CT 100

Strings, Regular Expressions
Strings

- Textual Data
  - Names
  - Id numbers
  - The SQL data type varchar is an example of a string data type
- A sequence of characters
  - Character code
- String literals
  - “Hello”
- String variables
  - Variables that hold or refer to a value that is a string
String Operations

• Indexing
  – “Hello”[2]

• Length
  – “Hello”.length

• Concatenation
  – “Hello” + “ How are you?”

• Searching
  – “Hello”.indexOf(“o”)
Patterns

- Defines properties of a set of strings
- Used for searching or checking if strings match expected formats
- Regular expressions
Regular Expression (R.E.) Patterns

- Any single character (except for a few special characters) is a regular expression.
- A period symbol (.) is a regular expression that matches any single character.
- If A and B are regular expressions then the following are also regular expressions:
  - AB: this R.E. matches strings where the first part of the string is matches A and the second part matches B
  - A|B: this R.E. matches strings that either match A or match B
  - (A): the R.E. matches strings that match A
Regular Expression (R.E.) Patterns

- $A^*$: this R.E. matches strings that are 0 or more repetitions of strings that match $A$
- $A^+$: this R.E. matches strings that are 1 or more repetitions of strings that match $A$
- $A?$: this R.E. matches the empty string or strings that match $A$
- $A^m$: this R.E. matches strings that match exactly $m$ repetitions of $A$
- $A^{m,n}$: this R.E. matches strings that match between $m$ and $n$ (inclusive) repetitions of $A$
- $A^{m,}$: this R.E. matches strings that match $m$ or more repetitions of $A$
R.E. Character Classes

- \d : this R.E. matches a single digit character
- \D : this R.E. matches a single character that is not a digit
- \w : this R.E. matches a single character that is a letter or digit
- \W : this R.E. matches a single character that is not a letter and not a digit
- \s : this R.E. matches a single whitespace character (space, tab, newline)
- \S : this R.E. matches a single character that is not a whitespace character
String Problems: Suppose S is a string variable whose values is “CT100”, T is a string variable whose value is “Computational Thinking” and i is an integer variable whose value is 4. What are the values of the following string operations?

- S.length
- T.length
- S[2]
- T[5]
- S[i]
- T[i]
- S.indexOf(“1”)
- S.indexOf(“0”)
- S.indexOf(“T”)
- T.indexOf(“p”)
- T.indexOf(“t”)
- T.indexOf(“T”)
- S+T
- T+S
- S+”:”+T
Which of the strings on the right match the R.E. on the left

- .
- .*
- \d
- \d*
- \D*
- \w*
- [0-9][0-9](\D)*
- .*(abc).*

- “x”
- “%23a”
- “9”
- “”
- “4192”
- “abc”
- “84#abcdef”
- “46”
- “81xyz”