MA/CSSE 474

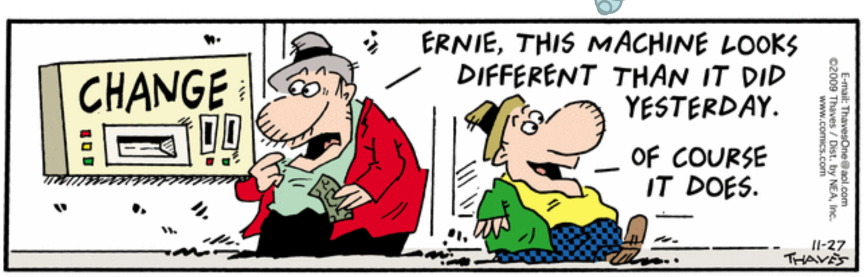
Theory of Computation

Languages, prefixes, sets, cardinality, functions

Your Questions?

- Syllabus
- Tuesday's discussion
- Reading Assignments
- HW1 or HW2
- Anything else

Must not be a FSM



Languages and Strings

Mostly very quick.

Some should be review of previous courses, and some others you should have gotten for Reading Quiz 2.

Ask questions if there are things I list here that you are not sure about.

Properties of Strings

- A ***string*** is a finite sequence (possibly empty) of symbols from some finite alphabet Σ.
- ε is the empty string (some books/papers use λ instead)
- Σ^* is the set of all possible strings over an alphabet Σ
- ***Counting:*** $|s|$ is the number of symbols in *s*. $|\varepsilon| = 0$ $|\texttt{1001101}| = 7$
- $\#_c(s)$ is the number of times that *c* occurs in *s*. $\#_a(\texttt{abbaaa}) = 4$.

More Functions on Strings

Concatenation: *st* is the ***concatenation*** of *s* and *t*.

If x = `good` and y = `bye`, then xy = `goodbye`.

Note that $|xy| = |x| + |y|$.

ε is the **identity** for concatenation of strings. So:

$\forall x\ (x\,\varepsilon = \varepsilon\, x = x)$.

Concatenation is **associative**. So:

$\forall s, t, w\ ((st)w = s(tw))$.

More Functions on Strings

Replication: For each string w and each natural number i, the string w^i is:

$w^0 = \varepsilon,\ \ w^{i+1} = w^i\, w$

Examples:

$\mathtt{a}^3$ = `aaa`

$(\mathtt{bye})^2$ = `byebye`

$\mathtt{a}^0\mathtt{b}^3$ = `bbb`

Reverse: For each string w, w^R is defined as:

if $|w| = 0$ then $w^R = w = \varepsilon$

if $|w| \geq 1$ then:

$\exists a \in \Sigma\ (\exists u \in \Sigma^*\ (w = ua))$.

So define $w^R = a\, u^R$.

Concatenation and Reverse of Strings

Theorem: If w and x are strings, then $(w\,x)^R = x^R\,w^R$.

Example:

$(\texttt{nametag})^R = (\texttt{tag})^R\,(\texttt{name})^R = \texttt{gateman}$

Proof on next slide

Concatenation and Reverse of Strings

Proof: By induction on $|x|$:

$|x| = 0$: Then $x = \varepsilon$, and $(wx)^R = (w\,\varepsilon)^R = (w)^R = \varepsilon\,w^R = \varepsilon^R\,w^R = x^R\,w^R$.

$\forall n \geq 0\ (((|u| = n) \rightarrow ((w\,u)^R = u^R\,w^R)) \rightarrow$
$((|x| = n + 1) \rightarrow ((w\,x)^R = x^R\,w^R)))$:

Consider any string x, where $|x| = n + 1$. Then $x = u\,a$ for some symbol a and $|u| = n$. So:

$(w\,x)^R$	$= (w\,(u\,a))^R$	rewrite x as ua
	$= ((w\,u)\,a)^R$	associativity of concatenation
	$= a\,(w\,u)^R$	definition of reversal
	$= a\,(u^R\,w^R)$	induction hypothesis
	$= (a\,u^R)\,w^R$	associativity of concatenation
	$= (ua)^R\,w^R$	definition of reversal
	$= x^R\,w^R$	rewrite ua as x

Relations on Strings:

Substring, ***proper substring***
Every string is a substring of itself.
ε is a substring of every string.

prefix, ***proper prefix***
Every string is a prefix of itself.
ε is a prefix of every string.

s is a ***suffix***, **proper suffix**, self, ε

Defining a Language

A ***language*** is a (finite or infinite) set of strings over a finite alphabet Σ. Examples for Σ = {a, b}

1. $L = \{x \in \{a, b\}^* :$ all a's precede all b's}
   ε, a, aa, aabbb, and bb are in *L*. aba, ba, and abc are not in *L*.
2. $L = \{x : \exists u \in \{a, b\}^* : x = u\text{a}\}$
   Simple English description:
3. $L = \{x\#y: x, y \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}^*$ and, when x and y are viewed as the decimal representations of natural numbers, $square(x) = y\}$.
   Examples (in L or not?):
   `3#9, 12#144, 3#8, 12, 12#12#12, #`
4. $L = \{\text{a}^n : n \geq 0\}$ **uses replication, simpler description of L?**
5. $A^nB^n = \{a^kb^k : k \geq 0\}$
6. L = Ø = { }
7. L = {ε}

You saw in Reading Quiz 2 that the last two examples are different languages

Natural Languages are Tricky

L = {w: w is a sentence in English}.

Examples:

Kerry hit the ball.

Colorless green ideas sleep furiously.

The window needs fixed.

Ball the Stacy hit blue.

A *Halting Problem* Language

L = {w: w is a Java program that, when given any finite input string, is guaranteed to halt}.

- Is this language well specified?
- Can we decide which strings L contains?

Languages and Prefixes

What are the following languages?

L = {$w \in$ {a, b}*: no prefix of w contains b}

L = {$w \in$ {a, b}*: no prefix of w starts with a}

L = {$w \in$ {a, b}*: every prefix of w starts with a}

Concatenation of Languages

If L_1 and L_2 are languages over Σ:

$L_1L_2 = \{w \in \Sigma^* : \exists s \in L_1 (\exists t \in L_2 (w = st))\}$

Alternate definition:

$L_1L_2 = \{ st : s \in L1 \wedge t \in L2 \}$

Simpler than the first definition, but the first one conveys the idea more precisely.

Example:

L_1 = {a, aa}
L_2 = {a, c, ε}

$L_1 L_2$ =

Concatenation of Languages

- L_1L_2
- L^R
- L^3 Is this the same as $\{w^3 : w \in L\}$
- L^0
- L^k
- L^*
- L^+

Formally: Kleene Star and + of a Language

$L^* = \{\varepsilon\} \cup$
$\{w \in \Sigma^* : \exists k \geq 1$
$(\exists w_1, w_2, \ldots w_k \in L\ (w = w_1 w_2 \ldots w_k))\}$

Alternate: $L^* = L^0 \cup L^1 \cup L^2 \cup \ldots = \bigcup_{k=0}^{\infty} L^k$

$L^+ = L\, L^*$

$L^+ = L^* - \{\varepsilon\}$ iff $\varepsilon \notin L$

L^+ is the closure of L under concatenation.

Concatenation and Reverse of Languages

Theorem: $(L_1 L_2)^R = L_2^R L_1^R$.

Proof:

$\forall x\ (\forall y\ ((xy)^R = y^R x^R))$ Theorem 2.1 we proved last time

$(L_1 L_2)^R = \{(xy)^R : x \in L_1 \text{ and } y \in L_2\}$ Definition of concatenation of languages

$= \{y^R x^R : x \in L_1 \text{ and } y \in L_2\}$ Thm 2.1

$= L_2^R L_1^R$ Definition of concatenation of languages

Sets and Relations

Sets of Sets

- The ***power set*** of *S* is the set of all subsets of *S*.

  Let *S* = {1, 2, 3}. Then:

  $\mathcal{P}(S) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$.

- $\Pi \subseteq P(S)$ is a ***partition*** of a set *S* iff:
  - Every element of Π is nonempty,
  - Every pair of elements of Π is disjoint , and
  - the union of all the elements of Π equals *S*.

  Some partitions of = {1, 2, 3}:

  {{1}, {2, 3}} or {{1, 3}, {2}} or {{1, 2, 3}}.
  How many *different* partitions of S?

Closure

- A set S is **closed** under binary operation *op* iff
  $\forall x,y \in S$ (x *op* y $\in$ S) ,
  closed under unary function f iff
  $\forall x \in S$ (f(x) $\in$ S)

If S is not closed under unary function f, a **closure** of S is a set S' such that
a) S is a subset of S'
b) S' is closed under f
c) No proper subset of S' contains S and is closed under f

- *Examples*
- ℕ+ (the set of all positive integers) is closed under addition and multiplication but not negation, subtraction, or division.
- What is the closure of ℕ+ under subtraction? Under division?
- The set of all finite sets is closed under union and intersection. Closed under infinite union?

Equivalence Relations

A relation on a set A is any set of ordered pairs of elements of A.

A relation $R \subseteq A \times A$ is an ***equivalence relation*** iff it is:

•reflexive,

•symmetric, and

•transitive.

Examples of equivalence relations:

•Equality

•Lives-at-Same-Address-As

•Same-Length-As

•Contains the same number of a's as

Show that $\equiv_3$ is an equivalence relation

Cardinality of a set.

The cardinality of every set we will consider is one of the following :

- a specific natural number (if *S* is finite),
- “countably infinite” (if *S* has the same number of elements as there are integers), or
- “uncountably infinite” (if *S* has more elements than there are integers).

The rest of today's slides

- We probably won't get to them today.
- But they are here just in case …

Functions on Languages

Functions whose domains and ranges are languages

maxstring(L) = $\{w \in L: \forall z \in \Sigma^* (z \neq \varepsilon \rightarrow wz \notin L)\}$.

Exercise for later:
What language is
maxstring($\{b^n a: n \geq 0\}$) ?

Examples:

- *maxstring*(A^nB^n)
- *maxstring*({a}*)

Let INF be the set of infinite languages.
Let FIN be the set of finite languages.

Are the language classes FIN and INF closed under *maxstring*?

Functions on Languages

chop(L) =
$\{w : \exists x \in L\ (x = x_1 c x_2,\ x_1 \in \Sigma_L^*,\ x_2 \in \Sigma_L^*,\ c \in \Sigma_L,$
$|x_1| = |x_2|,$ and $w = x_1 x_2)\}$.

What is *chop*(A^nB^n)?

What is *chop*($A^nB^nC^n$)?

Are FIN and INF closed under *chop*?

Functions on Languages

firstchars(L) =
$\{w : \exists y \in L\ (y = cx \wedge c \in \Sigma_L \wedge x \in \Sigma_L^* \wedge w \in \{c\}^*)\}$.
.

What is *firstchars*(A^nB^n)?

What is *firstchars*({a, b}*)?

Are FIN and INF closed under *firstchars*?