Chapter 8, Strings:

- General introduction; string sorting (last week Friday)
- Tries (Monday)
- Regular expression (Today)

Today:

- What regular expressions are
- How to use regular expressions practically
- Why regular expressions are important theoretically

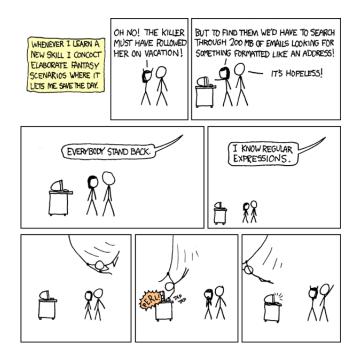
▲ロト ▲圖ト ▲画ト ▲画ト 三回 - のへで

Projects:

- Last regular project score update on Tues, Apr 25
- "Two minute warning" run of scripts on Fri, Apr 28 (no Schoology update—see report file /cslab/class/cs345/(your userid)/(your userid).report)
- All projects due on the last day of *classes*, midnight between Fri, Apr 28 and Sat, Apr 29—not last day of finals.

Final exam

- Our final exam block is Wed, May 3, 10:30am–12:30pm
- During our final exam block, we will meet in the CSCI lab
- Test 3 ("written" /conceptual part) will be like Test 1, but covering the second part (MST and SSSP) of graphs (ch 6), and then BSTs (ch 5) through strings (ch 8)
- Test 4 (programming part) will work the same way as Test 2, covering dynamic programming, hashing, and strings.



- An **alphabet** is a set of symbols, Σ .
- A string over an alphabet is a sequence of symbols from that alphabet. Σ* is the set of all strings over alphabet Σ.

▲ロト ▲圖ト ▲画ト ▲画ト 三回 - のへで

- A language over an alphabet is a set of strings, that is, a subset of Σ^* .
- Regular expressions constitute a system for specifying languages; a regular expression denotes a language.

base cases	\emptyset $arepsilon$	the empty set of strings the set containing the empty string, $\{""\}$ the set containing only the string with only <i>a</i> , for some $a \in \Sigma$, $\{"a"\}$
recursive cases	r s	the set of strings made from concatening strings from r and s , $\{x + y \mid x \in r \land y \in s\}$, for some regular expressions r and s the set of strings from r or s , $r \cup s$ for some regular expressions r and s the set of strings made from concatenating 0 or more strings from r for some regular expression r

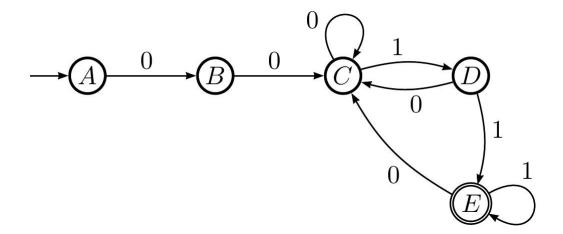
▲□▶ ▲□▶ ▲三▶ ▲三▶ 三三 のへで

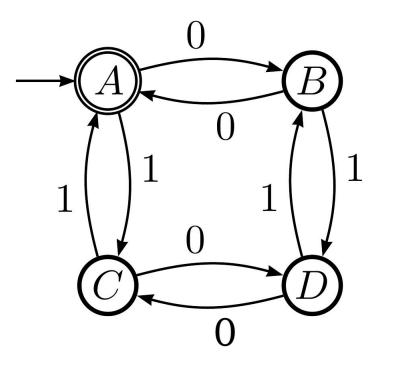
Abbreviation	Meaning	Equivalence
[abc]	One occurrence of any of these symbols	(a b c)
[a-c]	One occurrence of any symbol in this range	(a b c)
r?	Optionally an occurrence of a string defined by r	(r arepsilon)
r ⁵	5 occurrences of a string defined by r	rrrrr
r ^{3,5}	Between 3 and 5 occurrences of a string defined by r	(rrr rrrr rrrr)
r+	One or more occurrences of a string defined by <i>r</i>	rr*

▲□▶ ▲□▶ ▲三▶ ▲三▶ 三三 のへで

- DNA sequences: (A|C|G|T)*.
- Identifiers: ('?[A-Za-z][A-Za-z0-9_])|_.
- Phone numbers: $[2-9][0-9]^2 [2-9][0-9]^2 [0-9]^4$.
- ► Dates: ((1[0-2])|[1-9])/(30|31|([12][0-9])|[1-9])/[1-9][0-9]^{0,3}.
- US Postal Addresses: [0-9]+ [NSEW]^{0,2} [A-Z] [a-z]* (St|Ave|Rd|Ln|Dr| Blvd), ([A-Z] [a-z]*)*, [A-Z]² [0-9]⁵.

▲ロト ▲圖ト ▲画ト ▲画ト 三回 - のへで





▲□▶ ▲□▶ ▲三▶ ▲三▶ 三三 のへで