Problem Set 5

All parts are due on March 16, 2018 at 11PM. Please write your solutions in the LaTeX and Python templates provided. Aim for concise solutions; convoluted and obtuse descriptions might receive low marks, even when they are correct. Solutions should be submitted on the course website, and any code should be submitted for automated checking on `py.mit.edu/6.006`.

Problem 5-1. [35 points] **Dynamic Hash Browns**

(a) [5 points] Insert integer keys $(1, 6, 2, 4, 7, 8, 3)$ in order into a hash table using hash function $h(k) = (11 * k + 5) \mod 7$. Collisions should be resolved via chaining, where collisions are stored at the end of a chain. Draw a picture of the hash table after all keys have been inserted.

(b) [10 points] Given an array A containing n integers, describe an expected $O(n^2)$ time algorithm to determine whether two pairs of distinct integers from A have the same sum, i.e. there exists $\{a, b, c, d\} \subseteq A$ where $a + b = c + d$.

(c) [10 points] A dynamic array efficiently implements an **addressable stack** interface, supporting worst-case constant time array indexing, and insertion and removal of items at the back of the array (the largest index) in amortized constant time. But the same operations will take linear time when performed at the front of the dynamic array (at index zero), as items must be shifted down. Design a **double-ended addressable queue** data structure to store a sequence of items that supports **worst-case** constant time array indexing to the ith item in the stored sequence, as well as four dynamic operations: insertion and removal of items from both the front and back of the sequence, where each dynamic operation runs in at most **amortized** constant time.

(d) [10 points] Your favorite bookstore, Drytent, recently closed because of a fire. As a result of the fire, their books have been torn to pieces and scattered throughout the store. As you survey the damage, you find a browned scrap of paper upon which is written a passage of p beautifully written and particularly inspiring words, and you are determined to find the book it came from. The owners of Drytent give you access to an online repository containing text from the set B of books they sell, where the text from book $b \in B$ contains w_b words. You may assume that each word contains at most twenty characters. Describe an expected $O(p + \sum_{b \in B} w_b)$ time algorithm to determine which book contains the inspiring passage.

Problem 5-2. [20 points] **Trading Portfolios**

Herkshire Bathaway (HB) conducts business on the stock market, entering a huge number of trades every day. On the other side of each trade is some company which we call a **counterparty** of HB. Each trade that HB makes is assigned a portfolio name, though trades with different counterparties may be assigned the same portfolio, and trades with the same counterparty may be assigned different portfolios, according to an internal policy that cannot be questioned.

Whenever HB enters into a trade, a new line $(+, C, P)$ is appended to a daily log file: the plus sign indicates that a new trade was initiated by HB, the C indicates the counterparty identifier, and the P is the portfolio associated with that trade. However at any time during a day, a counterparty may cancel all trades with HB so far that day. Such a cancellation would be logged using a line such as $(-, C)$. Here, the minus sign indicates that a cancellation took place, with C identifying the counterparty who requested the cancellation. Here is an example of an HB daily log file.

```
1   +   MalWart      horizon
2   +   DomeHepot    emerald
3   +   MalWart      emerald
4   +   KolaKoka     diamond
5   -   MalWart
6   +   KolaKoka     horizon
7   +   FoleWhoods   platinum
8   +   KurgerBing   gold
9   -   DomeHepot
```

At the end of each day, HB needs to determine the set of distinct portfolios associated with trades that are still active, i.e. portfolios of trades that were not cancelled during the day. For the provided HB daily log, the portfolios associated with active trades at the end of the day are `horizon`, `diamond`, `platinum`, and `gold`.

(a) [10 points] Given an HB daily log containing n entries, describe a worst-case $O(n^2)$ time algorithm to determine the set of distinct portfolios still associated with active trades at the end of the day. You may assume that names of portfolios and counterparties can be compared in constant time.

(b) [10 points] Herkshire Bathaway is not happy with your quadratic solution and announces a round of layoffs. You decide that, for job security, you need to come up with a better algorithm. Describe an algorithm that produces a list of distinct portfolios associated with active trades which runs in expected $O(n)$ time.

Problem 5-3. [45 points] **Tedious Decryption**

Bob wants to send Alice a secret message of characters via a specially encoded sequence of integers. Each character of Bob's message corresponds to a **satisfying** triple of distinct integers from the sequence satisfying the following property: every permutation of the triple occurs consecutively within the sequence.

For example, if Bob sends the sequence $A = (4, 10, 3, 4, 10, 9, 3, 10, 4, 3, 10, 4, 8, 3)$, then the triple $t = \{3, 4, 10\}$ is satisfying because every permutation of t appears consecutively in the sequence, i.e. $(3, 4, 10)$, $(3, 10, 4)$, $(4, 3, 10)$, $(4, 10, 3)$, $(10, 3, 4)$ and $(10, 4, 3)$ all appear as consecutive sub-sequences of A.

For any satisfying triple, its **initial occurrence** is the smallest index i such that the triple is also satisfying, for the prefix of the input sequence ending at index i. For example, the initial occurrence of t in A is 10.

Bob's secret message will be formed by the characters corresponding to each satisfying triple in the encoded sequence, increasingly ordered by initial occurrence. To convert a satisfying triple (a, b, c) into a character, take the remainder of its sum divided by 27: remainders 0 to 25 correspond to the lower-case letters 'a' to 'z', while remainder 26 corresponds to a space character. For example, the triple $\{3, 4, 10\}$ corresponds to the letter 'r'.

(a) [5 points] Suppose Bob sends Alice the following sequence.

$$(10, 13, 9, 10, 13, 5, 2, 13, 5, 10, 9, 13, 10, 9, 13, 2, 5, 13, 2, 67, 23, 1, 2, 10, 1, 2, 1, 10, 2, 1)$$

Write down the list of satisfying triples contained in the sequence, as well as their associated characters, ordered by initial occurrence. What is the secret message?

(b) [10 points] Bob has started to send longer messages, and Alice is tired of decoding his messages by hand. Design an expected $O(n)$ time algorithm to decode a sequence of n integers from Bob, and output his secret message.

(c) [30 points] Write the Python function `decode_message(sequence)` that implements your decoding algorithm and returns the secret message. You may assume that all elements in the input list are non-negative integers. You can download a code template containing some test cases from the website. The template code contains a helper method `convert_to_char(k)`, which converts an integer k to a character according to the procedure described above. Submit your code online at `py.mit.edu/6.006`.

```
def decode_message(sequence):
    '''
    Decode hidden message based on permutation of triples
    Input:  list of non-negative integers
    Output: string corresponding to hidden message
    '''
```