Compositions of Tree Series Transformations

Andreas Maletti [a]

Technische Universität Dresden

Fakultät Informatik

D–01062 Dresden, Germany

`maletti@tcs.inf.tu-dresden.de`

December 03, 2004

1. Motivation

2. Semirings, Tree Series, and Tree Series Substitution

3. Distributivity, Associativity, and Linearity

4. Tree Series Transducers and Composition Results

Motivation

Tree series transducers are a straightforward generalization of

(i) tree transducers, which are applied in

- syntax-directed semantics,
- functional programming, and
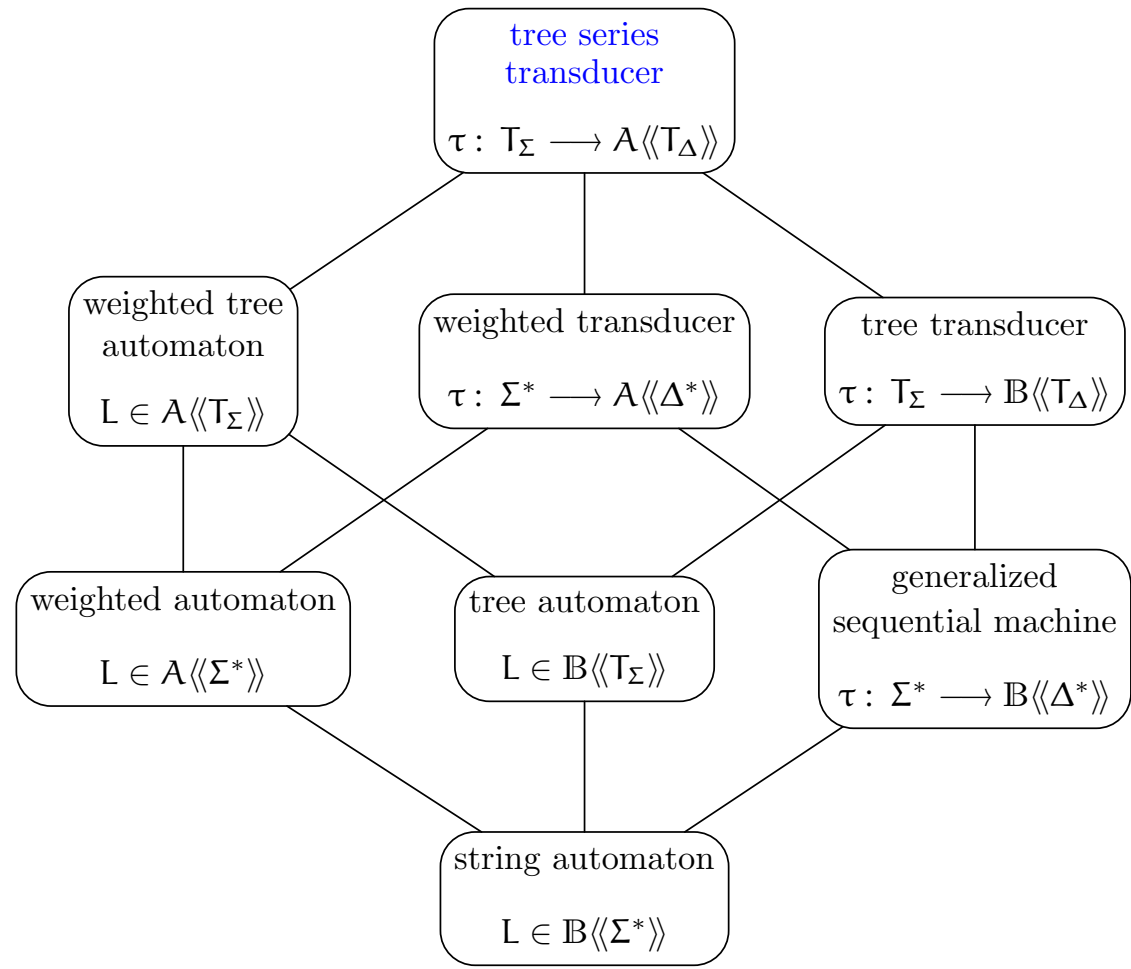- XML querying,

(ii) weighted automata, which are applied in

- (tree) pattern matching,
- image compression and speech-to-text processing.

Applications:

- can be used for code selection [Borchardt 04]
- potential uses in connection with tree banks

Generalization Hierarchy

tree series
transducer

$\tau : T_\Sigma \longrightarrow A\langle\!\langle T_\Delta \rangle\!\rangle$

weighted tree
automaton

$L \in A\langle\!\langle T_\Sigma \rangle\!\rangle$

weighted transducer

$\tau : \Sigma^* \longrightarrow A\langle\!\langle \Delta^* \rangle\!\rangle$

tree transducer

$\tau : T_\Sigma \longrightarrow \mathbb{B}\langle\!\langle T_\Delta \rangle\!\rangle$

weighted automaton

$L \in A\langle\!\langle \Sigma^* \rangle\!\rangle$

tree automaton

$L \in \mathbb{B}\langle\!\langle T_\Sigma \rangle\!\rangle$

generalized
sequential machine

$\tau : \Sigma^* \longrightarrow \mathbb{B}\langle\!\langle \Delta^* \rangle\!\rangle$

string automaton

$L \in \mathbb{B}\langle\!\langle \Sigma^* \rangle\!\rangle$

Trees

Σ ranked alphabet, $\Sigma_k \subseteq \Sigma$ symbols of rank k, $X = \{\, x_i \mid i \in \mathbb{N}_+ \,\}$

- $T_\Sigma(X)$ set of Σ-trees indexed by X,

- $T_\Sigma = T_\Sigma(\emptyset)$,

- $t \in T_\Sigma(X)$ is *linear* (resp., *nondeleting*) in $Y \subseteq X$, if every $y \in Y$ occurs at most (resp., at least) once in t,

- $t[t_1, \ldots, t_k]$ denotes the tree substitution of t_i for x_i in t

Semirings

A *semiring* is an algebraic structure $\mathcal{A} = (A, \oplus, \odot)$

- $(A, \oplus)$ is a commutative monoid with neutral element 0,

- $(A, \odot)$ is a monoid with neutral element 1,

- 0 is absorbing wrt. $\odot$, and

- $\odot$ distributes over $\oplus$.

Examples:

- semiring of non-negative integers $\mathbb{N}_\infty = (\mathbb{N} \cup \{\infty\}, +, \cdot)$

- Boolean semiring $\mathbb{B} = (\{0, 1\}, \vee, \wedge)$

- tropical semiring $\mathbb{T} = (\mathbb{N} \cup \{\infty\}, \min, +)$

- any ring, field, etc.

Properties of Semirings

We say that $\mathcal{A}$ is

- *commutative*, if $\odot$ is commutative,

- *idempotent*, if $a \oplus a = a$,

- *complete*, if there is an operation $\bigoplus_I : A^I \longrightarrow A$ such that

  1. $\bigoplus_{i \in I} a_i = a_{i_1} \oplus \cdots \oplus a_{i_n}$, if $I = \{i_1, \ldots, i_n\}$, and
  2. $\bigoplus_{i \in I} a_i = \bigoplus_{j \in J} \bigoplus_{i \in I_j} a_i$, if $I = \bigcup_{j \in J} I_j$ is a partition of I, and
  3. $\bigoplus_{i \in I} (a \odot a_i) = a \odot \bigoplus_{i \in I} a_i$ and $\bigoplus_{i \in I} (a_i \odot a) = \left(\bigoplus_{i \in I} a_i \right) \odot a$,

- *completely idempotent*, if it is complete with $\bigoplus_{i \in I} a = a$ for every non-empty I.

Semiring	Commutative	Idempotent	Complete	Completely Idempotent
$\mathbb{N}_\infty$	YES	no	YES	no
$\mathbb{B}$	YES	YES	YES	YES
$\mathbb{T}$	YES	YES	YES	YES

Tree Series

$\mathcal{A} = (A, \oplus, \odot)$ semiring, Σ ranked alphabet

Mappings $\varphi : T_\Sigma(X) \longrightarrow A$ are also called *tree series*

- the set of all tree series is $A\langle\!\langle T_\Sigma(X) \rangle\!\rangle$,

- the *coefficient* of $t \in T_\Sigma(X)$ in φ, i.e., $\varphi(t)$, is denoted by (φ, t),

- the *sum* is defined pointwise $(\varphi_1 \oplus \varphi_2, t) = (\varphi_1, t) \oplus (\varphi_2, t)$,

- the *support* of φ is $\mathrm{supp}(\varphi) = \{\, t \in T_\Sigma(X) \mid (\varphi, t) \neq 0 \,\}$,

- φ is *linear* (resp., *nondeleting* in $Y \subseteq X$), if $\mathrm{supp}(\varphi)$ is a set of trees, which are linear (resp., nondeleting in Y),

- the series φ with $\mathrm{supp}(\varphi) = \emptyset$ is denoted by $\widetilde{0}$.

Example: $\varphi = 1\,\alpha + 1\,\beta + 3\,\sigma(\alpha, \alpha) + \ldots + 3\,\sigma(\beta, \beta) + 5\,\sigma(\alpha, \sigma(\alpha, \alpha)) + \ldots$

Tree Series Substitution

$\mathcal{A} = (A, \oplus, \odot)$ complete semiring, $\varphi, \psi_1, \ldots, \psi_k \in A\langle\!\langle T_\Sigma(X) \rangle\!\rangle$

Pure substitution of $(\psi_1, \ldots, \psi_k)$ into φ:

$$\varphi \longleftarrow (\psi_1, \ldots, \psi_k) = \bigoplus_{\substack{t \in \mathrm{supp}(\varphi), \\ (\forall i \in [k]): t_i \in \mathrm{supp}(\psi_i)}} (\varphi, t) \odot (\psi_1, t_1) \odot \cdots \odot (\psi_k, t_k)\, t[t_1, \ldots, t_k]$$

o-*substitution* of $(\psi_1, \ldots, \psi_k)$ into φ:

$$\varphi \overset{o}{\longleftarrow} (\psi_1, \ldots, \psi_k) = \bigoplus_{\substack{t \in \mathrm{supp}(\varphi), \\ (\forall i \in [k]): t_i \in \mathrm{supp}(\psi_i)}} (\varphi, t) \odot (\psi_1, t_1)^{|t|_{x_1}} \odot \cdots \odot (\psi_k, t_k)^{|t|_{x_k}}\, t[t_1, \ldots, t_k]$$

Example: $5\,\sigma(x_1, x_1) \longleftarrow (2\,\alpha) = 10\,\sigma(\alpha, \alpha)$ and $5\,\sigma(x_1, x_1) \overset{o}{\longleftarrow} (2\,\alpha) = 20\,\sigma(\alpha, \alpha)$

Distributivity

$$\left(\bigoplus_{i \in I} \varphi_i \right) \xleftarrow{\ m\ } \left(\bigoplus_{i_1 \in I_1} \psi_{1 i_1}, \ldots, \bigoplus_{i_k \in I_k} \psi_{k i_k} \right) = \bigoplus_{\substack{i \in I, \\ (\forall j \in [k]):\, i_j \in I_j}} \varphi_i \xleftarrow{\ m\ } (\psi_{1 i_1}, \ldots, \psi_{k i_k})$$

Substitution	Sufficient condition for distributivity
pure substitution	always
o-substitution	φ_i linear, $\mathcal{A}$ completely idempotent
OI-substitution	φ_i linear and nondeleting [Kuich 99]

Associativity

$$\left(\varphi \xleftarrow{m}(\psi_1,\ldots,\psi_k)\right)\xleftarrow{m}(\tau_1,\ldots,\tau_n) = \varphi\xleftarrow{m}(\psi_1\xleftarrow{m}(\tau_1,\ldots,\tau_n),\ldots,\psi_k\xleftarrow{m}(\tau_1,\ldots,\tau_n))$$

Substitution	Sufficient condition for associativity
pure substitution	$\mathcal{A}$ completely idempotent, conforming partition $(I_j)_{j\in[k]}$ of $[n]$, τ_i boolean
	τ_i singletons with idempotent coefficients
	special associativity law
o-substitution	$\varphi, \psi_1,\ldots,\psi_k$ linear, $\mathcal{A}$ zero-divisor free and completely idempotent
	φ linear, $\mathcal{A}$ zero-divisor free, τ_i singletons
OI-substitution	φ_i linear and nondeleting [Kuich 99]

Special associativity law: partition $(I_j)_{j\in J}$ of I with $\mathrm{var}(\psi_j) \subseteq X_{I_j}$ for every $j \in J$

$$\left(\varphi \longleftarrow (\psi_j)_{j\in J}\right) \longleftarrow (\tau_i)_{i\in I} = \varphi \longleftarrow \left(\psi_j \longleftarrow (\tau_i)_{i\in I_j}\right)_{j\in J}$$

Linearity

$$(a \odot \varphi) \xleftarrow{\;m\;} (a_1 \odot \psi_1, \ldots, a_k \odot \psi_k) = a \odot a_1 \odot \cdots \odot a_k \odot \varphi \xleftarrow{\;m\;} (\psi_1, \ldots, \psi_k)$$

Substitution	Sufficient condition for distributivity
pure substitution	always
o-substitution	$a_i \in \{0, 1\}$ or special linearity law
OI-substitution	φ_i linear and nondeleting [Kuich 99]

Special linearity law: tree $t \in T_\Sigma(X_k)$

$$(a\,t) \xleftarrow{\;o\;} (a_1 \odot \psi_1, \ldots, a_k \odot \psi_k) = a \odot a_1^{|t|_1} \odot \cdots \odot a_k^{|t|_k} \odot \left(t \xleftarrow{\;o\;} (\psi_1, \ldots, \psi_k) \right)$$

Tree Series Transducers

Definition: A *(bottom-up) tree series transducer* (tst) is a system
$M = (Q, \Sigma, \Delta, \mathcal{A}, F, \mu)$

- Q is a non-empty set of *states*,

- Σ and Δ are input and output ranked alphabets,

- $\mathcal{A} = (A, \oplus, \odot)$ is a complete semiring,

- $F \in A\langle\!\langle T_\Delta(X_1)\rangle\!\rangle^Q$ is a vector of *final outputs*,

- $\mu = (\mu_k)_{k \in \mathbb{N}}$ with $\mu_k : \Sigma_k \longrightarrow A\langle\!\langle T_\Delta(X_k)\rangle\!\rangle^{Q \times Q^k}$.

If Q is finite and $\mu_k(\sigma)_{q,\vec{q}}$ is polynomial, then M is called *finite*.

Semantics of Tree Series Transducers

$m \in \{\varepsilon, o\}$, $q \in Q$, $t \in T_\Sigma$, $\varphi \in A\langle\!\langle T_\Sigma \rangle\!\rangle$

Definition: The mapping $h_\mu^m : T_\Sigma \longrightarrow A\langle\!\langle T_\Delta \rangle\!\rangle^Q$ is defined as

$$h_\mu^m(\sigma(t_1, \ldots, t_k))_q = \bigoplus_{q_1, \ldots, q_k \in Q} \mu_k(\sigma)_{q,(q_1, \ldots, q_k)} \xleftarrow{m} (h_\mu^m(t_1)_{q_1}, \ldots, h_\mu^m(t_k)_{q_k})$$

and $h_\mu^m(\varphi)_q = \bigoplus_{t \in T_\Sigma} (\varphi, t) \odot h_\mu^m(t)_q$.

- the m-*tree-to-tree-series transformation* $\|M\|^m : T_\Sigma \longrightarrow A\langle\!\langle T_\Delta \rangle\!\rangle$ computed by M is $(\|M\|^m, t) = \bigoplus_{q \in Q} F_q \xleftarrow{m} (h_\mu^m(t)_q)$ and

- the m-*tree-series-to-tree-series transformation* $|M|^m : A\langle\!\langle T_\Sigma \rangle\!\rangle \longrightarrow A\langle\!\langle T_\Delta \rangle\!\rangle$ computed by M is $(|M|^m, \varphi) = \bigoplus_{t \in T_\Sigma} (\varphi, t) \odot (\|M\|^m, t)$.

Extension

$(Q, \Sigma, \Delta, \mathcal{A}, F, \mu)$ tree series transducer, $m \in \{\varepsilon, o\}$, $\vec{q} \in Q^k$, $q \in Q$, $\varphi \in A\langle\!\langle T_\Sigma(X_k)\rangle\!\rangle$

Definition: We define $h_{\mu,m}^{\vec{q}} : T_\Sigma(X_k) \longrightarrow A\langle\!\langle T_\Delta(X_k)\rangle\!\rangle^Q$

$$h_{\mu,m}^{\vec{q}}(x_i)_q = \begin{cases} 1\,x_i & \text{, if } q = q_i \\ \widetilde{0} & \text{, otherwise} \end{cases}$$

$$h_{\mu,m}^{\vec{q}}(\sigma(t_1, \ldots, t_k))_q = \bigoplus_{p_1,\ldots,p_k \in Q} \mu_k(\sigma)_{q,p_1\ldots p_k} \overset{m}{\longleftarrow} (h_{\mu,m}^{\vec{q}}(t_1)_{p_1}, \ldots, h_{\mu,m}^{\vec{q}}(t_k)_{p_k})$$

We define $h_{\mu,m}^{\vec{q}} : A\langle\!\langle T_\Sigma(X_k)\rangle\!\rangle \longrightarrow A\langle\!\langle T_\Delta(X_k)\rangle\!\rangle^Q$ by

$$h_{\mu,m}^{\vec{q}}(\varphi)_q = \bigoplus_{t \in T_\Sigma(X_I)} (\varphi, t) \odot h_{\mu,m}^{\vec{q}}(t)_q$$

Composition Construction

$M_1 = (Q_1, \Sigma, \Delta, \mathcal{A}, F_1, \mu_1)$ and $M_2 = (Q_2, \Delta, \Gamma, \mathcal{A}, F_2, \mu_2)$ tree series transducer

Definition: The m-*product of* M_1 and M_2, denoted by $M_1 \cdot_m M_2$, is the tree series transducer

$$M = (Q_1 \times Q_2, \Sigma, \Gamma, \mathcal{A}, F, \mu)$$

- $F_{pq} = \bigoplus_{i \in Q_2} (F_2)_i \xleftarrow{m} h^q_{\mu_2, m}((F_1)_p)_i$

- $\mu_k(\sigma)_{pq,(p_1 q_1, \ldots, p_k q_k)} = h^{q_1 \ldots q_k}_{\mu_2; m}((\mu_1)_k(\sigma)_{p, p_1 \ldots p_k})_q.$

Main Theorem

$\mathcal{A}$ commutative and complete semiring, M_1 and M_2 tree series transducer

Theorem: $|M_1 \cdot_m M_2|^m = |M_1|^m \circ |M_2|^m$, if

- $m = \varepsilon$ and M_1 linear, or

- $m = \varepsilon$ and M_2 boolean and deterministic, or

- $m = \varepsilon$ and M_2 is deterministic and $\mathcal{A}$ is multiplicatively idempotent, or

- $m = o$ and M_1 is linear, M_2 is nondeleting and linear, and $\mathcal{A}$ is completely idempotent.

Main Corollary

Corollary:

- $\mathsf{l\text{-}BOT_{ts\text{-}ts}}(\mathcal{A}) \circ \mathsf{BOT_{ts\text{-}ts}}(\mathcal{A}) = \mathsf{BOT_{ts\text{-}ts}}(\mathcal{A})$.

- $\mathsf{BOT_{ts\text{-}ts}}(\mathcal{A}) \circ \mathsf{db\text{-}BOT_{ts\text{-}ts}}(\mathcal{A}) = \mathsf{BOT_{ts\text{-}ts}}(\mathcal{A})$,

- $\mathsf{BOT_{ts\text{-}ts}}(\mathcal{A}) \circ \mathsf{d\text{-}BOT_{ts\text{-}ts}}(\mathcal{A}) = \mathsf{BOT_{ts\text{-}ts}}(\mathcal{A})$, provided that $\mathcal{A}$ is multiplicatively idempotent,

- $\mathsf{l\text{-}BOT_{ts\text{-}ts}^{o}}(\mathcal{A}) \circ \mathsf{nl\text{-}BOT_{ts\text{-}ts}^{o}}(\mathcal{A}) = \mathsf{l\text{-}BOT_{ts\text{-}ts}^{o}}(\mathcal{A})$, provided that $\mathcal{A}$ is completely idempotent.

References

[Borchardt 04]   B. Borchardt: *Code Selection by Tree Series Transducers*. CIAA'04, Kingston, Canada, 2004. to appear

[Engelfriet et al 02]   J. Engelfriet, Z. Fülöp, and H. Vogler: *Bottom-up and Top-down Tree Series Transformations*. Journal of Automata, Languages, and Combinatorics 7:11–70, 2002

[Fülöp et al 03]   Z. Fülöp and H. Vogler: *Tree Series Transformations that Respect Copying*. Theory of Computing Systems 36:247–293, 2003

[Kuich 99]   W. Kuich: *Tree Transducers and Formal Tree Series*. Acta Cybernetica 14:135–149, 1999