Word automaticity of tree automatic ordinals is decidable

Martin Huschenbett

Institute of Theoretical Computer Science
Ilmenau University of Technology
Germany

September 12th, 2011
Theorem (H 2011)

1. Given a tree automatic presentation of an ordinal $\alpha$, it is decidable whether $\alpha$ is word automatic.
2. If $\alpha$ is tree automatic, then a word automatic presentation of $\alpha$ can be computed from the tree automatic presentation.

Proof sketch.

1. From a tree automatic presentation with bounded level width one can compute a word automatic presentation.
2. If the level width of a tree automatic presentation of an ordinal unbounded, then the ordinal is not word automatic.
3. It is decidable whether the level width of a given tree automatic presentation is (un)bounded.
Trees

Definition
Let $\Sigma$ be an alphabet. A $\Sigma$-tree, or just tree, is a map $t : D \to \Sigma$, where $D = \text{dom}(t)$ is a non-empty, finite, prefix-closed subset of $\{0, 1\}^*$. The set of all $\Sigma$-trees is denoted with $T_\Sigma$.

Two $\Sigma$ trees for $\Sigma = \{a, b, c\}$:
Tree automata and regular tree languages

**Definition**

A (deterministic bottom-up) **tree automaton** $A = (Q, \iota, \delta, F)$ (over $\Sigma$) consists of

- a non-empty, finite set of states $Q$
- an initial state $\iota \in Q$
- a transition function $\delta: Q \times Q \times \Sigma \rightarrow Q$
- a set $F \subseteq Q$ of accepting states

The state which is reached (at the root) after processing a tree $t \in T_\Sigma$ is denoted with $\delta(t)$.

**Definition**

A tree language $L \subseteq T_\Sigma$ is **regular** if it can be **accepted** by some tree automaton.
Convolution of trees

Definition

Let $\Box \notin \Sigma$ and $\Sigma_{\Box} = \Sigma \cup \{\Box\}$. For $t_1, t_2 \in T_\Sigma$ we define a tree $t_1 \otimes t_2 \in T_{\Sigma^2}$ as usual.
Definition

A relation $R \subseteq T_\Sigma \times T_\Sigma$ is regular if the tree language

$$\{ t_1 \otimes t_2 \mid (t_1, t_2) \in R \} \subseteq T_{\Sigma^2}$$

is regular. A tree automaton over $\Sigma^2$ which accepts this tree language is said to accept the relation $R$. 

---

Regular tree relations
Tree automatic ordinals

**Definition**

An ordinal $\alpha$ is **tree automatic** if there are

- an alphabet $\Sigma$,
- a regular tree language $L \subseteq T_\Sigma$, and
- a regular tree relation $R \subseteq L \times L$

such that $(L, R)$ is a (strict) well-ordering of type $\alpha$.

The pair $(L, R)$ is a **tree automatic copy** of $\alpha$.

A pair $(A, A_\prec)$ consisting of

- a tree automaton $A$ over $\Sigma$ accepting $L$ and
- a tree automaton $A_\prec$ over $\Sigma^2$ accepting $R$

is a **tree automatic presentation** of $\alpha$ respectively of $(L, R)$. 
The level width

Definition

The **level width** of a tree \( t \in T_\Sigma \) is

\[ \ell(t) = \max\{ | \text{dom}(t) \cap \{0, 1\}^d | \mid d \geq 0 \} \in \mathbb{N}. \]

The **level width** of a tree language \( L \subseteq T_\Sigma \) is

\[ \ell(L) = \sup\{ \ell(t) \mid t \in L \} \in \mathbb{N} \cup \{\infty\}. \]

If \( \ell(L) < \infty \), then \( L \) has **bounded level width**, otherwise \( L \) has **unbounded level width**.

\[ \ell \left( \begin{array}{ccc}
  b & a & c \\
  c & b & b \\
  a & b & c \\
\end{array} \right) = 4 \]

\[ \ell \left( \begin{array}{ccc}
  a & b & c \\
  b & b & b \\
  a & a & c \\
\end{array} \right) = 3 \]
Proposition 1

1. Let \((L, <)\) be a tree automatic copy of an ordinal \(\alpha\) such that \(L\) has bounded level width. Then, \(\alpha\) is word automatic.

2. From a tree automatic presentation of \(\alpha\) one can compute a word automatic presentation of \(\alpha\).
Encoding trees by words

Elements of $\Sigma \times 2\{0,1\}$

Elements of $\bigcup_{1 \leq \ell \leq 4} (\Sigma \times 2\{0,1\})^\ell$
Encoding trees by words

Elements of $\Sigma \times 2^{\{0,1\}}$

Elements of $\bigcup_{1 \leq \ell \leq 4} \left( \Sigma \times 2^{\{0,1\}} \right) ^\ell$
Encoding trees by words

Definition

A tree \( t \in T_\Sigma \) with level width \( k \) is encoded by a word

\[
C(t) \in \Gamma_k^*
\]

over the alphabet

\[
\Gamma_k = \bigcup_{1 \leq \ell \leq k} \left( \Sigma \times 2^{\{0,1\}} \right)^\ell.
\]
Encoding tree languages by word languages

Definition
A tree language \( L \subseteq T_\Sigma \) with level width \( k < \infty \) is encoded by the word language
\[
C(L) = \{ C(t) \mid t \in L \} \subseteq \Gamma_k^*.
\]

Lemma
1. If \( L \) is a regular, then \( C(L) \) is regular.
2. From a tree automaton accepting \( L \) one can compute a (non-deterministic) finite automaton accepting \( C(L) \).

Proof.
Let \( A = (Q, \iota, \delta, F) \) be a tree automaton accepting \( L \).
We construct a finite automaton accepting \( C(L) \) with state space
\[
\bigcup_{0 \leq m \leq k} Q^m.
\]
The transition $\overline{p} \xrightarrow{\overline{A}} \overline{q}$ exists precisely if $m = \ell$, $n$ “fits” $\overline{A}$, and for all $i = 1, \ldots, \ell$ one of the four conditions is met:

1. $A_i = \begin{bmatrix} a_i \\ 0 \end{bmatrix}$ and $p_i = \delta(q_j, q_{j+1}, a_i)$
2. $A_i = \begin{bmatrix} a_i \\ 0 \end{bmatrix}$ and $p_i = \delta(q_j, \iota, a_i)$
3. $A_i = \begin{bmatrix} a_i \\ 1 \end{bmatrix}$ and $p_i = \delta(\iota, q_{j+1}, a_i)$
4. $A_i = \begin{bmatrix} a_i \\ 1 \end{bmatrix}$ and $p_i = \delta(\iota, \iota, a_i)$

where $j$ is suitable (like in the picture).

Initial are the states $\begin{bmatrix} f \end{bmatrix}$ for $f \in F$.

Accepting is only the single state from $Q^0$. 
Encoding of tree relations by word relations

Definition

Let $L \subseteq T_\Sigma$ be a tree language with bounded level width. A tree relation $R \subseteq L \times L$ is encoded by the word relation

$$C(R) = \{ (C(t_1), C(t_2)) \mid (t_1, t_2) \in R \} \subseteq C(L) \times C(L).$$

Lemma

1. If $R$ is regular, then $C(R)$ is regular.
2. From a tree automaton accepting $R$ one can compute a (non-deterministic) finite automaton accepting $C(R)$.

Proof.

Similar to the proof for tree languages, but a bit more involved.
Let $A = (Q, \iota, \delta, F)$ be a tree automaton accepting $R$. States $p, q_1, q_2$ are from $Q$.

$$p = \delta(q_1, q_2, (a, b))$$

For each $f \in F$ the state below is initial.
Proof of Proposition 1

Proposition 1

1. Let $(L, <)$ be a tree automatic copy of an ordinal $\alpha$ such that $L$ has bounded level width. Then, $\alpha$ is word automatic.

2. From a tree automatic presentation of $\alpha$ one can compute a word automatic presentation of $\alpha$.

Proof.

- The encoding of trees as words is injective.
- $(C(L), C(<))$ is word automatic copy of $\alpha$, proving 1.
- From a tree automatic presentation $(A, A_<)$ of $(L, <)$ one can compute finite automata $M$ and $M_<$ which accept $C(L)$ and $C(<)$, respectively.
- $(M, M_<)$ is a word automatic presentation of $\alpha$, showing 2. $\Box$
Proposition 2

Let \((L, \prec)\) be a tree automatic copy of an ordinal \(\alpha\) such that \(L\) has unbounded level width. Then, \(\alpha\) is not word automatic.

Theorem (Delhommé 2001)

An ordinal \(\alpha\) is word automatic if, and only if, \(\alpha < \omega^\omega\).

Proposition 2’

Let \((L, \prec)\) be a tree automatic copy of an ordinal \(\alpha\) such that \(L\) has unbounded level width. Then,

\[ \alpha \geq \omega^\omega. \]
Proposition 2’

Let \((L, <)\) be a tree automatic copy of an ordinal \(\alpha\) such that \(L\) has unbounded level width. Then,

\[
\alpha \geq \omega^\omega.
\]

Lemma

Let \(n, r \geq 1\) and \((L, <)\) be a tree automatic copy of an ordinal \(\alpha\) such that \(L\) can be accepted by a tree automaton with \(n\) states and the level width of \(L\) is at least \(r \cdot 2^n\). Then,

\[
\alpha \geq \omega^r.
\]

Proof of Proposition 2’.

Since the level width of \(L\) is unbounded, we have \(\alpha \geq \omega^r\) for all \(r \geq 1\) and hence,

\[
\alpha \geq \sup\{ \omega^r \mid r \geq 1 \} = \omega^\omega.
\]
Lemma

Let $n, r \geq 1$ and $(L, \prec)$ be a tree automatic copy of an ordinal $\alpha$ such that $L$ can be accepted by a tree automaton with $n$ states and the level width of $L$ is at least $r \cdot 2^n$. Then, $\alpha \geq \omega^r$.

Proof of the lemma.

The proof idea is as follows:

- we construct infinite sets $A_1, \ldots, A_r \subseteq T_{\Sigma}$ in several rounds (the first approach yields unhandy sets and we shrink them twice)
- we define well-orderings $\prec_i$ on $A_i$ (which are of order type $\geq \omega$)
- the order type of $A_1 \times \cdots \times A_r$ ordered lexicographically is $\geq \omega^r$
- we give an embedding $f$ of $A_1 \times \cdots \times A_r \rightarrow L$ (ordered lexicogr.) into $L$ (ordered by $\prec$)

Let

- $A = (Q, \iota, \delta, F)$ be a tree automaton with $n$ states accepting $L$,
- $A_\prec = (Q_\prec, \iota_\prec, \delta_\prec, F_\prec)$ be a tree automaton accepting $\prec$, and
- $t \in L$ be a tree with level width at least $r \cdot 2^n$. 
\( t \in L \)

\[ n \geq r \cdot 2^n \text{ nodes} \]

\[ \geq r \text{ “roots”} \]

\[ u_1, u_i, u_r \]

Observations
\( t \in L \)

Let \( A \) run on \( f(\ldots)! \)

\[ n \geq r \text{ "roots"} \]

\[ \geq r \cdot 2^n \text{ nodes} \]

Observations

- the set \( A_i = \{ t_i \in T_\Sigma \mid \delta(t_i) = q_i \} \) is infinite
Let \( A \) run on \( f(\ldots) \)!

\[
f(t_1, \ldots, t_r)
\]

\[
\begin{align*}
q_f \in F \\
\end{align*}
\]

\[
\begin{align*}
t_1 \in A_1 \\
t_i \in A_i \\
t_r \in A_r
\end{align*}
\]

\[
\begin{align*}
A_i = \{ t_i \in T_\Sigma | \delta(t_i) = q_i \} \\
f(t_1, \ldots, t_r) \in L \text{ for all } t_1 \in A_1, \ldots, t_r \in A_r
\end{align*}
\]

**Observations**

- the set \( A_i = \{ t_i \in T_\Sigma | \delta(t_i) = q_i \} \) is infinite
- \( f(t_1, \ldots, t_r) \in L \) for all \( t_1 \in A_1, \ldots, t_r \in A_r \)
Lemma 1

The order type of

\[ f(A_1 \times \cdots \times A_r) \subseteq L \]

(ordered by \(<\)) is at least \(\omega^r\).

Observation

The value of

\[ \delta_<(f(s_1, \ldots, s_r) \otimes f(t_1, \ldots, t_r)) \]

is completely determined by the values

\[ \delta_<(s_1 \otimes t_1), \ldots, \delta_<(s_r \otimes t_r) \]

Objective

Reduce the number of values \(\delta_<(s_i, t_i)\) can take by shrinking the sets \(A_i\) in such a way that Lemma 1 remains valid.

Notice: This is no “cheating” since shrinking does not increase the order type of \(f(A_1 \times \cdots \times A_r)\).
Step 1

By the pigeonhole principle we can choose a state \( q_i^- \in Q_\prec \) such that there are infinitely many \( t_i \in A_i \) satisfying

\[
\delta_\prec(t_i \otimes t_i) = q_i^-.
\]

Throw all other \( t_i \) out of \( A_i \).

Observation

It does not depend on the choice of \( \tau_1, \ldots, \tau_{i-1}, \tau_{i+1}, \ldots, \tau_r \) whether

\[
f(\tau_1, \ldots, \tau_{i-1}, s_i, \tau_{i+1}, \ldots, \tau_r) < f(\tau_1, \ldots, \tau_{i-1}, t_i, \tau_{i+1}, \ldots, \tau_r) \quad (\star)
\]

holds true, but only on \( \delta_\prec(s_i \otimes t_i) \).

We define a well-ordering \( \prec_i \) on \( A_i \) as follows: we put

\[
s_i \prec_i t_i
\]

if, and only if, condition \((\star)\) above is met.
Result and problem

The state $\delta_<(s_i \otimes t_i)$ completely determines which possibility of

\[ s_i <_i t_i \quad s_i = t_i \quad s_i >_i t_i \]

holds true. There is exactly one state indicating $s_i = t_i$, but there are possibly several indicators for $s_i <_i t_i$ and $t_i <_i s_i$.

Step 2

By Ramsey’s theorem (for infinite undirected colored graphs) we can choose two states $q_i^<, q_i^> \in Q_<$ such that there is an infinite subset $B_i \subseteq A_i$ such that for all $s_i, t_i \in B_i$ we have:

- If $s_i < t_i$, then $\delta_<(s_i \otimes t_i) = q_i^<$.  
- If $s_i > t_i$, then $\delta_<(s_i \otimes t_i) = q_i^>$.  

Replace $A_i$ by $B_i$.  

Result

There are three states $q_i^<, q_i^=, q_i^> \in Q_<$ such that for all $s_i, t_i \in A_i$ we have:

\[
\begin{align*}
  s_i <_i t_i & \iff \delta_<(s_i \otimes t_i) = q_i^< \\
  s_i = t_i & \iff \delta_<(s_i \otimes t_i) = q_i^= \\
  s_i >_i t_i & \iff \delta_<(s_i \otimes t_i) = q_i^>
\end{align*}
\]

Step 3

Let $0_i$ and $1_i$ be the least and the second least (w.r.t. $<_i$) element of $A_i$. For $1 \leq i \leq r$ put

\[
e_i = f(0_1, \ldots, 0_{i-1}, 1_i, 0_{i+1}, \ldots, 0_r)
\]

W.l.o.g. assume $e_1 > \cdots > e_r$. 
Lemma 2

Let $<_{\text{lex}}$ be the lexicographical ordering on $A_1 \times \cdots \times A_r$ induced by $<_1, \ldots, <_r$. Then, for all $\bar{s}, \bar{t} \in A_1 \times \cdots \times A_r$ we have

$$\bar{s} <_{\text{lex}} \bar{t} \implies f(\bar{s}) < f(\bar{t}),$$

i.e., $f$ is an (order preserving) embedding of $A_1 \times \cdots \times A_r$ (ordered lexicogr.) into $L$ (ordered by $<$).

Consequence

Let $\beta$ denote the order type of $f(A_1 \times \cdots \times A_r)$ (ordered by $<$) and $\alpha_i$ the order type of $A_i$ (ordered by $<_i$). Then,

$$\alpha \geq \beta = \alpha_r \cdots \alpha_1 \geq \omega \cdots \omega = \omega^r.$$
Proof of Lemma 2 for $r = 2$.

We want to show

$$(s_1, s_2) \lessdot \text{lex} \ (t_1, t_2) \implies f(s_1, s_2) < f(t_1, t_2).$$

There are several cases:

1. $s_1 = t_1$ and $s_2 < t_2$: definition of $<_2$
2. $s_1 < t_1$ and $s_2 = t_2$: definition of $<_1$
3. $s_1 < t_1$ and $s_2 < t_2$: definitions of $<_1$ and $<_2$ imply

$$f(s_1, s_2) < f(t_1, s_2) < f(t_1, t_2)$$

4. $s_1 < t_1$ and $s_2 > t_2$: $0_1 < 1_1$ and $1_2 > 2_0$ imply

$$\delta_<(s_1 \otimes t_1) = \delta_<(0_1 \otimes 1_1) \quad \text{and} \quad \delta_<(s_2 \otimes t_2) = \delta_<(1_2 \otimes 0_2)$$

$$\delta_<(f(s_1, s_2) \otimes f(t_1, t_2)) = \delta_<(f(0_1, 1_2) \otimes f(1_1, 0_2)) = \delta_<(e_2 \otimes e_1)$$

Finally, $e_2 < e_1$ implies $f(s_1, s_2) < f(t_1, t_2)$. \triangle
Proposition 3

Given a tree automaton $A$, it is decidable whether the level width of tree language accepted by $A$ is (un)bounded. If it is bounded, then $(n + 1) \cdot 2^n$ is an upper bound, where $n$ is the number of states of $A$.

Proof.

Let $A = (Q, \iota, \delta, F)$ and $L \subseteq T_\Sigma$ be the language accepted by $A$. W.l.o.g. assume that $A$ is reduced, i.e., for each $q \in Q$ with $q \neq \iota$ there is a $t \in T_\Sigma$ such that $\delta(t) = q$.

We define a directed graph $(Q, E)$ by

$$(q, p) \in E \iff \exists r \in Q, a \in \Sigma: p = \delta(q, r, a) \text{ or } p = \delta(r, q, a).$$

For each transition $p = \delta(q, r, a)$ we get two edges:

```
q
  ^
  |
  v
r
```
Proof (cont’d).

In this graph \((Q, E)\) we define some special edges \(E_S \subseteq E\) by

\[(q, p) \in E_S \iff \exists r \in Q_\infty, a \in \Sigma: p = \delta(q, r, a) \text{ or } p = \delta(r, q, a),\]

where

\[Q_\infty = \{ q \in Q \mid \exists \infty t \in T_\Sigma: \delta(t) = q \} .\]

For each transition \(p = \delta(q, r, a)\) there are four possibilities:

---

**Lemma**

The following are equivalent:

1. The level width of \(L\) is unbounded.
2. There is a cycle in the graph \((Q, E)\) that contains at least one special edge and from which a state from \(F\) is reachable.

Notice: This property is decidable since \(Q_\infty\) is computable.
\( \exists p \in Q, a \in \Sigma : q_2 = \delta(p, q_1, a) \)

\( \ell(t_2) \geq \ell(t_1) \)
\( \Rightarrow \mathbb{1} \): The graph property implies unboundedness

\[ \ell(t_3) > \ell(t_2) \geq \ell(t_1) \]
Unboundedness implies the graph property $\geq (n + 1) \cdot 2^n$. 

Diagram showing $p_0, p_1, p_2, p_n, q_0, q_1, q_2, q_n$ connected in a certain pattern.
Unboundedness implies the graph property

\[ \geq (n + 1) \cdot 2^n \]
### Summary and extensions 1

<table>
<thead>
<tr>
<th>Theorem (H 2011)</th>
<th>Theorem (H 2011)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Given a tree automatic presentation of an ordinal $\alpha$, it is decidable whether $\alpha$ is word automatic.</td>
<td></td>
</tr>
<tr>
<td>Given a tree automatic presentation of a scattered linear ordering $\mathcal{L}$, it is decidable whether $\mathcal{L}$ is word automatic.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Theorem (Delhommé 2001)</th>
<th>Theorem (KRS 2003)</th>
</tr>
</thead>
<tbody>
<tr>
<td>An ordinal $\alpha$ is word automatic if, and only if, $\alpha &lt; \omega^\omega$.</td>
<td></td>
</tr>
<tr>
<td>The Hausdorff rank of every word automatic scattered linear ordering is finite.</td>
<td></td>
</tr>
</tbody>
</table>
Extensions 2 and outlook

Further results

- Proposition 1 is not valid only for ordinals but for all classes of structures:
  1. Let \((L, \ldots)\) be a tree automatic copy of a structure \(S\) such that \(L\) has bounded level width. Then, \(S\) is word automatic.
  2. From a tree automatic presentation of \(S\) one can compute a word automatic presentation of \(S\).

- For every tree automatic ordinal \(\alpha\) the set of all (codes of) \(\omega\)-limit points of \(\alpha\) is (effectively) regular.

Open problem

If we knew, that for each tree automatic ordinal \(\alpha\) and any \(n \geq 1\) the set of (codes of) \(\omega^n\)-limit points of \(\alpha\) is (effectively) regular, we had solved the isomorphism problem for tree automatic ordinals. However, “if”…