Indian Institute of Information Technology, Lucknow

Mid-semester Examination Feb 2018

Compiler Design (ICOD632C)

B.Tech (IT)-6th Sem

QPS: SC

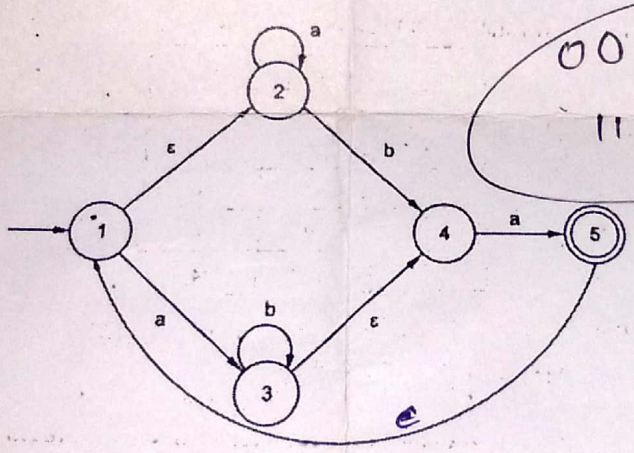
Max Time: 2 Hour **Max Marks: 30**

Note: This question paper contains ***SEVEN*** questions. Attempt ***ALL*** questions. Answers to all the questions should be written in the same order as of the respective questions. Marks are indicated against each question. Assume the suitable data and mention on answer sheet clearly, if found missing.

Q.1. a) Write a regular expression to denote a language L over Σ^*, where $\Sigma = \{a, b\}$, such that in each string the 3rd character from right end of the string a always *a*.

b) Write a regular expression to denote a language L over Σ^*, where $\Sigma = \{0,1\}$, such that each string in the language begin and end with 00 or 11. [2+2]

Q.2. Consider following NFA with ϵ moves. Find the equivalent DFA.



[2+4]

Q.3. *Show that for every regular expression there exist an NFA.* [3]

Q.4. *Consider the following program in LEX*

```
%%
^[A-Z]+ {ECHO; printf("\n");}
\n      ;
.       ;
%%
yywrap(){}
main(){yylex();}
```

For the following input, what will be the output of the given LEX program?

INPUT: ertSST *INPUT: SMTert* *INPUT: MertMS* *INPUT: srtMStr* *INPUT: Sdtr* [3]

Q.5. Write an algorithm that translates the transition diagrams for IDENTIFIER and INTEGER CONSTANTS (Decimal, Hex). The IDENTIFIER should start with a letter (capital or small alphabets) or an underscore followed by any number of letters or digits or underscore. IDENTIFIER also allows $ and # between the start and end symbols. Decimal constants should start with decimal digits and end with *u* or *U* or ϵ. Hex constants start with either *0x* or *0X* and end with *u* or *U* or ϵ. Hex constants contain one or more hex digits in between the start and end symbols. **[3+2+3]**

Q.6. What is context sensitive operator in LEX programming? Explain their uses. **[3]**

Q.7. Write a program in LEX that prints only vowels from the given input to a file named *out* and ignores all other letters, digits, whitespaces, and newline characters. **[3]**

Indian Institute of Information Technology, Lucknow

End-semester Examination April-May 2018

Compiler Design (ICOD632C)

B.Tech (IT)-6th Sem

QPS: Dr. Soumendu Chakraborty

Max Time: 3 Hour **Max Marks: 70**

Note: This question paper contains ***SEVEN*** questions. Attempt ***ALL*** questions. Answers to all the questions should be written in the same order as of the respective questions. Marks are indicated against each question. Assume the suitable data and mention on answer sheet clearly, if found missing.

Q.1. a) Explain with example the role of input buffering in lexical analyzer. Write an algorithm to compute ∈-closure of a set of states Q of an NFA. **[4+4]**

b) What is an ambiguous grammar? How the ambiguity can be removed from a given grammar? Explain with example. **[3+4]**

Q.2. Test whether the following grammar is LL(1) or not before parsing. Construct the predictive parsing table for the grammar using FIRST and FOLLOW.

$$S \rightarrow AaAb|BbBa, A \rightarrow \in, B \rightarrow \in$$ **[3+7]**

Q.3. Define the operator precedence relations. Write the operator precedence parsing algorithm. Perform precedence parsing on the following expression showing the parse table stack content and current input string.

$$id \div id \times (id + id)$$ **[2+3+4]**

Q.4. Consider the following grammar, and construct the LR(1) parsing table

$$S \rightarrow aSbS|bSaS| \in$$ **[12]**

Q.5. a) Construct the DAG of the following basic block.

$$a = b + c, b = b - d, c = c + d, e = b + c$$

b) Consider the following expression

$$w = (a - b) + (a - c) + (a - c)$$

(i) Translate it into three address code

(ii) Apply the code generation algorithm and show the code sequence.

[2+2+4]

Q.6. What is backpatching? Explain with example. **[4]**

Q.7. What is PEEPHOLE optimization? Elaborate different operations performed in PEEPHOLE optimization. **[12]**

Student Name: ..... Roll Number: .....

Indian Institute of Information Technology, Lucknow

Mid Semester Exam (IDMW632C)

Maximum Marks: 30 **Maximum Time:** 2 Hours

Date: 25th Feb, 2018 **Instructor:** Dr. Vishal Krishna Singh

Note: Please read the paper very carefully. All questions are compulsory.

1. Briefly discuss the following: (Marks: $5 \times 1 = 5$)

   (a) Data Characterization

   (b) Data Discrimination

   (c) Ordinal Attribute

   (d) Data Pre-processing

   (e) Numerosity Reduction

2. (a) Compare and contrast FP-Growth algorithm with Apriori algorithm. (Marks: 2)

   (b) Consider the market basket transactions given in the following table. Let $min_{sup} = 40\%$ and $min_{conf} = 40\%$. Find all the frequent item sets using Apriori algorithm. (Marks: 3)

Transaction Set	
Transaction ID	Items Bought
T_1	ABC
T_2	ABCDE
T_3	ABC
T_4	ACDE
T_5	ABCD

3. A business owner wants to improve employee relations in his company. He predicted that he met his goal of increasing employee satisfaction from 65% to 80%. Employees from four departments were asked if they were satisfied with the working conditions of the company. The results are shown in the following table: (Marks: 5)

1/3

P.T.O

IIIT - Lucknow

Roll Number: .....

Survey Data

	Finance	Sales	Human Re-source	Technology
Satisfied	12	38	5	8
Dis-satisfied	7	19	3	1
Total	19	57	8	9

- *Using the data provided above, determine whether the results support or reject the business owner's prediction. Consider the p-value to be 0.05. The following chi square table is provided for reference.*

DF	P-value								
	0.995	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	0	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750

4. In a TV Game show, a contestant selects one of three doors; behind one of the doors there is a prize, and behind the other two there are no prizes. After the contestant selects a door, the game-show host opens one of the remaining doors, and reveals that there is no prize behind it. The host then asks the contestant whether they want to SWITCH their choice to the other unopened door, or STICK to their original choice. **(Marks: 5)**

   (a) Is it probabilistically advantageous for the contestant to SWITCH doors, or is the probability of winning the prize the same whether they STICK or SWITCH? Justify.

   Note: *Assume that the host selects a door to open, from those available, with equal probability.*

5. Answer the following: **(Marks: $2 \times 5 = 10$)**

   (a) Discuss Information Gain and Gini Index for attribute selection in decision trees.

   (b) Considering the training example given below, calculate the gain in the Gini index when splitting on ~~A and B.~~ Which attribute would the decision tree algorithm choose?

a, b and c

2/3

IIIT - Lucknow

Roll Number: .....

Instance	a_1	a_2	a_3	Target Class
1	T	T	1.0	+
2	T	T	6.0	+
3	T	F	5.0	−
4	F	F	4.0	+
5	F	T	7.0	−
6	F	T	3.0	−
7	F	F	8.0	−
8	T	F	7.0	+
9	F	T	5.0	−

3/3

Indian Institute of Information Technology, Lucknow

End Semester Exam (IDMW632C)

Maximum Marks: 75 **Maximum Time: 3 Hours**

Date: 2nd May, 2018 **Instructor: Dr. Vishal Krishna Singh**

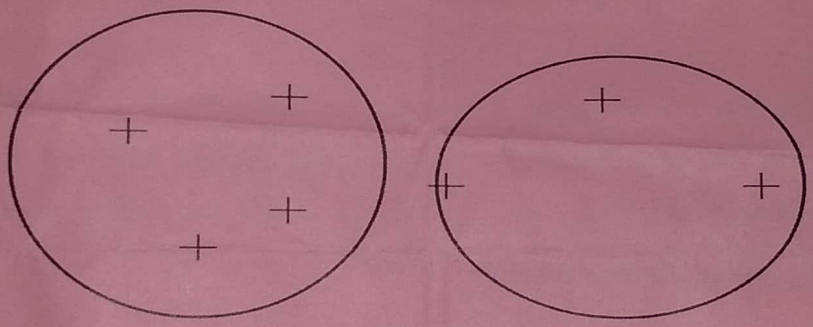
Note: Please read the paper very carefully. All questions are compulsory.

1. Answer the following: **(Marks: 2+2+2+4 = 10)**

   (a) We generally will be more interested in association rules with high confidence. However, often we will not be interested in association rules that have a confidence of 100%. Why ? Then specifically explain why association rules with 99% confidence may be interesting (i.e., what might they indicate)?

   (b) Discuss the basic difference between the agglomerative and divisive hierarchical clustering algorithms and mention which type of hierarchical clustering algorithm is more commonly used.

   (c) Are the two clusters shown below well separated ? Justify your answer.



   (d) State True or False. Give suitable reason for your answer.

   i. A density-based clustering algorithm can generate non-globular clusters.

   ii. In association rule mining the generation of the frequent itemsets is the computational intensive step.

2. Briefly discuss the following: **(Marks: 1 + 1 + 3 = 5)**

   (a) *Define* a core point in DBSCAN.

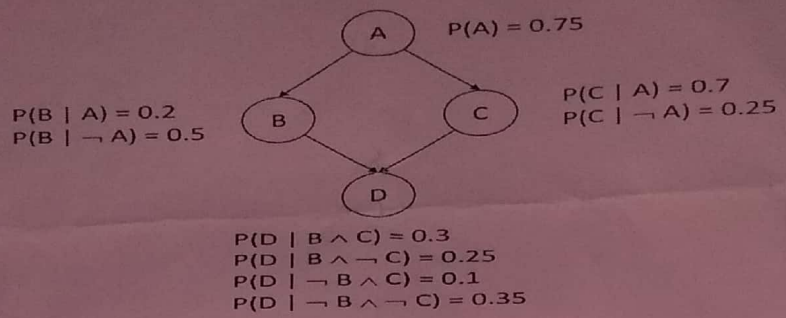
   (b) *Define* a centroid in k-means.

   (c) *The following* is a set of one-dimensional points:

   $[1, 1, 2, 3, 5, 8, 13, 21, 33, 54]$

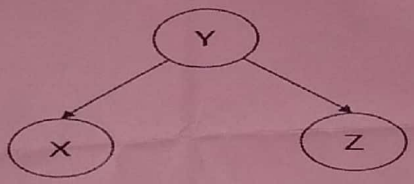
   Perform two iterations of k-means on these points using the two initial centroids 0 and 11

3. (a) Consider the following Bayesian network. A, B, C, and D are Boolean random variables. If we know that A is true, what is the probability of D being true? **(Marks: 7.5)**



(b) For the following Bayesian network, we know that X and Z are not guaranteed to be independent if the value of Y is unknown. This means that, depending on the probabilities, X and Z can be independent or dependent if the value of Y is unknown.

Construct probabilities where X and Z are independent if the value of Y is unknown, and show that they are indeed independent. **(Marks: 7.5)**



4. Give a detailed description of the following: **(Marks: $5 \times 3 = 15$)**

(a) ROLAP vs MOLAP

(b) HOLAP

(c) Multiway Array Aggregation Method for Computing Full Cubes.

(d) BUC Method For Cube Computation

(e) Star-Cubing method

5. (a) The typical unsupervised approach to detect anomalies in a dataset receives as input an anomaly scoring function $f(x)$ and a threshold t, and it returns all the data instances in the dataset with anomaly score greater than t. **(Marks: $2.5+2.5 = 5$)**

i. Given a scoring function $f(x)$, describe a method to determine a good value for the threshold t.

ii. Provide an example of a specific scoring function $f(x)$, and state if it can be used for statistical-based, proximity-based, clustering-based, or density-based anomaly detection.

(b) Assume that the dataset contains a Boolean attribute '*isAnomaly?*' with value '*yes*' if the data instance is an anomaly and '*no*' if it is not. **(Marks: 5+5 = 10)**

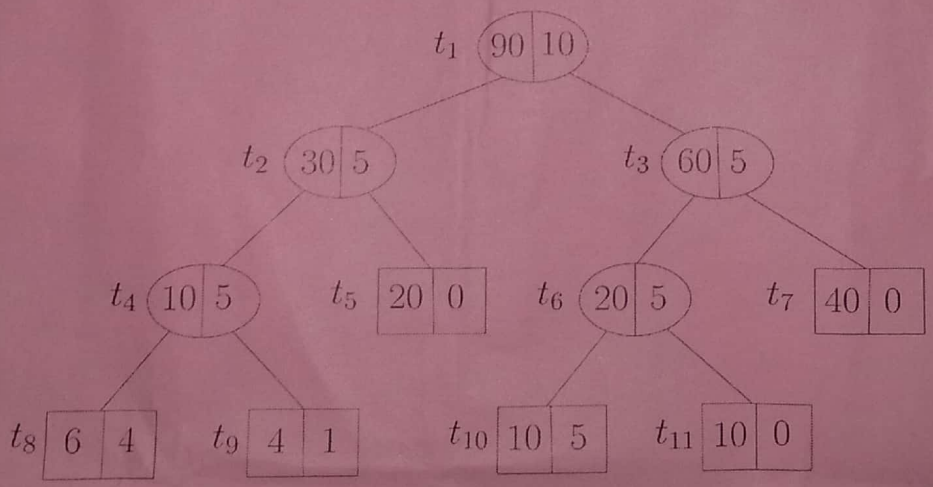
i. Describe how you would determine if a given new data instance (for which you dont know its *isAnomaly?* value) is an anomaly or not. Explain your answer.

ii. One expects anomalies to be relatively rare (i.e., infrequent) with respect to normal instances in a dataset. How does this affect the construction and/or the evaluation of the supervised method you proposed above? Explain.

6. (a) When the back propagation algorithm is used for classification, the objective function to be minimized is usually the same as in function approximation: *a sum of the squared error between the output and the target.* But in classification, the real objective (usually) is to minimize the number of misclassifications.

Why can this not be used directly, as the objective function? **(Marks: 5)**

(b) The tree given below, denoted by T_{max}, has been constructed on a training sample. In each node, the number of observations with class 0 is given in the left part, and the number of observations with class 1 in the right part. The leaf nodes have been drawn as rectangles. **(Marks: 10)**



i. Compute the impurity of nodes t_1, t_2 and t_3 using the gini-index.

ii. Give the impurity reduction achieved by the first split.

iii. Compute T_1, the smallest minimizing subtree of T_{max} for $\alpha = 0$

Indian Institute of Information Technology, Lucknow

Mid-semester Examination Feb 2018

Image and Video Processing (IIVP632C)

B.Tech (IT)-6th Sem

QPS: SC

Max Time: 2 Hour **Max Marks: 30**

Note: This question paper contains ***SIX*** questions. Attempt ***ALL*** questions. Answers to all the questions should be written in the same order as of the respective questions. Marks are indicated against each question. Assume the suitable data and mention on answer sheet clearly, if found missing.

Q.1. A common measure of transmission for a digital data is the Baud rate, Defined as the number of bits transmitted per second. Generally, transmission is accomplished in packets consisting of a start bit, a byte (8bits) of information, and a stop bit.

a) How many minutes it would take to transmit a 1024X1024 image with 256 gray levels, using a 56K baud modem?

b) Repeat for a phone digital subscriber line (DSL) at 750K baud? **[2+1]**

Q.2. a) If V = {2,3,4}, compute the 4 and 8 connected components of the pixels (p), (q), (r) in the sub-image shown in Fig.1(a). Show the components by giving unique names to unnamed pixels.

b) If V={2}, show 4 and 8 connected regions of the sub-image shown in Fig.1(a).

c) For V={1} and V={2}, Determine whether S2 in Fig.2(b) is 4 and 8 connected regions or not. Also find out whether S1 and S2 are 4and 8 adjacent or not

0	1	0	1
0	0	2(p)	1
0	3(q)	2	0
4(r)	2	0	2
0	3	0	0

(a)

S1				S2			
0	1	0	1	1	1	0	1
0	0	2(p)	1	0	1	0	1
0	3(q)	2	0	2	2	2	0
4(r)	2	0	2	4	2	0	2
0	3	0	0	0	2	0	0

(b)

Fig.1.

[3+2+3]

Q.3. Explain bilinear and bi-cubic interpolation with example. Show that summation is a linear transformation. **[2+2+2]**

Q.4. Define n^{th} moment of a random variable. The values of 0^{th} and 1^{st} moments are 1 and 0 respectively, Justify. What do the positive and negative 3^{rd} moments signify? **[1+3+2]**

Q.5. Show that the product of cumulative distribution of an input image with maximum intensity level transform the image into an image with uniform probability density. **[4]**

Q.6. Derive the following Laplacian kernel shown in Fig.2. from the 3×3 neighborhood of a pixel at location (x,y). **[3]**

-1	0	-1
0	4	0
-1	0	-1

Fig.2.

Indian Institute of Information Technology, Lucknow

End-semester Examination April-May 2018

Image and Video Processing (IIVP632C)

B.Tech (IT)-6th Sem

QPS: Dr. Soumendu Chakraborty

Max Time: 3 Hour **Max Marks: 70**

Note: This question paper contains ***FIVE*** questions. Attempt ***ALL*** questions. Answers to all the questions should be written in the same order as of the respective questions. Marks are indicated against each question. Assume the suitable data and mention on answer sheet clearly, if found missing.

Q.1. a) Consider two 8-bit images whose intensity levels span the full range from 0 to 255.

(i) Discuss the limiting effect of repeatedly subtracting image (2) from image (1). Assume that the result is represented also in eight bits.

(ii) Would reversing the order of the images yield a different result?

b) Show that the Laplacian operator is rotation invariant for

$$u = x\cos\theta + y\sin\theta, v = -x\sin\theta + y\cos\theta$$

c) Discuss the limiting effect of repeatedly applying a 3×3 low-pass spatial filter to a digital image. You may ignore border effects. **[3+2+3+4]**

Q.2. a) Show that

$$f(t) = h(t) \star [\sum_{n=-\infty}^{\infty} f(t)\delta(t - n\Delta T)] = \sum_{n=-\infty}^{\infty} f(n\Delta T) sinc(\frac{t - n\Delta T}{n\Delta T})$$

Where the Fourier Transform of $h(t)$ is defined as

$$\Im(h(t)) = \begin{cases} \Delta T & -\mu_{max} \leq \mu \leq \mu_{max} \\ 0 & otherwise \end{cases}$$

b) Show that aliasing is inevitable in continuous signal. Mathematically show that limiting an infinite continuous signal to a finite interval introduces infinite frequency components.

[8+3+4]

Q.3. a) Show that $F(u) = F(u + kM) and\ f(x) = f(x + kM)$
Where k is an integer and M is the period.

b) Find the equivalent filter *H(u,v)*, that implements the following filter in frequency domain

-1	0	-1
0	4	0
-1	0	-1

c) What is a Finite Impulse Response filter? Explain with example. **[5+7+3]**

Q.4. a) How the filtered image is computed using Homomorphic Filtering? Show with proper Derivation.

b) Show that the variance σ^2 of Laplacian of Gaussian (LOG) can be defined as follows

$$\sigma^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 - \sigma_2^2} \ln\left[\frac{\sigma_1^2}{\sigma_2^2}\right]$$

Where σ_1^2, σ_2^2 are the variances of Difference of Gaussians. **[7+8]**

Q.5. a) Compute global optimum threshold using Otsu's method.

b) Prove the validity of the duality expression $(A \cdot B)^c = (A^c \circ \hat{B})$ and $(A \circ B)^c = (A^c \cdot \hat{B})$. **[8+5]**