

Ordinary Written exam

MED8

Algorithms, Data Structures and Software Engineering for Media Technology

Wednesday June 13th 2018

Name: _____

Cpr.no.: _____

Study no.: _____

Algorithms, Data Structures and Software Engineering for Media Technology

Ordinary Examination

13 June 2018

Instructions

- You have 3 hours to complete this examination.
- Neither electronic devices nor written material are allowed in the examination room.
- This examination consists of 10 questions. The maximum number of marks that can be obtained is 50. You must obtain at least 25 marks to pass.
- Do not write any answers on this question paper—answers written on the question paper will be ignored by the examiner. Write all your answers on the writing paper provided.
- Do not write your answers in pencil and do not use a pen with red or green ink. Use a pen with blue or black ink.
- Hand in no more than one answer to each question.
- Do not turn over until you are told to do so by the invigilator.

Question 1

For each of the following equations, state whether it is true or false.

a) $3n^2 + 2n + 100 = \Theta(n^2)$

True

b) $4n \log(n^2) = \Theta(n \log n)$

True

c) $4n^2 = O(n^3)$

True

d) $2^n = O(n^2)$

False

e) $100^{\log n} = \Omega(2^n)$

False

f) $0.0001n^3 = \Omega(n^3)$

True

g) $4n^2 = o(n^2)$

False

h) $3n \log n = o(n^2 \log n)$

True

i) $9n^3 + 3n^2 + 2n + 5n \log n + 1000000 = \omega(n^3)$

False

j) $4n^2 \log n = \omega(n^2)$

True

[1/2 mark for each correct part]

Question 2

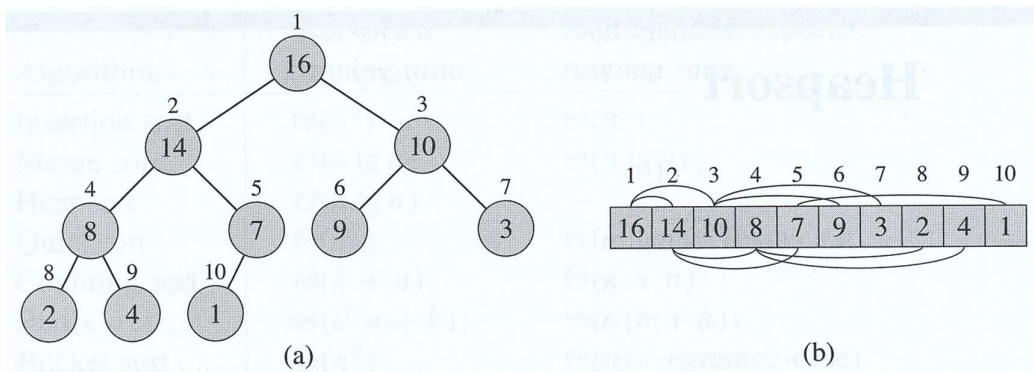


Figure (a) above shows a binary tree and figure (b) shows a representation of this tree as an array (1-based indexing is used).

- a) State the max-heap property. [1 mark]

The key at every node is less than or equal to that of its parent.

- b) Is the tree in Figure (a) above a max-heap? [1 mark]

Yes.

- c) Suppose n is a node in a binary tree and let i be the index of node n in an array representation of the tree, when the tree is represented in the same manner as that in which the array in Figure (b) above represents the binary tree in Figure (a). Write down algorithms that take i as input and efficiently compute the index in the array for (1) the parent of n , (2) the left child of n and (3) the right child of n . For each of the three algorithms, explain how it can be implemented efficiently using bit-shifting. [3 marks]

PARENT(i)

```
1 return [i/2]
```

LEFT(i)

```
1 return 2i
```

RIGHT(i)

```
1 return 2i + 1 [1/2 mark each]
```

Parent can be implemented by shifting binary representation of i one bit to the right. [1/2 mark]

Left can be implemented by shifting binary representation of i one bit to the left. [1/2 mark]

Right can be implemented by shifting binary representation of i one bit to the left and then adding 1 as the low-order bit. [1/2 mark]

Question 3

- a) Write down the operations that a data structure must support if it is to function as a *dictionary*. [1 mark]

Insert, delete and test for membership.

- b) What is the worst-case running time for searching for an element in a hash table? Under what circumstances does this worst case occur. [2 marks]

$\Theta(n)$ where there n items stored in the table. This occurs when all values have the same hash code and are thus hashed to the same slot.

- c) What is the expected running time for searching for an element in a hash table? Under what conditions does this expected running time occur? [2 marks]

$O(1)$. Occurs when there is simple uniform hashing, that is, when the expected length of a chain hanging off a hash slot is equal to the load factor, $\alpha = n/m$ where n is the number of items in the hash table and m is the number of slots in the hash table.

Question 4

Study the following code example and answer the questions that follow it.

```
1 #include <stdio.h>
2 #include <stdlib.h>
3 #define N 3
4
5 __global__ void addVectors(float* dev_a, float* dev_b, float* dev_c) {
6     int i = blockIdx.x;
7     if (i < N) {
8         dev_c[i] = dev_a[i] + dev_b[i];
9     }
10 }
11
12 int main(void) {
13     float a[N] = {1,2,3}, b[N] = {4,5,6}, c[N];
14     float *dev_a, *dev_b, *dev_c;
15
16     cudaMalloc((void**)&dev_a, N*sizeof(float));
17     cudaMalloc((void**)&dev_b, N*sizeof(float));
18     cudaMalloc((void**)&dev_c, N*sizeof(float));
19
20     cudaMemcpy(dev_a, a, N*sizeof(float), cudaMemcpyHostToDevice);
21     cudaMemcpy(dev_b, b, N*sizeof(float), cudaMemcpyHostToDevice);
22
23     addVectors<<<N,1>>>(dev_a, dev_b, dev_c);
24
25     cudaMemcpy(c, dev_c, N*sizeof(float), cudaMemcpyDeviceToHost);
26
27     cudaFree(dev_a);
28     cudaFree(dev_b);
29     cudaFree(dev_c);
30
31     for(int i = 0; i < N; i++)
32         printf("%.0f ", c[i]);
33
34     return EXIT_SUCCESS;
35 }
```

- a) In which line or lines is memory allocated on the GPU? [1 mark]

Lines 16-18.

- b) How many blocks of threads is the kernel run on? [1/2 mark]

N or 3

How many threads are there in each block? [1/2 mark]

1

- c) What is the purpose of line 25? [1 mark]

Copies the contents of the dev_c array on the GPU (device) to the c array which is in memory accessible from the CPU (host).

- d) In which line or lines is the kernel defined and how do we indicate to the compiler that the function is to be used as a GPU kernel? [1 mark]

Lines 5-10. Indicated by preceding the definition with the __global__ keyword.

- e) What is the purpose of the variable blockIdx, used in line 6 and what is its type? [1 mark]

blockIdx is a dim3 type variable that stores the identifying index of the current block. The block is identified by its co-ordinates within a 1-, 2- or 3-dimensional grid.

Question 5

- a) What is IaaS, PaaS and SaaS? Provide some brief descriptions and examples.
- ***Infrastructure as a Service (IaaS) provides a capability for users to provision processing, storage, and network resources on demand. Examples: cloud computing (Amazon EC2), storage services (Amazon S3), etc.***
 - ***In Platform as a Service (PaaS) the service provider makes certain core components such as databases, queues, workflow engines, email, and so on, available as services to the customer. Examples: cloud databases (eg Amazon DynamoDB), etc***
 - ***In Software as a Service (SaaS), typically, third-party providers using a subscription model provide end user applications to their customers. Examples: Machine learning, Computer vision (eg Amazon Rekognition)***
- b) `ssh -i file.pem ec2-user@{IP_Address}`
After setting up an AWS EC2 server, what is the use of the above shell command? What is the use of “file.pem”?
- This command is used to login (ssh) to the EC2 server. The file.pem is the key pair file used to identify the user who logs in to the EC2 server.***
- c) What is the python library Boto?
- Boto is a Python package that provides interfaces to Amazon Web Services***
- d) With which Amazon service you can create the AWS Access Key ID and AWS Secret Access Key?
- With AWS Identity and Access Management (IAM)***
- e) What is the advantage of using a Content Delivery Network (CDN)?
- Most cloud providers maintain a distributed set of servers in multiple data centers around the globe. These servers are used to make it easy to use Content Delivery Network (CDN) to serve content to end users from locations closest to them.***

[1 mark for each correct part]

Question 6

- a) What is the difference between Amazon’s EC2 and Lambda?
- EC2 is providing a virtual machine (server) which is also called instance.***

***Lambda is following a serverless architecture (lets you run code without provisioning or managing servers: serverless).
So: server vs serverless***

- b) What is Multi-tenancy in software architecture?
Multi-tenancy is a principle in software architecture where a single instance of the software serves multiple tenants or customers.
- c) Aiming at sizing the infrastructure for better performance, how can you do vertical scaling and how can you do horizontal scaling?
Vertical scaling: bigger machines
Horizontal scaling: more machines being deployed.
- d) What is a container and which tool you can use for this?
A container is a class, a data structure, or an abstract data type (ADT) whose instances are collections of other objects
Docker is a widely used container tool that can package an application and its dependencies in a virtual container that can run on any Linux server
- e) What is polyglot persistence?
Using multiple data storage technologies, chosen based upon the way data is being used by applications, is called polyglot persistence

[1 mark for each correct part]

Question 7

- a) What are semantic whitespaces in Python? [1 mark]

The use of indentations to define statements and scopes in the code.

- b) Which libraries, for Python, are used for working with multidimensional arrays? [1 mark]

Numpy, Scipy and (optional) Pandas

- c) In the following code, assume that **a** is an array. Explain what elements of the array each of the three lines would return. [1 marks]

```
a[-2:-1]
a[-10:]
a[10:]
```

1. Returns from the second to last to the last element of the array →
The second to last element
2. Returns the last ten elements
3. Returns from element 10 to the end of the array

Question 8

- a) Explain what structured data is? [1 mark]

Structured data consists of fields with defined type and size. Fields could be name, age, gender and address in a CSV file or database.

- b) Explain what unstructured data is? [1 mark]

Unstructured data doesn't have a predefined data model or schema, and only has an internal structure. Examples are text, images, video and audio.

- c) In the Python library **Scikit**, what is an estimator object? Name at least one specific class. [1 mark]

An estimator is a classifier, regressor or unsupervised learning model. An example could be LogisticRegression or RandomForestClassifier.

- d) In the Python library **Scikit**, what is a transformer object? Name at least one specific class. [1 mark]

Transformer objects are using to manipulate data e.g. preprocessing, such as scaling or reducing dimensionality. Examples could be PCA or MinMaxScaler.

- e) In the Python library **Scikit**, what are the arguments to create a machine learning pipeline? [2 marks]

The arguments are a list of (name, transformer) tuples with the final tuple being (name, estimator).

Question 9

- a) Explain the concept of **ensemble learning**. [2 marks]

Ensemble learning uses multiple learning algorithms, or same algorithms with different hyperparameters, to obtain better predictive performance than using a single algorithm.

- b) How does **bootstrap aggregation** work? [2 marks]

Given a training set D, bootstrap aggregation (bagging) generates M new datasets using uniform sampling with replacement. M models are fitted using the M datasets and their output is averaged for regression and voting is used for classification.

- c) How does **boosting** work? [2 marks]

Boosting works by combining multiple weak classifiers into a single strong. Emphasis on learning misclassified examples is done by increasing sample weights when training a new weak classifier. The final classifier is a linear combination of the weak classifiers.

Question 10

Study the following code and answer the questions that follow.

```
from keras.models import Sequential
from keras.layers import Dense

model = Sequential()
model.add(Dense(128, input_dim=128, activation='relu'))
model.add(Dense(64, activation='relu'))
model.add(Dense(128, activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dense(2, activation='softmax'))

model.compile(optimizer='sgd', loss='binary_crossentropy')
```

a) How many layers are in the neural network **model**? [1 mark]

5 layers

b) How many units does the last hidden layer contain? [1 mark]

32 units

c) What algorithm is used to train the neural network? [1 mark]

Stochastic Gradient Descent

d) What error measure is used for optimizing the weights of the neural network? [1 mark]

Log Loss (Binary Crossentropy)

e) What is the activation function of the output layer? [1 mark]

Softmax activation function

END OF EXAMINATION