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 $L \subseteq \Sigma^*$. $SUBSEQ(L) = \bigcup_{w \in L} SUBSEQ(w)$.

End of Definition

Example: *abaab* has the following subsequences: *a*, *b*, *aa*, *ab*, *ba*, *bb*, *aaa*, *aab*, *aba*, *abb*, *baa*, *bab*, *aaab*, *abaa*, *abab*, *abaab*.

Easy Theorems

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: TRUE of FALSE.

The Surprising Truth

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

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

Well Quasi Orderings

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

Equiv to WQO

Lemma: The following are equivalent:

- $(X, \preceq)$ is a wqo,
- For any sequence $x_1, x_2, \ldots \in X$ there exists an *infinite* ascending subsequence.

End of Lemma

Try yourself in groups.

Proof

Let $x_1, x_2, \ldots,$ be an infinite sequence. Define the following coloring:
$COL(i,j) =$

- UP if $x_i \preceq x_j$.
- DOWN if $x_j \prec x_j$.
- INC if x_i and x_j are incomparable.

By Ramsey there is homog set. If colored DOWN or INC then violates wqo. So must be UP.

Cross Product

Definition: If $(X, \preceq_1)$ and $(Y, \preceq_2)$ are wqo then we define $\preceq$ on $X \times Y$ as $(x,y) \preceq (x',y')$ if $x \preceq_1 y$ and $x' \preceq_2 y'$.

Closed Under Cross Product

Lemma: If $(X, \preceq_1)$ and $(Y, \preceq_2)$ are wqo then $(X \times Y, \preceq)$ is a wqo ($\preceq$ defined as in the above definition).
End of Lemma
Try yourself in groups.

Proof

Let $(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots$ be an infinite sequence of elements from $A \times B$.
Define the following coloring:
$COL(i, j) =$

- UP-UP if $x_i \preceq x_j$ and $y_i \preceq y_j$.
- UP-DOWN if $x_i \preceq x_j$ and $y_j \preceq y_i$.
- UP-INC if $x_i \preceq x_j$ and y_j, y_i are incomparable.
- DOWN-UP, DOWN-DOWN, DOWN-INC, INC-UP, INC-DOWN, INC-INC are defined similarly.

Use Ramsey's Theorem. UP-UP is the only possible color of a homog set, else either X or Y is not a wqo.

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.*)
End of Lemma
Try yourself in groups.

Proof

Let OBS be the set of elements z such that

1. $z \notin Y$.
2. Every $y \preceq z$ is in Y.

Claim 1: *OBS* is finite
Try yourself in groups.

OBS finite

Claim 1: ***OBS*** **is finite**
Try yourself in groups.
Proof of Claim 1: Every $z, z' \in OBS$ are incomparable: Assume NOT. Then $(\exists z, z')[z \preceq z']$. $z \in Y$ by part 2 of the definition of OBS. But if $z \in Y$ then $z \notin OBS$. Contradiction.
Assume that OBS is infinite. Then the elements of OBS (in any order) form an infinite anti-chain. Contradicts wqo.
End of Proof of Claim 1

Finish it Up

Let $OBS = \{z_1, z_2, \ldots\}$.
Claim 2: For all y:

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

Try yourself in groups.

Finish it Up

Let $OBS = \{z_1, z_2, \ldots\}$.
Claim 2: For all y:

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

Try yourself in groups.
Proof of Claim 2: Contrapositive:

$$y \notin Y \text{ iff } (\exists i)[z_i \preceq y].$$

Assume $y \notin Y$. If $y \in OBS$ DONE. If $y \notin OBS$ then $(\exists z_1)[z_1 \notin Y \wedge z_1 \prec y]$. If $z \in OBS$ DONE. If not then repeat. If process STOPS then DONE. If not then
$\cdots z_{17} \prec z_{16} \prec \cdots \prec z_1 \prec y]$, violates wqo.
End of Proof of Claim 2 and of Proof

Subsequence Order

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

Main Theorem

Theorem: $(\Sigma^*, \preceq_{\rm subseq})$ is a wqo.

Proof

Assume not. Obtain MIN BAD SEQUENCE

$$y_1, y_2, \ldots$$

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

Claim: Y is a wqo.
Proof of Claim: Assume not.
Bad Sequence: $y'_{k_1}, y'_{k_2}, \ldots$ (can take $k_1 \le \{k_2, k_3, \ldots\}$).
Consider: $y_1, y_2, \ldots, y_{k_1-1}, y'_{k_1}, y'_{k_2}, \ldots$.
This is BAD:
if $i < j \le k_1 - 1$ and $y_i \preceq y_j$ then $y_1, y_2, \ldots$ is not BAD.
if $i < j$ and $y'_{k_i} \preceq y'_{k_j}$ then $y'_{k_1}, y'_{k_2}, \ldots$ is not BAD.
if $i \le k_j$ an $y_i \preceq y'_{k_j}$ then $y_i \preceq y'_{k_j} \preceq y'_{k_j} \sigma k_j = y_{k_j}$. KEY: $i < k_j$. So $y_1, y_2, \ldots$ is not BAD.
SO $y_1, y_2, \ldots$ is BAD. It begins $y_1, y_2, \ldots, y_{k_1-1}$. Its 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

$$(y_1', \sigma_1), (y_2', \sigma_2), \ldots.$$

where $y_i = y_i'\sigma_i$.
There exists $i < j$ with $(y_i', \sigma_i) \prec_{\rm cross} (y_j', \sigma_j)$. Hence $y_i'\sigma_i \prec_{\rm subseq} y_j'\sigma_j$. Hence $y_i \prec y_j$. Contradicts $y_1, y_2, \ldots$ being BAD.

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