CS 121, Section 1

Representing objects as strings

by Joanna Boyland
on September 10-16th, 2020

Section 1

* Review Lecture 3
* Exercises

» Representation Schemes

Definition: For a set of objects $\mathcal{O}$, a representation scheme is a one-to-one function E: $\mathcal{O} \to \{0,1\}^*$.

Examples:

- Binary representation $E : \mathbb{N} \to \{0,1\}^*$
- Representing rationals $E : \mathbb{Q} \to \{0,1\}^*$

» Prefix freeness

Definition: A representation scheme $E : \mathcal{O} \to \{0,1\}^*$ is prefix-free if for every $x \neq x'$, $E(x)$ is not a prefix of $E(x')$.
Two important theorems:

1. Prefix free $\implies$ can encode ordered pairs and thus lists
2. Every encoding can be made prefix-free
    a. "C style": null terminated
    b. "Pascal style": $|x|, x$

» Practice Problem

Show that we can transform any representation to a prefix-free one by a modification that takes a k bit string into a string of length at most $k + O(\log k)$.[1]

[1] For now, we will assume that the length of the standard binary encoding of a natural number is $\log k + O(1)$.

» Practice Problem

Represent

* 0 as 11010
* 1 as 11011
* 10 as 11000110
* 11 as 11000111
* 100 as 111101100
* etc

» Practice Problem Solution

Let E be a representation E: $\mathcal{O} \to \{0,1\}^*$. Let $G : \{0,1\}^* \to \{0,1\}^*$ be the C-style prefix-free modification of binary strings, sending 0 to 00, 1 to 11, and terminating strings with 01. Let $F : \{0,1\}^*$ be the Pascal-style modification, sending a string x to $|x|$ concatenated with x itself, with $|x|$ represented by the standard binary representation with the C-style modification.

We will prove that F is a prefix-free encoding that sends a k bit string to a string of length at most $k + O(\log k)$.

Lemma: We will prove that G is a prefix-free encoding. Take strings x and y. If $G(x) = G(y)$ then $x = y$, so G is an encoding. For every x, $G(x)$ has an even number of digits and ends with 01, but 01 never appears earlier in $G(y)$ for any y, if preceded by an even number of digits. Thus G is prefix-free.

» Practice Problem Solution, cont'd

Correctness: We will prove that F is an encoding. Suppose two strings x and y satisfy F(x) $=$ F(y). Thus, the strings |x|, x and |y|, y, where |x| and |y| are represented using the standard binary representation, modified by G, are the same. Then, since G is prefix-free, we have that x and y are the same. Thus F is an encoding.

We will prove that F is prefix-free. Suppose that $|x|, x$ is the prefix of $|y|, y$ for some strings x and y. Then (some prefix of) $|x|, x$ is the prefix of $|y|$. Since G is prefix-free, we must have that $|x| = |y|$. Thus the length of $|x|, x$ is the same as the length of $|y|, y$. Since we assumed that $|x|, x$ is the prefix of $|y|, y$, we have that $x = y$. Thus F is prefix-free.

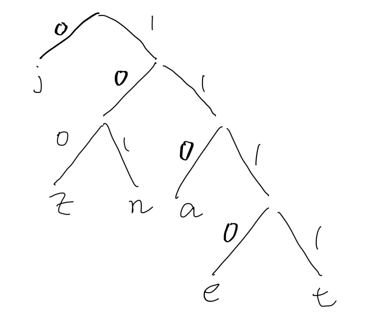
Efficiency: We can see clearly that G takes a string of length k to a string of length $2k + 2$. Thus, we have that F sends a string of length k to a string of length $2(\log k + O(1)) + 2 + k$, which is at most $k + O(\log k)$, as desired.

» Can we represent everything?

Short Answer: No
Longer Answer: For example, we can't represent the real numbers. Why? It would take an even longer answer. If you want to look it up for yourself, search for Cantor's Theorem. (This is not needed for class!)

1. Representing real numbers with binary? No!
2. Representing rational numbers with binary? Also no!

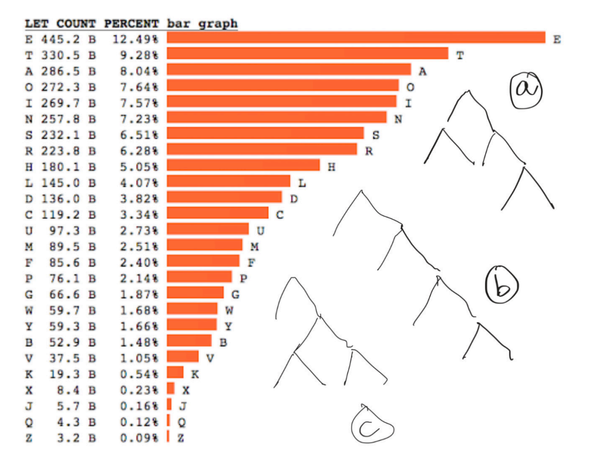
» Useful Skill: Representing things well



Representing representations!
Idea: Use a binary tree

» Problem 1

For each of the following sets of numbers, add the letters to the tree that is the most efficient representation:
{E, H, G, K, Z}, {O, I, N, D, C}, {T, L, U, M, P}.



LET	COUNT	PERCENT	bar graph
E	445.2 B	12.49%	E
T	330.5 B	9.28%	T
A	286.5 B	8.04%	A
O	272.3 B	7.64%	O
I	269.7 B	7.57%	I
N	257.8 B	7.23%	N
S	232.1 B	6.51%	S
R	223.8 B	6.28%	R
H	180.1 B	5.05%	H
L	145.0 B	4.07%	L
D	136.0 B	3.82%	D
C	119.2 B	3.34%	C
U	97.3 B	2.73%	U
M	89.5 B	2.51%	M
F	85.6 B	2.40%	F
P	76.1 B	2.14%	P
G	66.6 B	1.87%	G
W	59.7 B	1.68%	W
Y	59.3 B	1.66%	Y
B	52.9 B	1.48%	B
V	37.5 B	1.05%	V
K	19.3 B	0.54%	K
X	8.4 B	0.23%	X
J	5.7 B	0.16%	J
Q	4.3 B	0.12%	Q
Z	3.2 B	0.09%	Z

» Problem 2

Show that we can transform any representation to a prefix-free one by a modification that takes a k bit string into a string of length at most $k + \log k + O(\log \log k)$.[2]

[2] Hint: Think recursively how to represent the length of the string

» Problem 3

Let NtS be the function NtS: $\mathbb{N} \to \{0,1\}^*$ that sends a number to its binary representation.

(a) Prove that NtS satisfies that for every $n \in \mathbb{N}$, if x $=$ NtS(n) then $|x| = 1 + \max(0, \lfloor \log_2 n \rfloor)$

(b) Prove that NtS is in fact a representation, by showing that it is one-to-one: that is, find a function $StN : \{0,1\}^* \to \mathbb{N}$ such that $StN(NtS(n)) = n$.