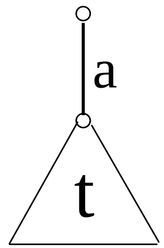
TQL Logic

- We consider a fragment of the TQL logic of [CG01]:
  - tree constructs,
  - Boolean connectives,
  - fixed point operator and star operator.
- The interpretation of the formula φ is $[\![\varphi]\!]_\delta \subseteq \mathsf{Tree}$ with $\delta : \xi \mapsto S \subseteq \mathsf{Tree}$ (valuation for free recursion variables).
- The tree t satisfies φ under valuation δ, $(t \models_\delta \varphi)$ if $t \in [\![\varphi]\!]_\delta$.

TQL Logic : Tree Constructs

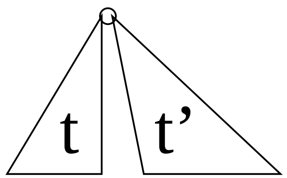
- $\mathbf{0}$ (leaf) $[\![\mathbf{0}]\!]_\delta = \mathbf{0}$
- $a[\varphi]$ (element tree)

$$[\![a[\varphi]]\!]_\delta = \left\{ \quad \text{(tree)} \quad \middle| \quad t \in [\![\varphi]\!]_\delta \right\}$$



- $\varphi \,|\, \psi$ (composition)

$$[\![\varphi \,|\, \psi]\!]_\delta = \left\{ \quad \text{(tree)} \quad \middle| \quad t \in [\![\varphi]\!]_\delta, \quad t' \in [\![\psi]\!]_\delta \right\}$$



φ : logic formula $\quad \xi$: recursion variable $\quad \delta : \xi \mapsto S \subseteq$ Tree

$[\![\varphi]\!]_\delta \subseteq$ Tree : interpretation of $\varphi \quad t \models_\delta \varphi$ iff $t \in [\![\varphi]\!]_\delta$

TQL Logic : Boolean Connectives

- $\top$ (true)

$$[\![\top]\!]_\delta = \text{Tree}$$

- $\neg\varphi$ (negation)

$$[\![\neg\varphi]\!]_\delta = \text{Tree} \setminus [\![\varphi]\!]_\delta$$

- $\varphi \vee \psi$ (disjunction)

$$[\![\varphi \vee \psi]\!]_\delta = [\![\varphi]\!]_\delta \cup [\![\psi]\!]_\delta$$

φ : logic formula $\quad \xi$: recursion variable $\quad \delta : \xi \mapsto S \subseteq \text{Tree}$

$[\![\varphi]\!]_\delta \subseteq \text{Tree}$: interpretation of $\varphi \quad t \models_\delta \varphi$ iff $t \in [\![\varphi]\!]_\delta$

TQL Logic : Recursion and Star Operator

- ξ (recursion variable)

$$[\![\xi]\!]_\delta = \delta(\xi)$$

- $\mu\xi.\varphi$ (least fixed point)

$$[\![\mu\xi.\varphi]\!]_\delta = \text{ the least } S \subseteq \mathsf{Tree} \text{ s.t. } S = [\![\varphi]\!]_{\delta[\xi \mapsto S]}$$

- $\varphi*$ (star)

$$\varphi* = \{\mathbf{0}\} \cup \bigcup_{n \in \mathbb{N}_*} \{\ t_1 | \cdots | t_n \quad | \quad t_i \in [\![\varphi]\!]_\delta \text{ for all } i \in 1..n\}$$

φ : logic formula $\quad$ ξ : recursion variable $\quad$ $\delta : \xi \mapsto S \subseteq \mathsf{Tree}$

$[\![\varphi]\!]_\delta \subseteq \mathsf{Tree}$: interpretation of φ $\quad$ $t \models_\delta \varphi$ iff $t \in [\![\varphi]\!]_\delta$

TQL Logic : Recursion and Star Operator

- ξ (recursion variable)

$$[\![\xi]\!]_\delta = \delta(\xi)$$

- $\mu\xi.\varphi$ (least fixed point)

$$[\![\mu\xi.\varphi]\!]_\delta = \text{ the least } S \subseteq \mathsf{Tree} \text{ s.t. } S = [\![\varphi]\!]_{\delta[\xi \mapsto S]}$$

- $\varphi*$ (star)

$$\varphi* = \{\mathbf{0}\} \cup \bigcup_{n \in \mathbb{N}_*} \{\, t_1 | \cdots | t_n \quad | \quad t_i \in [\![\varphi]\!]_\delta \text{ for all } i \in 1..n\}$$

$$\varphi* = \mu\xi.\varphi \,|\, \xi \vee \mathbf{0}$$

φ : logic formula $\qquad \xi$: recursion variable $\qquad \delta : \xi \mapsto S \subseteq \mathsf{Tree}$

$[\![\varphi]\!]_\delta \subseteq \mathsf{Tree}$: interpretation of $\varphi \qquad t \models_\delta \varphi$ iff $t \in [\![\varphi]\!]_\delta$

TQL Logic : Examples

- Path $a.b.c$ starting from the root

$$a[b[c[\top] \mid \top] \mid \top] \mid \top$$

TQL Logic : Examples

- Path $a.b.c$ starting from the root

$$a[b[c[\top] \mid \top] \mid \top] \mid \top$$

- At least two a-labelled edges and at most one b-labelled edge under the root node

$$a[\top] \mid a[\top] \quad \mid \quad \neg(b[\top] \mid b[\top] \mid \top)$$

TQL Logic : Examples

- Path $a.b.c$ starting from the root

$$a[b[c[\top] \mid \top] \mid \top] \mid \top$$

- At least two a-labelled edges and at most one b-labelled edge under the root node

$$a[\top] \mid a[\top] \quad \mid \quad \neg(b[\top] \mid b[\top] \mid \top)$$

- As many edges labelled by a as edges labelled by b under each node

$$\mu\xi.a[\xi] \mid b[\xi] \mid \xi \vee \mathbf{0}$$

Satisfiability of the TQL logic

Satisfiability is undecidable for a closed TQL formula.

Fragments of the TQL Logic

- $\mathrm{TQL}^*_{\mathrm{guarded}}$: guarded fragment for recursion
  For $\mu\xi.\varphi$, occurrences of ξ in φ are under an element operator

  $$\mu\xi.a[\xi]* \mid (b[\mathbf{0}] \vee \xi) \vee \mathbf{0} \qquad \mu\xi.a[b[\xi] \mid \xi \mid a[\mathbf{0}]] \vee \mathbf{0}$$

- $\mathrm{TQL}_{\mathrm{guarded}}$: guarded fragment without star operator
  Guarded for recursion and without star operator

- $\mathrm{TQL}_{\mathrm{pos}}$: positive fragment
  The logic without negation

Automata for the TQL Logic

- $\text{TQL}_{\text{guarded}} \equiv$ automata with numerical constraints as sets of vectors of intervals
- $\text{TQL}^*_{\text{guarded}} \equiv \text{TQL}_{\text{pos}} \equiv$ automata with numerical constraints as semilinear sets

MSO and PMSO Logics

Trees considered as graphs.

- MSO (monadic second-order) logic
  - domain: E the set of edges
  - signature: $\tau = \{\mathsf{label}_a \mid a \in \Lambda\} \cup \{<\}$

$\mathsf{label}_a(u)$	the edge u is labelled with b
$u < v$	the destination of u and the source of v coincide

- PMSO (Presburger monadic second-order) logic [SSM03]
  - formulae u/f
  - atomic Presburger formula $\#U$

$$u/(5 + \#U = \#U')$$

$u:$ edge variable $U:$ second order variable $f:$ Presburger formula

Automata for MSO and PMSO

- PMSO logic $\equiv$ automata with numerical constraints as semilinear sets [SSM03]
- MSO logic $\equiv$ automata with numerical constraints as sets of vectors of intervals

Summary

Ambient logic	Second-order logic	Automata
$TQL^*_{guarded}$, TQL_{pos}	PMSO	semilinear sets
$TQL_{guarded}$	MSO	vectors of intervals

References

[CG00] L. Cardelli and A. D. Gordon. Anytime, Anywhere: Modal Logics for Mobile Ambients. In *POPL'00*, pages 365–377. Springer, 2000.

[CG01] L. Cardelli and G. Ghelli. A Query Language Based on the Ambient Logic. In *ESOP'01*, volume 2028 of *LNCS*, pages 1–22. Springer, 2001.

[DZL03] S. Dal Zilio and D. Lugiez. XML Schema, Tree Logic and Sheaves Automata. In *RTA 2003*, volume 2706 of *LNCS*, pages 246–263. Springer, 2003.

[DZLM04] S. Dal Zilio, D. Lugiez, and C. Meyssonnier. A logic you can count on. In *POPL'04*, pages 135–146. ACM, 2004.

[Lug03] D. Lugiez. Counting and equality corstraints for multitree automata. In *FOSSACS 2003*, volume 2620 of *LNCS*, pages 328 – 342. Springer, 2003.

[SSM03] H. Seidl, T. Schwentick, and A. Muscholl. Numerical Document Queries. In *PODS 2003*, pages 155–166. ACM, 2003.