

Data Structures, exercises week 8, 22-26.3

You can choose the programming languages freely, except in exercise 5 you need to use Java.

And once more: **do not execute the programs with these machines: melkki, melkinkari, melkinpaasi**

1. Change strings `QUEUE` and `STACK` into integers using the method of Cormen at page 230 (2nd edition) page 262 (3rd edition). Which string corresponds to integer 17625834700?

The following Java code tells the ASCII code of a character:

```
char merkki = 'A';
int koodi = (int)merkki;
System.out.println(koodi);
```

ASCII codes have been listed also in the following page:

<http://www.asciitable.com/>

2. Make a hash function (using any programming language) that maps a string into
 - (a) integer $i \in \{1, 2, \dots, 2^k\}$ where k is an integer
 - (b) integer $i \in \{1, 2, \dots, m\}$ where m is a prime
 - (c) so that it is suitable (when using double hashing) as the second hash function together with the function of b.
3. Implement (using any programming language) a hash table where collisions are resolved by chaining or that is based on open addressing. You can assume, that the keys in the hash table are strings. In this exercise it is adequate to implement operations `hash-search` and `hash-insert` only. It is adequate, that your table has a constant size, so rehashing is not needed.
4. Test the hash table in the previous exercise by inserting all the words in the novel Seven Brothers by Aleksis Kivi into the table. The novel is available here:

<http://www.gutenberg.org/files/11940/11940-8.txt>

Test different hash functions. If your implementation is based on chaining, find out what is the length of the longest chain and an average chain. How does the size of the table effect the length of chains?

If your implementation is based on open addressing, evaluate the performance using different sized arrays and different collision management strategies.

Reading a file and processing it word by word is done like this with Java:

```

import java.io.File;
import java.util.Scanner;

public class Main {
    public static void main(String[] args) throws Exception {
        // open file called seitseman.txt
        Scanner teksti = new Scanner(new File("seitseman.txt"));

        // continue until no words left
        while ( teksti.hasNext() ) {
            String mj = teksti.next(); // get the next word
            System.out.println(mj);
        }
    }
}

```

5. Compare the performance of your hash table with Java HashSet (hash table) and Java TreeSet (balanced search tree). If you didn't do the previous exercise, you can mark this exercise by comparing HashSet with TreeSet plus doing at least the first of the bonuses below. In week 4 exercises you can find the instructions how to test the running time of algorithms.

Do the following measurements:

- Add the words of Seven Brothers into the structures and measure the time that went into it.
- After doing the previous task, test with every word of Novel Crime and Punishment, is the word found in the set of words of Seven Brothers. Measure the running time time. Run through all the words of Crime and Punishment from the start to the end, the same word can repeat itself in the query as many times as it is found in the text.

Crime and Punishment can be found here:

<http://www.gutenberg.org/files/2554/2554-8.txt>

Bonus questions:

- TreeSet maintains keys in order. The keys found in TreeSet are possible to go through using an *iterator*. Find out how iterator works and print the words in Seven Brothers in alphabetical order.
Also the keys in HashSet are possible to go through with iterator. What is then the order of the keys?
- Print the common words of Seven Brothers and Crime and Punishment. Print every single same word only once.

6. Binary search trees are being used as keys of a hash table. You can assume, that the keys of the tree are integers. Design a hash function that maps a binary tree into an integer. Argue the following things in your solution: Can two different binary trees have the same integer values? Does every integer correspond to some binary tree?