Homework #5: Strings
Due Date: written 5pm Monday 4/25, program 11:59pm Tuesday 4/26

This is your last regular homework, concerning topics from Chapter 5. Our last lecture is Thursday 4/21. After that I’ll offer a review meeting on Tuesday 4/26 (during reading period), at our usual time and room. Our final exam is as scheduled by the registrar, at 3pm-5:30pm Monday 5/2 in our usual room. For the final exam we will have assigned seating and allow one sheet of notes, expect more instructions on that later.

Program Part

The program part of this homework will concern the Burrows-Wheeler Transform, which we will discuss in lecture and Piazza. Once it is available, it should be documented in share/hw5/Notes.txt.

Written Part

Problem 1. (Radix Sorts) This is based on exercises from page 726.

1(a). Give a trace (like bottom of page 707) of LSD sort on these strings:

```
no is th ti fo al go pe to co to th ai of th pa
```

1(b). Give a trace (like page 714) of MSD sort on these strings.

```
now is the time for all good people to come to the aid of their party
```

1(c). Indicate which characters of the above strings were NOT examined by MSD sort.

Problem 2. (Tries) Consider these strings again:

```
now is the time for all good people to come to the aid of their party
```

2(a). Draw the R-way trie that results after these strings inserted as keys into an empty trie (use a diagram like in the book or slides, where nulls are omitted). Assume each key is inserted with an integer value, indicating its position in the insertion order: so key “now” is inserted with value 0, key “is” is inserted with the value 1, key “the” is inserted (the first time) with the value 2, and so on. If a key (like “the”) is re-inserted, then its new value replaces its old value.

2(b). Draw the TST that results when the same strings and values are inserted into an empty TST. (Just do naive insertion, do not do any “rebalancing”).

Problem 3. (Huffman) Consider the string $S=\text{TOMGOTOTOGO!}$ (without the quotes).

3(a). Work out the Huffman code for $S$. Write down the code as both a codeword table, and as a binary trie. (Different students could get different answers here, since you could break ties and order siblings differently.)

3(b). How long is the bitstring encoding $S$, and how long is the bitstring encoding the code trie, assuming 8 bits per ASCII symbol? (You do not have to write out these two bitstrings.)
Problem 4. (LZW) For both parts below, assume that in our initial LZW code table 'A' is 41, 'B' is 42, end-of-file is 80, and the first available (undefined) code is 81. (In fact, these are ASCII codes written in hexadecimal.)

4(a). Show the complete sequence of codes emitted to compress “ABABABAABAAB”. Also, list the new code table entries (the first: 'AB' is 81).

4(b). Find the string encoded by this sequence, and also show your new code table definitions: 42, 41, 41, 81, 83, 85, 86, 42, 80