

## Qualify Exam - Data Mining

Notice: Close book. 總分: 100 分

(共 2 頁)

1. (15%) Answer the following questions:
  - A. Give the equation of *Lift* and explain how it can be used to measure the interestingness of association rules.
  - B. Explain what is "Constrained Clustering"? Also give a brief example.
  - C. Describe how "*Ensemble*" method works for building a classifier and explain why it can reach higher accuracy normally.
2. (25%) Answer the following questions on clustering methods:
  - A. "Partition-based clustering", "Hierarchical clustering" and "Density-based clustering" are among the most popular types of clustering methods. Explain briefly how the three types of clustering methods work by using one well-known clustering method as example for each type, respectively.
  - B. Give comparisons on advantages and drawbacks among the three types of clustering methods.
  - C. Explain how to validate the quality of clustering results. Also, give an efficient way to store the clustering results.
3. (30%) Answer the following questions about classification modeling:
  - A. Compare Decision Tree, SVM (Support Vector Machine) and CBA (Classification Based on Association) in terms of accuracy, efficiency and interpretability.
  - B. Explain what is "overfitting" problem in classification modeling and how to deal with it.
  - C. Given a dataset with highly imbalanced class distribution. Describe how to deal with the datasets for improving the classification accuracy.
4. (30%) Answer the following questions:
  - A. Given a database  $D$  consisting of a set of transactions  $T_i: \{t_i, C_{id}, (I_a, I_b, \dots, I_n)\}$ , where  $t_i$  is the purchase time of transaction  $T_i$ ,  $C_{id}$  is the customer id and  $(I_a, I_b, \dots, I_n)$  are the items contained in this transaction. Defining *Sequential Pattern Mining* as "Finding all maximal sequences that meet the user-specified minimum support  $S_{min}$ ", please describe the main processes involved in conducting the *Sequential Pattern Mining*.
  - B. AprioriAll and AprioriSome are the well-known methods for mining sequential

patterns. Describe how AprioriAll and AprioriSome work, respectively, and point out the main differences between them briefly.

- C. Compare AprioriAll and AprioriSome in terms of execution efficiency and memory requirements (also explain why).
- D. Describe an approach for mining sequential patterns without generating candidates.

## DBMS Qualify 2014 Spring

1. Answer the following queries in SQL using the following given schema.

Teacher(T#, TName)

Student(S#, SName)

Course(C#, CName)

Taking(S#, T#, C#, Score)

- (a) Give the student names for those who do not take any courses. (5%)
- (b) Give the student names for those who take at least one course. (5%)
- (c) Give the teacher names for those who teach exactly two courses. (5%)
- (d) Give the student names for those who take all the courses taught by teacher "Peter Chang". (5%)
- (e) Give the average score and the course name for each course taught by teacher "Peter Chang". (10%)

2. (20%) Answer the following queries in relational algebra using the above schema.

- (a) Give the name of every student whose average score of all the courses he/she takes is higher than the average score of all the courses that teacher "Peter Chang" teaches.
- (b) For each student, list the student name, the teacher name, and the average score of the courses that the student takes under that teacher. For example, student A take two courses from teacher M and three courses from teacher N. Then, the result should be like this:

Student	teacher	average score
A	M	75
	N	80

3. (10%) Given a relation schema and their functional dependencies as follows.

R(A, B, C, D, E, F)

(A, B  $\rightarrow$  C), (A, B, C  $\rightarrow$  D), (B  $\rightarrow$  E), (F  $\subseteq$  D)

- (a) Normalize the relation to 2NF.
- (b) Normalize the relation to 3NF.

背面有題

4. (16%) Give two advantages and two disadvantages for the two-phase locking and the time-stamping protocols.
  
5. (24%) Answer the following questions/terms.
  - (a) Referential integrity constraint.
  - (b) Explain the concept of cloud computing.
  - (c) Explain the “map/reduce” concept used in current cloud computing systems.
  - (d) Give two research issues (and briefly explain the issues) that you can think of in cloud computing systems.
  - (e) Give two research issues (and brief explanations) that you can think of in social networks.
  - (f) In research, usually how do we compute the “similarity” between two terms or concepts?

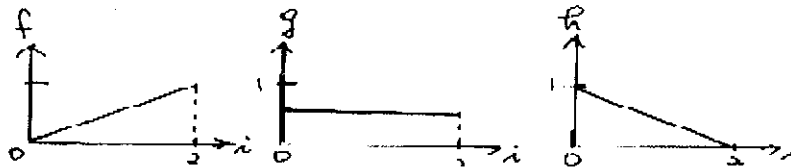
## OS 資格考題 (102 年度第 2 學期)

1. [20%] Consider a memory system in which the memory consists of the following holes in order: 11KB, 4KB, 20KB, 18KB, 10KB, 7KB, 9KB, 12KB and 15KB. Which holes are taken for a sequence of segment requests of 12 KB, 18KB, and 9KB, under the following memory allocation policies?
  - a. First-fit
  - b. Best-fit
  - c. Worst-fit
2. [15%] A computer with a 32-bit address uses a two-level page table. Virtual addresses are split into a 10-bit top-level page table field, a 10-bit second-level page table field, and a page offset.
  - (a) [5%] What's the page size?
  - (b) [10%] If a program has its code and data together fitted in the lowest 20KB and stack fitted in the highest 8K of its virtual address space, how many pages are needed for storing the page table for this program?
3. [15%] Please describe the reason and a possible solution for *thrashing*.
4. [15%] Please describe the advantages and disadvantages of a pure demand paging system and a pre-paging system.
5. [20%] Please write pseudo code to solve the Dining Philosophers problem by using the *monitor* mechanism.
6. [15%] Please describe the tradeoff between using processor affinity and load balancing mechanisms.

**Digital Image Processing,  
PhD Qualification Exam,  
Department of CSIE**

April 2014

1. If a medical image is given and you are asked to detect (or segment) an organ area from the image, you then have to develop a program to solve this problem. Suppose the given image is a sonogram and you have to do three main image processing steps which are image preprocessing, feature extraction, and segmentation, please explain how you want to accomplish these three steps and why you select these approaches. (20%)
2. We have an image whose histogram is  $f(i)$ . Please do histogram equalization to make the histogram transferred to  $g(i)$ , (10%) then do histogram modification to make the histogram transferred to  $h(i)$ . (10%)



3. If you want to remove noises from a given image, you can usually do filtering in the spatial or frequency domains. Please describe the procedures of filtering in spatial and frequency domains. (10%) Are there differences between the two filtering methods? (5%) How can you evaluate the filtering results in signal noise ratio (give equation)? (5%)
4. Please describe what the Hough transform is. (5%) Please explain why the  $(r, \theta)$  domain is used instead of the  $(a, b)$  domain in Hough transform. (5%) Please describe how to use Hough transform to detect coins with three different diameters from an image. (10%)
5. If a Gaussian filter is  $G(x, y) = e^{-\frac{x^2+y^2}{2\sigma^2}}$ . If Laplacian is the sum of second derivatives of  $x$  and  $y$  variable, please show Laplacian of Gaussian (LOG) filter is

$$h(x, y) = \frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} e^{-\frac{x^2+y^2}{2\sigma^2}}. \quad (10\%)$$

If the difference of Gaussian is used to approximate the LOG with  $\sigma_1 > \sigma_2$ , please show the two filters will have the same

zero crossing if  $\sigma^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 - \sigma_2^2} \ln \left[ \frac{\sigma_1^2}{\sigma_2^2} \right]. \quad (10\%)$

1. Explain the following terms in detail: (60%)  
(a) power dissipation (power consumption) (b) setup time  
(c) clock skew problem (d) functional simulation  
(e) hold time (f) soft IP
2. Suppose you have completed a circuit design with hardware description language. Now, you must decide to implement your circuit whether with (a) ASIC or (b) CPLD/FPGA. Which way is better? Explain the advantages and disadvantages of the two ways in detail. (20%)
3. Describe the difference between full custom and Cell-based design flow. (20%)

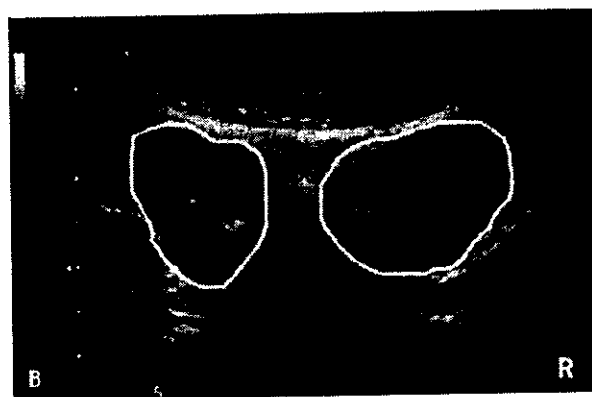
國立成功大學資訊工程系一〇二學年度第二學期博士班資格考命題紙

科目名稱：生醫超音波訊號與成像技術

考試時間：一〇三年四月十八日

本試題共 1 張，本張為第 1 張

1. Ultrasound image is frequently applied in clinical diagnosis.
  - (a). Plot a block diagram for a typical ultrasound B-mode imaging system, and then explain functions of each component in the system (15%); (b). The resolution of ultrasound image is determined by the resolution cell of the transmitted wave. Define the resolution cell. (5%); (c). Speckle sometimes is considered to be a noise in ultrasound image. What is the origin of speckle? (5%) Provide imaging processing techniques to remove speckle in a 2D B-mode ultrasound image. (5%)
2. For the following original ultrasound image (left), (a). provide detail image processing scheme to develop a software program for automatically detecting the regions of interest as the white lines in the right image. Please include your considerations, such as the feature of the original image, preprocessing, filtering, edge detection, contouring... (25%); (b). The image of those deeper regions is either darker or ambiguous. Describe possible reasons leading to this effect. (5%) Provide method to enhance image of those regions. (5%)



3. (a). Plot a diagram about the internal structure of a single element transducer. Describe functions of each part of a transducer. (10%) (b). A 20 MHz transducer was made by a round shape of PZT piezoelectric disk with its diameter of 4 mm. Calculate the transition point between the near-field and far-field of the acoustic field when the transducer was immersed in a water tank. Plot ideal axial and lateral profiles for this transducer. (15%) (c). Describe feasible methods or modalities to measure acoustic field of an ultrasound transducer. (10%)



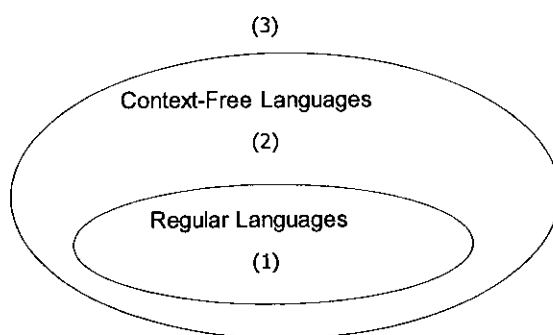
Ph.D. Qualify Examination 2014  
Theory of Computation

- This examination is closed books.
- Please turn off your cell phones.
- Remember that there are 2 pages of the qualify examination.
- Answer all questions as possible. You may have a partial score if you answer the correct direction.

1. Deterministic Finite Acceptors (DFAs) (10 pts)  
For  $\Sigma = \{0, 1\}$ , construct DFA that accepts the set consisting of:  
All strings whose binary interpretation is divisible by 5.
2. Nondeterministic Finite Acceptors (NFAs) (20pts. 10 pts each)  
Draw NFA to accept the following sets of strings over  $\{a, b\}$ :  
(a) Find an NFA with three states that accepts the language  
 $L = \{a^n : n \geq 1\} \cup \{b^m a^k : m \geq 0, k \geq 0\}$ .  
  
(b) All strings where every odd position is "a".
3. Please use the pumping lemma to show that the language is nonregular: (15 pts)  
 $L = \{a^n b^l : n/l \text{ is an integer}\}$ .
4. Suppose  $r_1$  and  $r_2$  are regular expressions, and  $L(r_1)$  and  $L(r_2)$  are regular languages, correspondingly. Prove that  $L(r_1 + r_2)$  is also a regular language (15 pts).  
hint: Note that  $L(r_1 + r_2) = L(r_1) \cup L(r_2)$ . You can consider the solution by using NFA.
5. Show that the following grammar is ambiguous. (10 pts)  
 $S \rightarrow aSbS \mid bSaS \mid \lambda$ .
6. Construct a nondeterministic pushdown automata that accept the following language over  $\Sigma = \{a, b, c\}$ : (10 pts)  
 $L = \{a^n b^m c^{n+m} : n \geq 0, m \geq 0\}$ .

7. Fill the following languages into the language hierarchy (If  $L_i$  is a regular language and also a context-free language, please fill  $L_i$  in the set of regular languages): (20 pts)

- $L_1 = \{a^n b^m : n \geq m\},$
- $L_2 = \{a^n b^j c^k : k = jn\},$
- $L_3 = L(a^* b^* c^*),$
- $L_4 = \{a^n b^n c^n : n \geq 0\},$
- $L_5 = \{ac, abd, a\},$
- $L_6 = \{a^3 b^n c^n : n \geq 0\},$
- $L_7 = \{a^n w w^R a^n : n \geq 0, w \in \{a, b\}^*\},$
- $L_8 = \{a^{n^2} b^n : n \geq 0\},$
- $L_9 = \{a^n b^j a^j b^n : n \geq 0, j \geq 0\},$
- $L_{10} = \{ww : w \in \{a, b\}^*\},$



**4/2014 博士班資格考： 機率與統計 Show All Details.**

1. (25%) A binomial parameter  $p$  describes the probability of success. The corresponding probability of failure is therefore  $q=1-p$ .
  - a. In a random sample of 1000 families which subscribing newspaper in Taipei, it was found that 660 families subscribing to Daily News. Please find the 95% confidence interval for the actual proportion of families subscribing to Daily News.
  - b. How large the sample is required if we want to be 95% confident that the estimate of  $p$  be within 0.02?
2. (25%) Show all details.
  - (a) A random sample of  $n$  observations,  $X_i$  for  $i=1,2,\dots,n$ , is taken from a normal population with mean  $\mu$  and variance  $\sigma^2$ . Let 
$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$
 Find the pdf of  $\bar{X}$ .
  - (b) Let 
$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$
 Find the pdf of  $S^2$ .
  - (c) Let  $F$  be uniformly distributed with pdf,  $f(x) = \begin{cases} 1/2 & , 0 \leq x \leq 2 \\ 0 & , \text{otherwise} \end{cases}$ .  
Let  $F_1$  and  $F_2$  be 2 observations of  $F$  and  $G=F_1+F_2$ . Find the pdf of  $G$ .
3. (20%) Let  $X$  and  $Y$  be two random variables with uniform distribution in the interval  $[0,1]$ . Let  $Z=X+Y$ . Find the probability distribution of  $Z$ .
4. (10%)
  - (a) State the differences among independent, uncorrelated, and orthogonal
  - (b) What is Strict Sense Stationary (SSS) and what is Wide Sense Stationary?
5. (20%) Given a distribution with probability mass function  $f(x, \lambda) = \frac{e^{-\lambda} \lambda^{2x}}{(2x)!}$ ,  $x=0,1,2,\dots$ . Suppose that a random sample  $x_1, x_2, \dots, x_N$  is taken from the process. What is the maximum likelihood estimate of  $\lambda$ ?

**2014 April NCKU CSIE PH.D. Qualification Examination**  
**Computer Architecture**

1. Some operations on two operands (e.g., subtraction) are not commutative. What are the advantages and disadvantages of the stack, accumulator, and load-store architectures when executing noncommutative operations? (15 points)
2. List all the dependences (output, anti, and true) in the following code fragment. Indicate whether the true dependences are loop carried or not. Show why the loop is not parallel. (20 points)  
for (i=2; i<100; i++) {  
    a[i] = b[i] + a[i];     /\* S1 \*/  
    c[i - 1] = b[i] + d[i]; /\* S2 \*/  
    a[i - 1] = 2 \* b[i];   /\* S3 \*/  
    a[i + 1] = 2 \* b[i];   /\* S4 \*/  
}
3. The classical approach to improving cache behavior is to reduce miss rates. Please summarize the techniques that can reduce miss rates. (15 points)
4. Explain why the cache miss penalties increase as the processor becomes much more faster than DRAMs? Briefly describe five optimizations to reduce cache miss penalty. (20 points)
5. Describe two major instruction set characteristics that can further divide general purpose register (GPR) instruction set architecture into three classes, based on whether the instruction operands are used explicitly or implicitly. And show the advantages and disadvantages of these three further divided classes. (15 points)
6. Explain the following synchronization primitives: atomic exchange, test-and-set, and fetch-and-increment. Also, explain what is the pair of instructions, load linked (LL) and store conditional (SC) and how this pair of instructions can be used to implement atomic exchange and fetch-and-increment. (15 points)

Algorithms 資格考 2014 Spring

1. (30%) Answer each part TRUE or FALSE for the little  $o$  notation.
  - a)  $n = o(2n)$ .
  - b)  $2^n = o(n^2)$ .
  - c)  $2^n = o(3^n)$ .
  - d)  $1 = o(n)$ .
  - e)  $n = o(\log n)$ .
  - f)  $1 = o(\frac{1}{n})$ .
2. (10%) How many strongly connected components are there in a  $n$ -node directed path?
3. (10%) Using the master theorem to show the recurrence  $T(n) = 3T(\frac{n}{4}) + n \lg n$ .
4. (20%) Describe a  $\Theta(n \lg n)$ -time algorithm that, given a set  $S$  of  $n$  integers and another integer  $x$ , determine whether or not there exist two elements in  $S$  whose sum is exactly  $x$ .
5. (10%) (a) (5%) Determine which one of the 0-1 knapsack problem and the fractional knapsack problem cannot be solved using the greedy strategy? (b) (5%) Give an example to explain that.
6. (20%) Find a feasible solution or determine that no feasible solution exists for the following system of different constraints:

$$x_1 - x_2 \leq 1,$$

$$x_1 - x_4 \leq -4,$$

$$x_2 - x_3 \leq 2,$$

$$x_2 - x_5 \leq 7,$$

$$x_2 - x_6 \leq 5,$$

$$x_3 - x_6 \leq 10,$$

$$x_4 - x_2 \leq 2,$$

$$x_5 - x_1 \leq -1,$$

$$x_5 - x_4 \leq 3,$$

$$x_6 - x_3 \leq -8.$$