

Chapter 8, Strings:

- ▶ General introduction; string sorting (last week Friday)
- ▶ Tries (Monday)
- ▶ Regular expressions (**Today**)

Today:

- ▶ A few follow-up comments about tries
- ▶ Context: Alphabets, strings, and languages
- ▶ What regular expressions are
- ▶ Practical use of regular expressions
- ▶ Fun examples

End-of-semester important dates

- ▶ Tues, Dec 2: Test 4 practice problems made available. ✓
- ▶ Thurs, Dec 4: Last “normal” lab ✓
- ▶ Mon, Dec 8: Last project assigned ✓
- ▶ Tues, Dec 9: Last “normal” running of project grading script ✓
- ▶ Wed, Dec 10: Test 3 & 4 Review sheet distributed. ✓
- ▶ Thurs, Dec 11: Review lab (pick practice problems for Test 4)
- ▶ Fri, Dec 12, AM: “Two-minute warning” running of project grading script (Canvas gradebook will not be updated—see project report in your turn-in file)
Note that Fri, Dec 12 is the Last Day of Classes.
- ▶ Fri, Dec 12, 11:59 PM: Official project deadline
- ▶ Sat, Dec 13, when I wake up: Permissions to turn-in folders turned off
- ▶ Mon, Dec 15: Project grading script run for final/semester grades
- ▶ Thurs, Dec 18, 10:30am-12:30pm: Tests 3 and 4 (in lab)

WHENEVER I LEARN A
NEW SKILL I CONCOCT
ELABORATE FANTASY
SCENARIOS WHERE IT
LETS ME SAVE THE DAY.

OH NO! THE KILLER
MUST HAVE FOLLOWED
HER ON VACATION!



BUT TO FIND THEM WE'D HAVE TO SEARCH
THROUGH 200 MB OF EMAILS LOOKING FOR
SOMETHING FORMATTED LIKE AN ADDRESS!

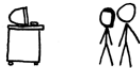


IT'S HOPELESS!

EVERYBODY STAND BACK.



I KNOW REGULAR
EXPRESSIONS.



- ▶ 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 .
- ▶ 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 $\left\{ \begin{array}{ll} \emptyset & \text{the empty set of strings} \\ \varepsilon & \text{the set containing the empty string, } \{""\} \\ a & \text{the set containing only the string with only } a, \\ & \text{for some } a \in \Sigma, \{ "a" \} \end{array} \right.$

recursive cases $\left\{ \begin{array}{ll} rs & \text{the set of strings made from concatenating strings from } r \text{ and } s, \\ & \{x + y \mid x \in r \wedge y \in s\}, \text{ for some regular expressions } r \text{ and } s \\ r|s & \text{the set of strings from } r \text{ or } s, r \cup s \\ & \text{for some regular expressions } r \text{ and } s \\ r^* & \text{the set of strings made from concatenating 0 or more strings from } r \\ & \text{for some regular expression } r \end{array} \right.$

Abbreviation	Meaning	Equivalence
$[abc]$	One occurrence of any of these symbols	$(a b c)$
$[a-g]$	One occurrence of any symbol in this range	$(a b c d e f g)$
$r?$	Optionally an occurrence of a string defined by r	$(r \epsilon)$
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	$(rrr rrrr rrrrr)$
r^+	One or more occurrences of a string defined by r	rr^*

- ▶ *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:*
 $[0-9]^+ [NSEW]^{0,2} [A-Z][a-z]^* (St|Ave|Rd|Ln|Dr|Terr|Blvd),$
 $([A-Z][a-z]^*)^*, [A-Z]^2[0-9]^5$.