Subsequence Languages: An Exposition

William Gasarch-U of MD

Definitions

Defintion: Let Σ be a finite alphabet.

1. Let $w \in \Sigma^*$. $SUBSEQ(w)$ is the set of all strings you get by replacing some of the symbols in w with the empty string.
2. Let $L \subseteq \Sigma^*$. $SUBSEQ(L) = \bigcup_{w \in L} SUBSEQ(w)$.

Example: *abaa* has the following subsequences: *a*, *b*, *aa*, *ab*, *ba*, *aaa*, *aab*, *aba*, *baa*, *abaa*.

Easy Theoremss

1. If L is regular than $SUBSEQ(L)$ is regular.
2. If L is context free than $SUBSEQ(L)$ is context free.
3. If L is c.e. than $SUBSEQ(L)$ is c.e.

What about: *If L is decidable then* $SUBSEQ(L)$ *is decidable.*

VOTE:

1. If L is decidable then $SUBSEQ(L)$ is decidable.
2. $(\exists L)[$ L decidable but $SUBSEQ(L)$ is NOT decidable$]$.
3. Anyone but Trump. Or Cruz. Or Rubio. Or ... Or Gillmore.

The Surprising Truth

If L is ANY subset of Σ^* then SUBSEQ(L) is regular.

Higman first proved this theorem in the 1950's using different terminology. He used:

Definition: A set together with an ordering $(X, \preceq)$ is a *well quasi ordering* (wqo) if for any sequence $x_1, x_2, \ldots$ there exists i, j such that $i < j$ and $x_i \preceq x_j$.

Note: If $(X, \preceq)$ is a wqo then its both well founded and has no infinite antichains.

Closed Downwards

Lemma: Let $(X, \preceq)$ be a countable wqo and let $Y \subseteq X$. Assume that Y is closed downward under $\preceq$. Then there exists a finite set of elements $\{z_1, \ldots, z_k\} \subseteq X - Y$ such that

$$y \in Y \text{ iff } (\forall i)[z_i \not\preceq y].$$

(The set $\{z_1, \ldots, z_k\}$ is called an *obstruction set (OBS)*.)

Note: Really hard example: Graph Minor is a wqo, Planar graphs are closed downward, and $OBS = \{K_{3,3}, K_5\}$.

Subsequence Order

Definition: The *subsequence order on* Σ^*, which we denote $\preceq_{\text{subseq}}$, is defined as $x \preceq_{\text{subseq}} y$ if x is a subsequence of y.

Theorem: $(\Sigma^*, \preceq_{\text{subseq}})$ is a wqo.
Proof: Assume not. Obtain MIN BAD SEQUENCE:
y_1 is shortest 1st element of a bad sequence.
y_2 is shortest 2nd element of a bad sequence that begins y_1.
etc.

$$y_1, y_2, \ldots$$

Let $y_i = y_i' \sigma_i$ where $\sigma_i \in \Sigma$.
Let $Y = \{y_1', y_2', \ldots\}$.

Y is a wqo

Claim: Y is a wqo.
Proof of Claim: Assume not. Bad Seq: $y'_{k_1}, y'_{k_2}, \ldots$
$(k_1 \le \{k_2, k_3, \ldots\})$.
Consider:

$$\mathrm{SEQ} : y_1, y_2, \ldots, y_{k_1 - 1}, y'_{k_1}, y'_{k_2}, \ldots$$

We show there is no $i < j$ with $y_i \preceq y_j$, so SEQ is BAD:
$i < j \le k_1 - 1 \wedge y_i \preceq y_j \implies y_1, y_2, \ldots$ not BAD
$i < j \wedge y'_{k_i} \preceq y'_{k_j} \implies y'_{k_1}, y'_{k_2}, \ldots$ not BAD
$i \le k_j \wedge y_i \preceq y'_{k_j} \implies y_i \preceq y'_{k_j} \preceq y'_{k_j} \sigma k_j = y_{k_j}$. So $y_1, y_2, \ldots$ not BAD
SO SEQ is BAD.
SEQ begins $y_1, y_2, \ldots, y_{k_1 - 1}$.
SEQ k_1th element is y'_{k_1} which is SHORTER than y_{k_1}.
Contradicts $y_1, y_2, \ldots,$ being a MINIMAL bad sequence.
End of Proof of Claim

Y is a wqo, Σ is a wqo...

Y is a wqo. Σ is a wqo. So $Y \times \Sigma$ is a wqo.
Look at the sequence

$$y_1, y_2, \ldots,$$

There exists σ such that infinity many end with σ.

$$y'_{k_1}\sigma, y'_{k_2}\sigma, \ldots$$

Since $Y = \{y'_1, y'_2, \ldots\}$ is wqo $(\exists i < j$ with $y'_{k_i} \le y'_{k_j}$.
So $y'_{k_i}\sigma \le y_{k_j}\sigma$ so $y_{k_i} \le y_{k_j}\sigma$.
SO $y_1, y_2, \ldots$ AINT BAD! Contradiction.

Subseq Theorem

Theorem: Let Σ be a finite alphabet. If $L \subseteq \Sigma^*$ then $SUBSEQ(L)$ is regular.
Proof: Σ is a wqo. Hence $(\Sigma^*, \preceq_{\rm subseq})$ is a wqo.
If $L \subseteq \Sigma^*$ then $SUBSEQ(L)$ is closed under $\preceq_{\rm subseq}$. So $SUBSEQ(L)$ has a finite obstruction set. Hence regular.

Nonconstructive?

Given a DFA, CFG, P-machine, NP-machine, TM (decidable), TM (c.e.) for a language L, can one actually obtain a DFA for $SUBSEQ(L)$? For that matter, can you obtain a CFG, etc for $SUBSEQ(L)$? Gasarch, Fenner, Postow showed all of the NCON below. Leeuwen the CFG/REG CON result. The rest are easy.

	$SBSEQ(REG)$	$SBSEQ(CFG)$	$SBSEQ(DEC)$	$SBSEQ(C.E.)$
REG	CON	CON	CON	CON
CFG	CON	CON	CON	CON
DEC	$NCON$	$NCON$	$NCON$	CON
$C.E.$	$NCON$	$NCON$	$NCON$	CON