CS 3100 – Models of Computation – Fall 2010

August 26, 2010

Notes 3, Handed out: August 31, 2010 during Lecture 3

- Work out solutios for Assignment 1

Concept / mathematical notation practice: Write these mathematically:

- Empty string
- Empty set – write this in two different ways
- Write out the powerset of $\{a, b\}$
- How many elements are there in the powerset of $\{a, b, c, d\}$?
  - Each item is present or absent
  - Let 0000 represent "no elements are there" i.e. $\emptyset$
  - Let 0001 represent $\{d\}$
  - Let 1001 represent $\{a, d\}$
  - You get the idea now. Each subset of $\{a, b, c, d\}$ is represented by each four-bit bit vector.
  - How many such bit vectors are there? 0000, 0001, 0010, etc. all the way to 1111?

- Set containining empty string
- Set containining empty set
- Set containining an empty string and an empty set (now, why would you do that?!

  This is strictly allowed, but a set usually has only items of the same type – either all sets or all strings.)
- String containing an empty set (Groan! Can you do this? No! Strings are not sets.)
- Union of set $\{a\}$ and itself
- The result of inserting a and b into $\{a\}$

Notions centered around languages

- A language is a set of strings
- Empty language : $\emptyset$
- The language containing the empty string : $\{\varepsilon\}$
- One is often interested in the beginning and ending patterns within strings. To specify this, one can split the string into two pieces that can be *concatenated* to form the whole.
- Concatenation of *ab* and *cde* is *abcde*
- Concatenation of ε and *abd* is *abd*
- Now I'm going to attempt to motivate the notion of *concatenation of languages*
- When one talks about "all possible former halves of strings" and "all possible latter halves of strings," one is talking about the *language* of the first halves and the language of the last halves
  - Example : telephone number = area-code followed by main-number
  - `area-code =  ( d d d )`
  - `main-number =  ddd-dddd`
- One can then talk about the language of area codes and language of telephone numbers, i.e., L_{ac} and L_{telno}, and concat. these languages

Language concatenation

- If $L_1 = \{ab, b, varepsilon\}$ and $L_2 = \{b, dd, \varepsilon\}$ then $L_1 L_2$ has all strings where the first string comes from L_1 and the second from L_2
- So what is $\{a, aa\}\{bb, b\}$?
- What is $\{\}\{\varepsilon\}$?
- What is $\{\}\{a, aa\}$?
- What is $\{a, aa\}\{\}$?
- What is $\{\varepsilon\}\{a, aa\}\{\varepsilon\}\{bb\}$?
- For $L = \{ab, b, \varepsilon\}$, we have $L^2 = \{ab, b, \varepsilon\}\{ab, b, \varepsilon\}$
- For the above L, we have $L^3 = \{ab, b, \varepsilon\}\{ab, b, \varepsilon\}\{ab, b, \varepsilon\}$
- For the above L, we have $L^0 = \{\varepsilon\}$ by definition (to allow the concat not to entirely disappear)!

Language union:

- What is $\{bb, cc\} \cup \{a, aa\}$?

Kleene star of a language

- $L^* = L^0 \cup L^1 \cup ...$
- i.e. $L^* = \cup_{k \geq 0} L^k$
- This means $\emptyset^* = \{\varepsilon\}$

Regular expressions

- Shorthands for (regular) languages
- $\emptyset$ is an RE denoting language $\emptyset$
- ε is an RE denoting language $\{\varepsilon\}$
- 0 is an RE denoting language $\{0\}$
- 1 is an RE denoting language $\{1\}$
- For R_1 and R_2 as REs, $R_1 + R_2$ is an RE denoting...
- For R_1 and R_2 as REs, R_1R_2 is an RE denoting...
- For R as an RE, (R) denotes...
- For R as an RE, R^* denotes...
- Do problem 2.1

  Which of $\varepsilon, abba, bababb, baaaa$ are in the language of $(a + b)^*ab(a + b)^*$?

- **Short break**
- Discuss Assignment 2, introducing solving similar problems.