Announcements:

- gradescope.com ⇝ @ugrad.cs.ubc.ca
  → CWL

Today:

Ch 0 – Languages (Alphabet, String/Word) "Problems" are languages

Ch 0 – Directed graph as algorithms

In article Information Theory, Counting Walks

Languages:

- Alphabet = Finite Set
- String over an alphabet, $\mathbb{A}$, is a finite sequence with elements from $\mathbb{A}$
- Symbols (letters) are elements of the alphabet
- Words means strings
- A language over an alphabet is a set of strings

EXAMPLES

$\{0,1\}$, $\{0,1,\ldots,9\}$, $\{a,b\}$

Strings over $\{0,1\}$:

$(0,1,1,0)$	0110
$(1,1,1)$	111
$(\ )$	ε Empty String
$(0,1,1,1,1)$	01111

Languages over $\{0,1\}$:

- All strings: $\{\varepsilon, 0, 1, 00, 01, 10, 11, \ldots\} = \{0,1\}^*$
- No strings: $\emptyset = \{\ \}$
- Strings with same # of 1's as 0's: $\{01, 10, 11100010, \ldots\}$

choose some representatives

SAME_NUM_1S_0S

Use all caps (typically)

LANGUAGES OVER $\{0,1,2,3,4,\ldots,9\} = A$

DIVISIBLE_BY_3 = { all strings that represent base 10 integers divisible by 3 }

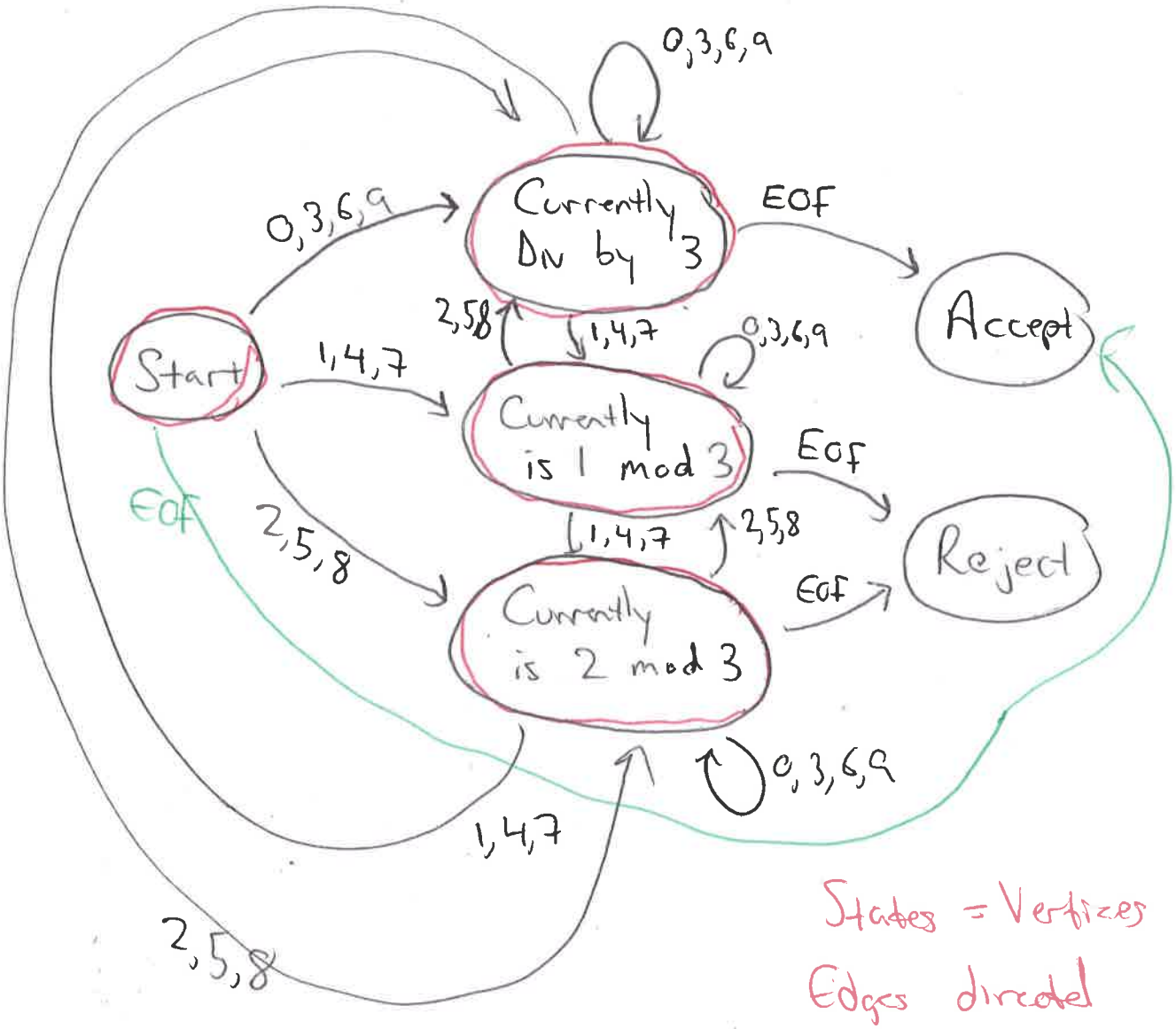
$= \{\varepsilon, 0, 3, 6, 9, 03, 06, 09, 12, 15, \ldots, 96, 99,$ 003, 006, $\ldots\}$

convention

If DIVISIBLE_BY_3 has fast algorithm, there are sublanguages that can be impossible to solve.

→ We say L_1 is a sublanguage of L_2 if $L_1 \subseteq L_2$.

E.g. L_1 = { all strings divisible by 3, that ~~when~~ viewed as only contain 0's and 1's that interpreted as binaries are programs that halt }



States = Vertices

Edges directed

Extra Info

Input: 1739216754 $\in$ DIV_BY_3

Algorithm = Directed Graph & Extra Info

Walks in directed graphs

As motivation: Morse Code

Morse: Dot - Space • 0.1 seconds
Dash - Space — 0.2 seconds

Alphabet = { •, — }

What is the information rate of Morse Code?

how many strings of Morse Code that take 20 seconds to transmit?

Idea: Alphabet {0,1}: # strings length n is 2^n

0 0 0 ... 0
1 1 1 ... 1

Alphabet {0,1,...,255} (byte) # strings length n is 256^n