Ph.D. Comprehensive Examination

Computer Science Department
University of Miami

January 14, 2022

Student Name: Jiasong Chen Student Number:

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

1. **Data organization; 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, Search $O(\log n)$ Insert $O(n)$ Successor $O(1)$ Predecessor $O(1)$
   (b) Ordered doubly linked list, $O(n)$, $O(n)$, $O(1)$, $O(1)$
   (c) Min-Heap, and $O(n)$, $O(\log n)$, $O(n)$, $O(n)$
   (d) Hash table $O(1)$, $O(1)$, $O(n)$, $O(n)$

2. **Program control and structure; 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

Call by value, the input of function swap(i, A[i]) is a copy of the actual parameters. The actual parameter i, A[i] doesn't change even if the formal parameter changes in the function. The actual parameter and formal parameter are stored at different memory location.

Call by reference, the input of function swap(i, A[i]) is the address of actual parameters. The actual parameters and formal parameters are stored at the same address, so the change of the formal parameter will result the change of the actual parameters.

Call by value result, the actual parameter don't change after the new assignment of value.

3. **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: it is helping to find the target customer to meet their requirements. for example, the NABC analysis, the requirement analysis helps find the core competitions. forexample, if we are developping a cat recognization project. The customer might be the people who loves cats. The requirements of those customer might be good looking user interface, if the customers are computer scientist, the requirement might be the recognition algorithm robustness. The requirement analysis could be divided into different level such as UI level, prior (core competitive) requirements.

(c) Architectual design: It is about the framework design according to the requirements, for example, the design of the website, the choice of architecture, front end and back end interactions. the choice of algorithms.

(d) Implementation: Based on the documentation, for example, the differences evaluation between waterfall and agile development, Code management Github merge.

4. **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. **Software, Programming Techniques**

Given that

$B(x)$ means "x is a bear"
$F(x)$ means "x is a fish", and
$E(x, y)$ means "x eats y",

what is the best English translation of
$\forall x[F(x) \rightarrow \forall y(E(y, x) \rightarrow B(y))]$?

(a) All fish eat bears.

(b) Every fish is eaten by some bear.

(c) Bears only eat fish.

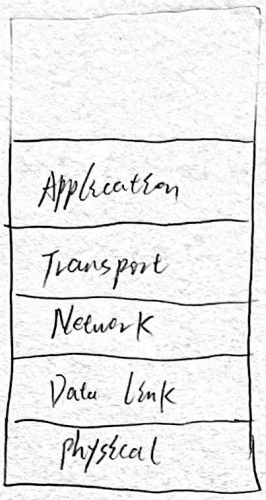
(d) Every bear eats fish.

(e) Only bears eat fish.

6. **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).



physical layer: the unit is bits, it transmits the raw bits stream through physical medium.

Data Link: the unit is frame. the raw bits data from physical layer is organized as frame and deliver to next layer.

Network: the unit is packet, this layer deliver packet to next layer.

Transport: the unit is segment. this layer transmits data using transmission protocol such as TCP, UDP.

Application: this layer handles the human-computer interface and applications that connect internet services.

From bottom to up, the physical layer transmits the raw bits through physical medium like cables. Then, the data link layer performs framing, physical address, error detection. The raw bit data is organized with the frame header and trailer along with these information like MAC address. After that, the network layer receives frame and generate packets. The packet is generated with the frame data and a header including the source and destination information like IP address. Then, FTP uses two TCP connections during transport layer. One is data connection on server port 21, which to do data transfer. The other one is control connection on server port 20, which to verify user identification and remote directory. For the application layer, ↘

the FTP protocol acts as an interface that transfer data from different file systems

7. **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. Sort list A with Quick sort $O(n \log n)$

2. Sort list B with Quick Sort $O(m \log m)$

3. Linear Scan A and B lists with following steps: $O(n+m)$

let list C be the result, i is the index of list A, j is the index of list B.

① if $A[i] < B[j]$, then add $A[i]$ into list C and move i to next element, that is $i = i+1$.

② if $A[i] = B[j]$, then move both i and j to next element, that is $i = i+1$, $j = j+1$

③ if $A[i] > B[j]$, then move j to next element until $A[i] \leq B[j+n]$,

If A is iterated through, then the list C is the result. If B is iterated through and A is not over, then move all the remaining elements in A into list C, C is the result. the total complexity of above algorithm should run in $O(n \log n + m \log m)$

8. **Automata and language theory**

Consider the following grammar:

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

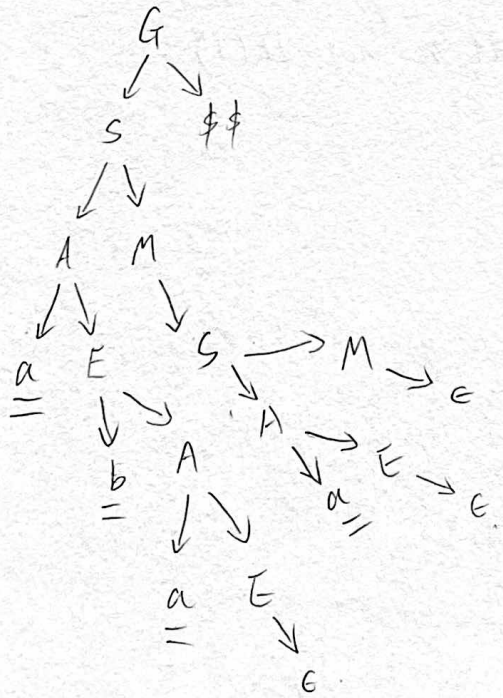
(a) Describe the language that the grammar generates in English.

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

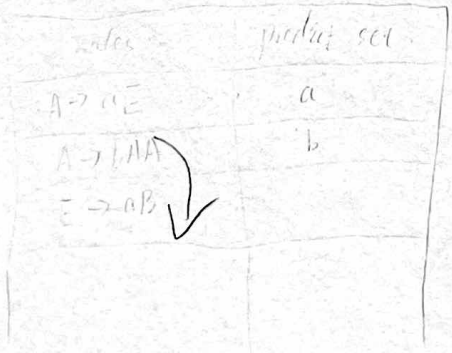
(c) Is the grammar LL(1)? If so, show the parse table; if not, identify a prediction conflict.

(a) The grammer only generates a string consist of "a" and "b". The number of "a" is greater than the number of "b"

(b)



(c)



(c)

	Rules	predict set
1	$A \to aE$	a
2	$A \to bAA$	b
3	$E \to aB$	a
4	$E \to bA$	b
5	$E \to \epsilon$	a, b, $\$$
6	$B \to bE$	b
7	$B \to aBB$	a

Review the 4th and 5th rules in above table, the predict set of $E \to bA$ is b, the predict set of $E \to \epsilon$ is a, b, $\$$, the intersection of these two is b, not empty set, so it is not LL(1).

9. **Discrete Structures**

A **permutation** of a set A is an *ordered* arrangement of the elements in A. An ordered arrangement of just r elements from A is called an r-permutation of A. For non-negative integers $r \leq n, P(n,r)$ denotes the number of r-permutations of a set with n elements.

What is $P(n,r)$? This is about counting.

$\binom{n}{r}$

the number of r-permutations of a set with n elements

$\binom{n}{r}$

10. **Other Topics**

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

K-mean algorithm.

1. Specify the number of clusters K.

2. Randomly Initialize K clusters' centers.

3. Repeat following steps until no changes to the cluster centers.

   ① Calculate the distance between all data points to each cluster center.

   ② Assign each data point to the nearest cluster

   ③ Calculate the new clusters' centers by averaging the data points in each cluster.

For the choice of cluster number K, we could try different values for K and make an elbow plot to choose the best choice of K.