Ph.D. Comprehensive Examination

Computer Science Department
University of Miami

August 14, 2020

Student Name: Hao Zhu Student Number: C12203432

Problem number	Points (10 max)
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
Total:	

1. **I.A Data organization; III.A Algorithms and complexity**

Each item x in a set S has a unique key $key[x]$. We need to implement the following operations.

(a) Search (S, key)
(b) Insert (S, x)
(c) Successor (S, x)
(d) Predecessor (S, x)

Give the 4 running times as an O() for the following implementations:

(a) Ordered (sorted) array,
(b) Ordered doubly linked list,
(c) Min-Heap, and
(d) Hash table

(a.) (a). search: O(log n) (b) Insert: O(n) (c) Succ: O(1) (d:). pred: O(1)

b. (a). search: O(n) (b) Insert: O(n). (c). Succ: O(1). (d) Pred: O(1)

c. (a). search: O(n), (b) Ins: O(log n). (c). Succ: O(n), (d). Pr: O(n)

d. (a) Search: O(1), (b) Insert: O(1), c: Succ: O(n) (d). Pred: O(n)

2. **I.B Program control and structure; I.C Programming language and notations**

Suppose that procedure *swap* is declared as follows:

```
procedure swap( x, y: integer);
    procedure f(): integer;
        var z: integer;
        begin // f
            z = x; x = y; return z;
        end // f
    begin // swap
        y = f();
    end // swap
```

Describe the effect of the procedure call *swap*(i, A[i]) under each of the following parameter passing methods:

(a) Call-by-value

(b) Call-by-reference

(c) Call-by-value-result

(a). Call-by-value:
pass the copy of x, y, into swap(). the globel x, y's value won't be change. return to a new z value = x

(b). Call-by-reference
pass the reference of x, y, into swap(). the globel x, y, value will be changed. return to z value = globel y value

(c). pass the value of x, y, into swap() before the swap() finished. the value of globel x, y won't be changed. When swap() finish, x is into globel y, so the return z should equals to y.

3. **I.D Software engineering**

From the software engineering point of view, any software development process can be divided into several sub-disciplines:

(a) Requirement Analysis
(b) Functional Specification
(c) Architectural Design
(d) Implementation
(e) Testing and Evaluation
(f) Maintenance

Choose three sub-disciplines or tasks within these sub-disciplines that involve a mathematical approach, and illustrative them with examples.

(a). Requirement Analysis.

Gathering the requirements from the customers, and try to fulfill their expections. Analysis other fectors like money cost, finish time etc.

Goal $= \sum_{i=1}^{n}$ requirments,

Min(Cost)

(b). Functional specification.

Due to the goal, make sure how many functions the program should be contained. Like, a online game should be included with, Art, programming, drame, background network, communication parts.

Main software $= \sum_{i=1}^{n}$ functions.

(c). Implementation.

After having the architectural design, need to Convert the design into code. Usually in a big company or studio, it should be seperate in to different modules. P.Code $= \sum_{i=1}^{n}$ modeli,

4. **I.E Systems**

(a) _____ linked libraries can support shared library code, allowing one copy of a library routine to be used by several different processes.

absolute **relative** **static** **dynamic** **none of these is correct**

(b) When it is not known at compile time where a process will reside in memory, _____ code must be generated.

logical **physical** **absolute** **relocatable**

(c) A UNIX process calls *fork()* to create a child process as shown: *pid = fork()*;

i. What value will be assigned to pid in the parent process by the call to *fork()*?

the parent's process id **the child's process id** **zero none of these**

ii. What value will be assigned to pid in the child process by the call to *fork()*?

the parent's process id **the child's process id** **zero none of these**

(d) The Banker's algorithm is used for deadlock _____.

denial **prevention** **avoidance** **recovery**

(e) Belady's anomaly can affect the performance of the _____ page replacement algorithm.

FIFO **LRU** **optimal** **SJF**

(f) _____ access files are made of fixed length records that allow programs to read and write records in no particular order.

sequential **direct** **logical** **none of these is correct**

(g) When an I/O request is being handled for a user's process, which term refers to the policy of returning control to the user process before the I/O is completed?

synchronous I/O **asynchronous I/O** **delayed I/O** **none of these**

(h) Which multithreading model requires that a new kernel thread be created for each new user thread?

many-to-one **one-to-one** **many-to-many** **none of these is correct**

(i) A process that does not affect, and is not affected by, another process is referred to as:

static **independent** **cooperating** **dynamic** **unbounded**

5. **II Computer Organization**

(a) Draw an architecture of a quad-core processor and discuss the role of each module in your diagram.

(b) Find a binary representation of the decimal number 0.1.

in each Core: CPU, Alu, registers, controll units.

system bus connects Cores with, thee main memories and I/O system.

CPU contains, Alu and controll units,

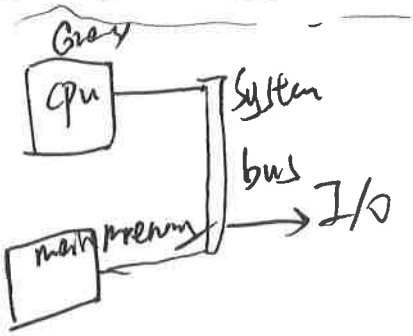
main memory contains the date and instructions

I/O system is controlled by the control units in CPU.

in L1 cache: contains date and instructions from mem

in L2 cache: contains left addresses in L1

in L3 cache: contains shared address for cores



b)

$0.1 \times 2 = 0.2$
$0.2 \times 2 = 0.4$
$0.4 \times 2 = 0.8$
$0.8 \times 2 = 1.6$
$0.6 \times 2 = 1.2$
$0.2 \times 2 = 0.4$
$0.4 \times 2 = 0.8$
$0.8 \times 2 = 1.6$
$0.6 \times 2 = 1.2$
$0.2 \times 2 = 0.4$
$0.4 \times 2 = 0.8$
$0.8 \times 2 = 1.6$
$0.6 \times 2 = 1.2$
$0.2 \times 2 = 0.4$
$0.4 \times 2 = 0.8$
$0.8 \times 2 = 1.6$

, so 0.1 could be represented as

0.0001|0011|0011|0011|

6. **II.D Networking and Communications**

(a) Draw a diagram showing layers of the Internet Protocol Stack and briefly discuss role of each layer.

(b) Describe functions of each layer when a file is transferred from a source to destination using (*file transfer protocol* (FTP).

(a) 1. Physical: the medium like cabel, wire, and wifi which like the package truck used to deliver pakage

2. Datalink: imputed on network interne card. each has eMAL, when connectc between host and router, pick which one to connected based on 4 type services. different

3. Neework: search the way established by Datalink, by the protocol TCP/IP or other rules. In this process. like to find a way for A to B by address. network will find the right way. fast without conflict.

4. Transport. Get the connection between the hose and the remote There has two type, TCP and UDP, one need get full-connect they other not.

5. Applicaton. You see the package, and open it, like open Email or play oplen client of online game.

(b). Transport layer FTP use two ports of TCP on transport layer. Port 20 and Port 21. One used to the trans data, others used to trans controll information, like username, password.

In. It need to get open a gate for doing transpore, when finished, close the gate. Applicaton layer may be used to creer the account

7. **III.A Algorithms and complexity**

Describe an algorithm that takes two input lists of integers $A = a_1, ..., a_n$ and $B = b_1, ..., b_m$ and delivers the list of all the elements that belong to A but not to B. A and B do not contain redundant elements, however, the elements of A and B might have a large range.

The algorithm should run in $O(n \log m + m \log m)$ time.

(1) first sort A and sort B.
this will take $O(n\log m + m\log m)$ time

(2). iterate all the item in A, and compare with the item in B one by one. i,j, from 0. j++.

if A[i] < B[j], A[i] is not in B, i++

if A[i] == B[j], A[i] is in B, skip.

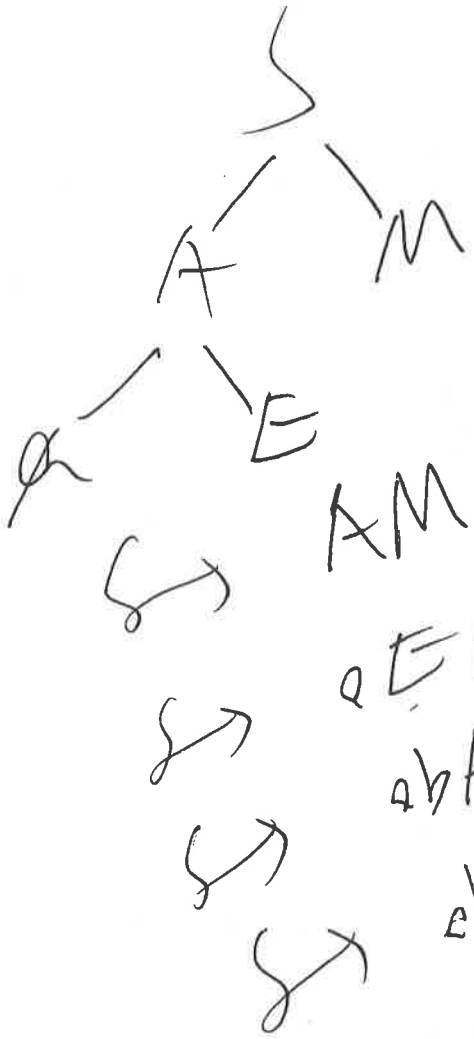
if A[i] > B[j], A[i] is not in B, j++

when the iteration is done, all the parsed A are not in B. if their has item rest in A, but B is finished, the rest A can't be in B. add rest A and Parsed A.

the time for search will are $O(n+m)$.

So the algorithm is $O(n\log m + m\log m) + O(n+m)$

$= O(n\log m + m\log m)$



$\rightsquigarrow AM$

$\rightsquigarrow aEM$

$\rightsquigarrow abAM$

$\rightsquigarrow abaEM$

$abaeM$

$\rightsquigarrow AM$

$\rightsquigarrow aEM$

$\rightsquigarrow abAM$

$\rightsquigarrow abaEM$

$abeS$

$abeAM$

$abaeM$

8. **Automata and language theory**

Consider the following grammar:

$G \longrightarrow S\ \$\$$
$S \longrightarrow A\ M$
$M \longrightarrow S \mid \epsilon$
$A \longrightarrow a\ E \mid b\ A\ A$
$E \longrightarrow a\ B \mid b\ A \mid \epsilon$
$B \longrightarrow b\ E \mid a\ B\ B$

(a) Describe the language that the grammar generates in English.

(b) Show a parse tree for the string a b a a.

(c) Is the grammar LL(1)? If so, show the parse table; if not, identify a prediction conflict.

(a) in the ~~[illegible]~~ grammar, steve from S, S, G, M, A, E, B are the nonterminals. a b are the terminals, ~~[illegible]~~ and there has productive Rules for ~~those~~ G, S, M, A, E, B

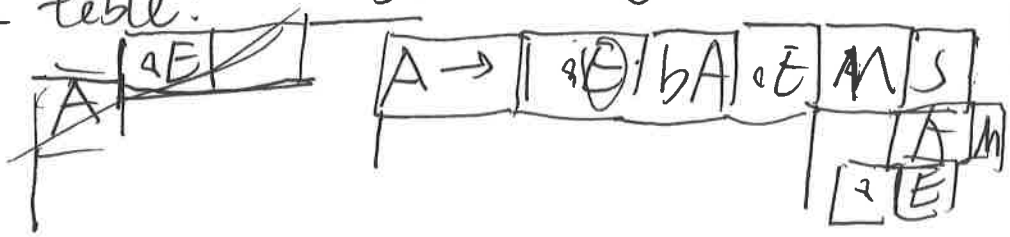
(b)

(c) Yes

$S \to AM$
$S \to aEM$
$S \to abAM$
$S \to abaEM$
$S \to ababM$
$S \to abaS$
$S \to abaAM$
$S \to abaaEM$
$S \to abaaM$
$S \to ababaa\varepsilon$

each step is left e-right and I store

parse table:



9. **III.C Discrete Structures**

Recall that the Hamiltonian Cycle Problem is the problem of deciding, on input graph G, whether G has a cycle that visits all the nodes exactly once. Show that this problem is polynomial time decidable if the input is restricted to the graphs with the property that each node has at most two neighbors (i.e., at most two adjacent nodes).

$G(V, E)$ V is the node number,
is each node connect 2 nodes.
$E = 2V$,

Start from a radom node, visit each nodes once, take $O(V+E) = O(3V)$.

from a node to do depth first search to find next unvisited node, after all the nodes been visited, by DFS it will take $O(V^2)$.

So the whole time will take $O(V^2) + O(3V)$
$= O(V^2)$

$\therefore O(V^2)$ is polynomial

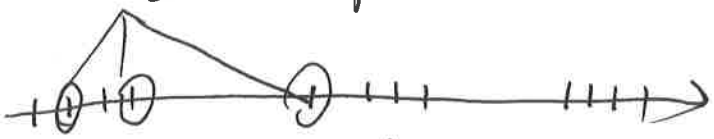
10. **IV Other Topics**

Give a detailed explanation of any one approach to machine learning. Give a substantial example that illustrates the technical operation of the approach, and demonstrates interesting knowledge learned.

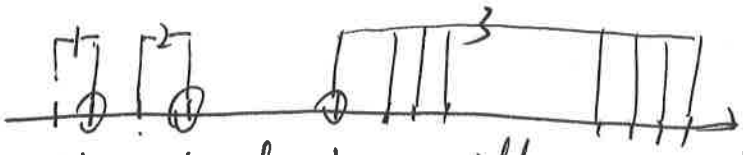
K-mean clustering.

if there has 12 nodes in three clusters in a line like

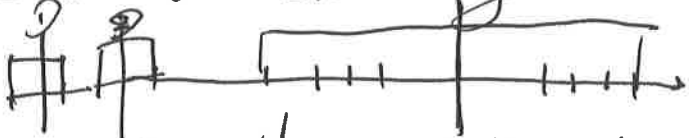
1. Randomly pickup three nodes as 3 clusters



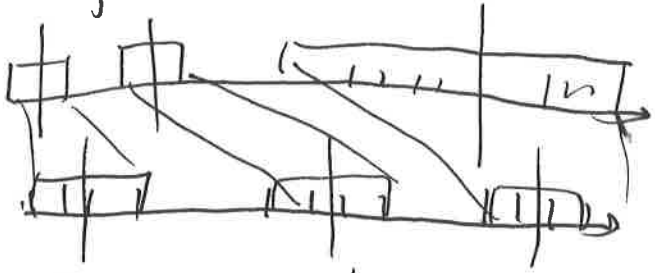
2. find the ~~dth~~ distance between other points to three points cluster the ~~par~~ nodes ~~with~~ to three clusters with the distance between the closest cluster



3. then find the ~~mid~~ ~~mean~~ mid points of 3 clusters, then re-do the clustering.



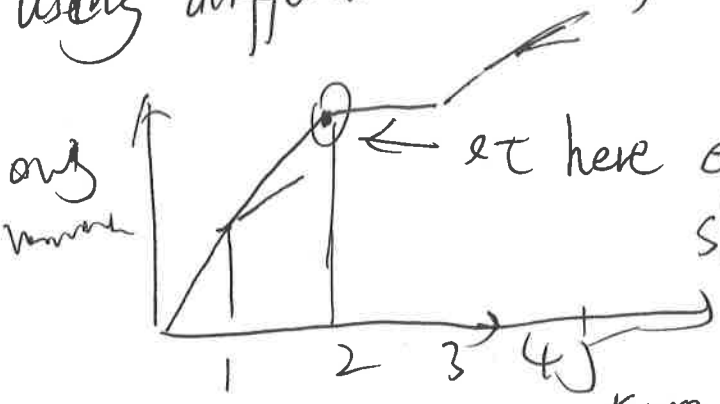
4. Compare the variation between the results with actual one



5. To get a better results, need to do more k-means clustering back

6.. usually, k's value is unkown, need to do testing with different k values to find a bese k.

when using different k-value, the results could be as



at here the curve slowly increase, we usually pick this as the value.

K-mean is a easy, usefull algorithm, some times combine it with others like Spectrum to display to results.