

# Exam, DAT038/TDA417 (and DAT037/TDA416)

## Data structures and algorithms

Tuesday, 2020-01-14, 14:00–18:00, Johanneberg campus

---

**Examiner(s)** Peter Ljunglöf, tel. 0766–075561  
Nick Smallbone, tel. 0707–183062  
(will visit around 15:00 and 16:30)

**Allowed aids** None, but see the appendix (the last two pages)  
for a list of things to know, data structures, algorithms, etc.

**Exam review** When the exams have been graded they are available for review  
in the student office on floor 4 in the EDIT building.  
If you want to discuss the grading, please come to the exam review:  
  
Monday, 3 February, 10–12 and 13–15, in room 5128 (EDIT building)  
  
In that case, you should leave the exam in the student office until after the review.  
We will bring all exams to the review meeting.

**Notes** Write your anonymous code (not your name) on every page.  
You may answer in English or Swedish.  
Excessively complicated answers might be rejected.  
Write legibly – we need to be able to read your answer!  
You can write optional explanations on the question sheet or on a separate paper.

There are **6 basic questions**, and **3 advanced questions**, and **two points per question**.  
So, the highest possible mark is 18 in total (12 from the basic questions and 6 from  
the advanced questions). Here is what you need to do to get each grade:

- To pass the exam, you must get 8 out of 12 on the basic questions.
- To get a four, you must get 9 out of 12 on the basic questions,  
and also get 2 out of 6 on the advanced questions.
- To get a five, you must get 10 out of 12 on the basic questions,  
and also get 4 out of 6 on the advanced questions.

Grade	Total points	Basic points	Advanced
3	≥ 8	≥ 8	–
4	≥ 11	≥ 9	≥ 2
5	≥ 14	≥ 10	≥ 4

**Good luck!**

Write your anonymous code (*not* your name):

---

## Basic question 1: Complexity

The *symmetric difference* of two sets  $A$  and  $B$  is the set of elements which are in either  $A$  or  $B$  but not both. For example, the symmetric difference of the sets  $\{1, 2, 3\}$  and  $\{3, 4\}$  is  $\{1, 2, 4\}$ . In mathematical notation it is written  $A \oplus B$  and is defined as:

$$A \oplus B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$$

The following algorithm computes the symmetric difference of  $A$  and  $B$ :

```
S = new empty set
for every element x in A:
    if x is not in B:
        add x to S
for every element y in B:
    if y is not in A:
        add y to S
return S
```

### 1A) What is the order of growth (complexity) of this algorithm, if the sets are implemented as unordered linked lists?

Assume that  $A$  and  $B$  have the same size  $N$ , and give your answer in terms of  $N$ . You can assume that the size of the final set  $S$  is approximately  $N$  too.

Order of growth: \_\_\_\_\_

*Optional explanation:*

### 1B) What if the sets are implemented as red-black trees?

*For old students:* If the course did not include red-black trees when you studied it, you may replace red-black trees with AVL trees.

Order of growth: \_\_\_\_\_

*Optional explanation:*

---

Points for question 1 (to be filled by the grader):

Write your anonymous code (*not* your name):

## Basic question 2: Quicksort

### 2A) Is the following statement true or false?

“When using quicksort with the median-of-three strategy for choosing the pivot, the worst-case runtime is linearithmic, i.e.  $O(n \log n)$ .”

(Assume that the version of quicksort used does *not* shuffle the array before sorting.)

Answer: \_\_\_\_\_

*Optional explanation:*

### 2B) Perform a quicksort partitioning of the following array, using the median-of-three strategy for choosing the pivot

0	1	2	3	4	5	6	7	8
11	24	27	51	32	5	44	8	20

Show what the partitioned array looks like afterwards.

Remember to swap the pivot with the first element before partitioning.

Also note which sub-arrays need further sorting in a recursive call (e.g., by drawing curly braces under each sub-array, or by crossing over the cells that do not need sorting).

**Answer:**

0	1	2	3	4	5	6	7	8

*Optional explanation:*

Points for question 2 (to be filled by the grader):

Write your anonymous code (*not* your name):

---

### **Basic question 3: Binary search trees**

**3A) Insert the items 3, 4, 5 (in that order) into an empty red-black tree, and draw the resulting tree below.**

If you don't have a coloured pen, draw red nodes as a square and black nodes as a circle. Use a separate sheet of paper if you run out of space.

*For old students:* If the course did not include red-black trees when you studied it, you may replace red-black trees with AVL trees.

**3B) Now insert the items 1, 2 (in that order) into your answer from part A, and draw the resulting tree.**

---

Points for question 3 (to be filled by the grader):

Write your anonymous code (*not* your name):

## Basic question 4: Hash tables

Suppose you have the following hash table, implementing using *linear probing*.

0	1	2	3	4	5	6	7
12345			658	31	74		124

The hash function is the digit sum modulo 8:

$$h(d_k \dots d_2 d_1 d_0) = (d_k + \dots + d_2 + d_1 + d_0) \bmod 8.$$

Which means that e.g.  $h(7403) = (7+4+0+3) \bmod 8 = 14 \bmod 8 = 6$ .

**4A) Think about the order in which the numbers could have been inserted into the hash table. Which numbers could have been the first one to be inserted?**

(Assuming that it hasn't been resized, and no elements have been deleted)

There are several possible answers; **circle all** that could have been the first one.

12345      658      31      74      124

*Optional explanation:*

**4B) Insert the numbers 374 and 5294, in that order, and give their hash values.**

Assume that no resizing occurs.

0	1	2	3	4	5	6	7
12345			658	31	74		124

$h(374) =$  \_\_\_\_\_       $h(5294) =$  \_\_\_\_\_

*Optional explanation:*

Points for question 4 (to be filled by the grader):

Write your anonymous code (*not* your name):

## Basic question 5: Priority queues

Take a look at the following three arrays of integers:

	1	2	3	4	5	6	7	8	9	10
<b>A:</b>	1	2	3	3	5	7	3	6	8	4

<b>B:</b>	3	6	4	5	7	5	8	6	8	9
-----------	---	---	---	---	---	---	---	---	---	---

<b>C:</b>	2	4	5	4	6	8	7	9	5	6
-----------	---	---	---	---	---	---	---	---	---	---

**5A) One of these arrays is a binary heap (a min-heap). Which one?**

Answer: \_\_\_\_\_

*Optional explanation:*

**5B) Remove the smallest item from the heap and show how it looks afterwards.**

	1	2	3	4	5	6	7	8	9	10
Answer:										

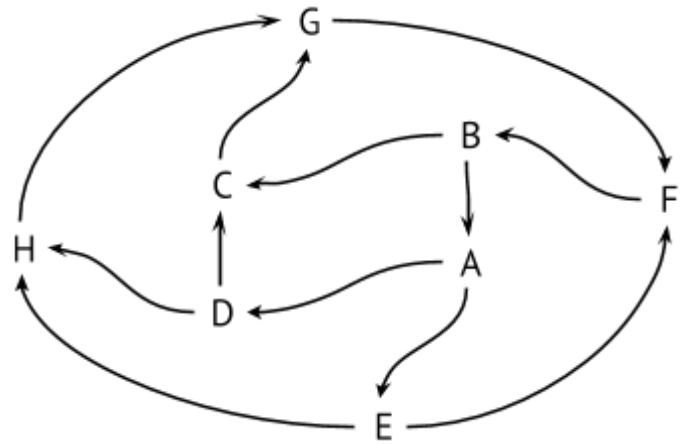
*Optional explanation:*

Points for question 5 (to be filled by the grader):

Write your anonymous code (*not* your name):

## Basic question 6: Graphs

You are given the directed graph to the right.



### 6A) Perform a depth-first search over the graph, starting from node A.

Break ties alphabetically (i.e., node M should be tried before node N).  
In which order are the nodes visited?

Visiting order: \_\_\_\_\_

*Optional explanation:*

### 6B) Make the graph a DAG by removing one edge, and give its topological order.

Remove one edge from the graph, so that it becomes a DAG.

Edge to remove: \_\_\_\_\_

Give one possible topological order of the resulting DAG, as an ordered list of nodes.

Topological order: \_\_\_\_\_

*Optional explanation:*

Points for question 6 (to be filled by the grader):

Write your anonymous code (*not* your name):

---

## Advanced question 7: Mystery program

The following algorithm takes two lists of integers  $A$  and  $B$ , and returns a Boolean result:

```
M = new hash table mapping integers to integers
for every element x in A: M.put(x, 0)
for every element x in B: M.put(x, 0)
for every element x in A:
    n = M.get(x)
    M.put(x, n+1)
for every element x in B:
    n = M.get(x)
    M.put(x, n-1)
for every key x in M:
    if M.get(x) != 0:
        return false
return true
```

**7A) What does M contain after running the algorithm with the following inputs?**

A = {1, 2, 3, 2, 3, 4}  
B = {2, 4, 5, 3, 2, 2}

**Answer:** M = {1: \_\_\_\_\_, 2: \_\_\_\_\_, 3: \_\_\_\_\_, 4: \_\_\_\_\_, 5: \_\_\_\_\_}

**7B) For what inputs does the algorithm return true?**

**Answer:**

---

Points for question 7 (to be filled by the grader):



Write your anonymous code (*not* your name):

---

## Advanced question 8: Symmetric difference

*Answer on a separate sheet of paper.*

Question 1 was about calculating the symmetric difference  $S = A \oplus B$  of two sets.

In this question, you will design an algorithm for calculating the symmetric difference, assuming that all sets are implemented as *sorted linked lists*. To simplify the pseudocode, we assume that the elements are integers.

Assume that the sets  $A$  and  $B$  are objects of a class `List` which represents a linked list of integers. The `List` class provides the following constructor and methods which you are allowed to use:

- `new List()`: create a new, empty list
- `void add(int x)`: append an element to the end of the list
- `boolean isEmpty()`: return **true** if the list is empty
- `int remove()`: remove and return the head (first element) of the list (throws an error if the list is empty)

All these operations take constant time.

Note that `add()` and `remove()` do not care if the list is sorted or not, that is entirely up to you to ensure. However, you can assume that  $A$  and  $B$  are sorted.

**To get one point**, you should design an algorithm that calculates  $S = A \oplus B$ .

Note that the list  $S$  that you create must be sorted, and must be an object of class `List`. Also note that you will destroy the lists  $A$  and  $B$  when iterating over them, but that's OK.

Write down your answer either as pseudocode or Java code. The time complexity (order of growth) of your algorithm must be linear in  $|A|+|B|$ , and you have to explain why this is the case.

**To get two points**, your algorithm should, in addition to calculating  $S$ , also calculate the intersection  $T = A \cap B$ , and the union  $U = A \cup B$ , *in one single pass*.

As before, the time complexity must be linear in  $|A|+|B|$ , and you must explain why.

---

Points for question 8 (to be filled by the grader):

Write your anonymous code (*not* your name):

---

## Advanced question 9: Sum of $K$ smallest values

*Answer on a separate sheet of paper.*

Design a data structure `MinSum` which allows you to find the sum of the  $K$  smallest values in a collection of integers. It should support the following operations:

- `new MinSum(int K)`: create a new collection, initially empty
- `void add(int x)`: add an integer  $x$  to the collection
- `int minSum()`: return the sum of the smallest  $K$  values in the collection

You only need to support these three operations, so you do not need to include e.g. operations which find what integers have been added to the collection.

Write down:

- The names and types of the fields that your data structure will use.
- How each operation should be implemented, either in pseudocode or in Java code. You may use any of the data structures and algorithms listed in the appendix, without explaining how they are implemented.
- What order of growth each operation has, and why.

**To get 1 point**, your operations should have the following order of growth (complexity):

- `new MinSum(K)`: constant time
- `add(x)`:  $\log K$  (logarithmic time)
- `minSum()`:  $K$  (linear time)

**To get 2 points**, your operations should have the following order of growth:

- `new MinSum(K)`: constant time
- `add(x)`:  $\log K$  (logarithmic time)
- `minSum()`: constant time

---

Points for question 9 (to be filled by the grader):

# Appendix: List of concepts, algorithms & data structures

You may use any of the algorithms or data structures mentioned below (and on next page) in your answers, including the standard methods for each data structure, unless the question specifies otherwise.

**Orders of growth:** (in alphabetical order)

constant, cubic, exponential, linear, linearithmic, logarithmic, quadratic

**Trees:** size, depth, (un)balanced tree, complete tree, traversal (inorder, preorder, postorder)

**Hash tables:** load factor, uniform hashing assumption

**Graphs:** connected components, cycles/acyclic, topological order, minimum spanning tree (MST), shortest path tree (SPT)

**Insertion sort:**

1. select next element
2. insert in its correct position
3. repeat

**Selection sort:**

1. select smallest element
2. add to the end of the sorted array
3. repeat

**Mergesort:**

1. split array
2. recurse
3. merge results

**Quicksort:**

1. select pivot
2. partition array
3. recurse

**Key-indexed counting:**

1. count frequencies
2. calculate indices
3. move elements to new positions

**LSD string sort:**

1. perform key-indexed counting on every character position, starting from least significant position

**Linear search:**

1. compare each element in array

**Binary search:**

1. find middle element
2. compare
3. repeat

**Binary heap insert:**

1. add at end
2. swim it up

**Binary heap delMin:**

1. replace root with end element
2. sink it down

**DFS graph search:**

1. mark node
2. recursively visit unmarked neighbours

**BFS graph search:**

1. select next node in queue
2. mark it
3. add unmarked neighbours to queue

**Prim's algorithm:**

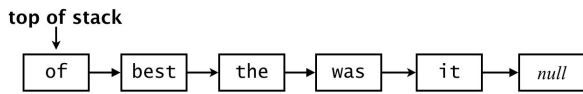
1. add to MST the min-weight edge that grows MST by one node, that is not yet in the MST
2. repeat

**Dijkstra's algorithm:**

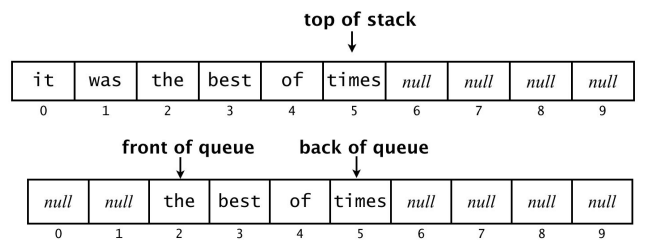
1. add to SPT the node with lowest cost from start node, that is not yet in the SPT
2. repeat

# Appendix: List of concepts, algorithms & data structures

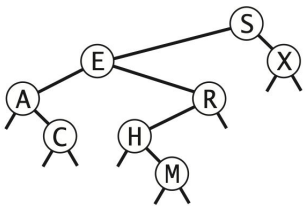
## Stack/queue as a linked list



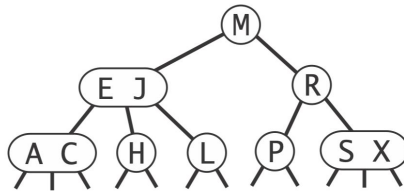
## Stack/queue as an array



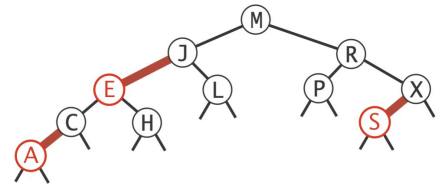
## Binary search tree



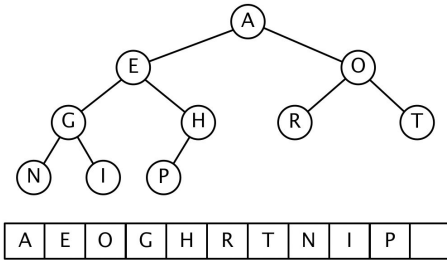
## 2-3 tree



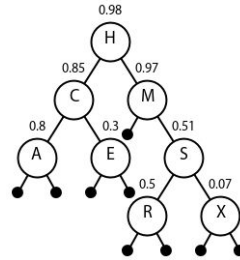
## Red-black tree



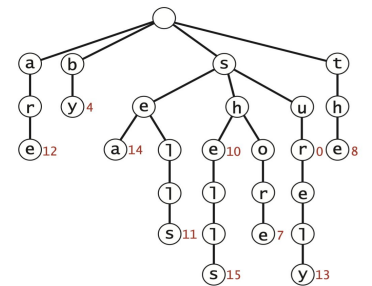
## Binary (min) heap



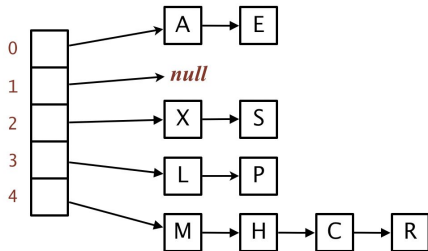
## Treap



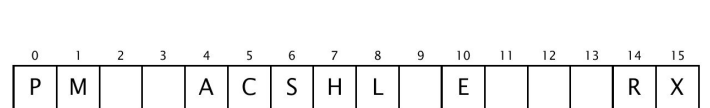
## R-way trie



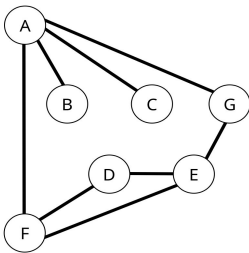
## Hash table: Separate chaining



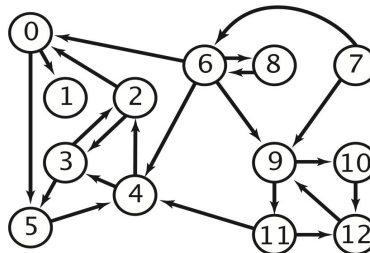
## Hash table: Linear probing



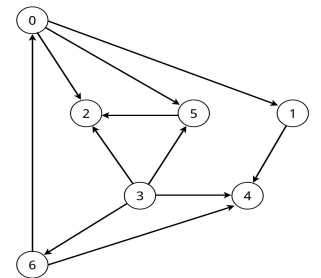
## Undirected graph



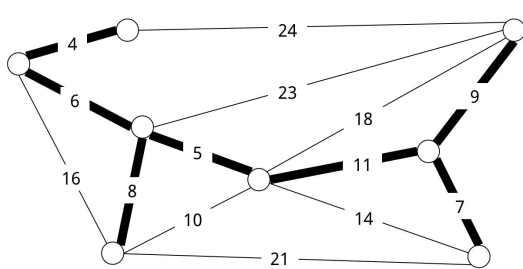
## Directed graph



## DAG



## Minimum spanning tree



## Shortest path tree

