V.S.B. ENGINEERING COLLEGE, KARUR.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

ACADEMIC YEAR 2018 -2019 (ODD SEMESTER)

ASSIGNMENT QUESTIONS

CS6701 - Cryptography and Network Security

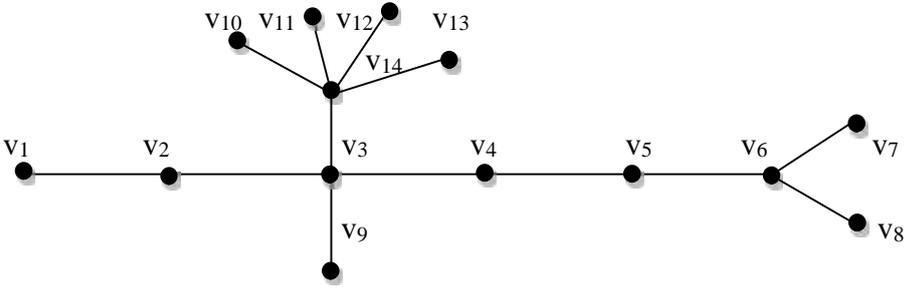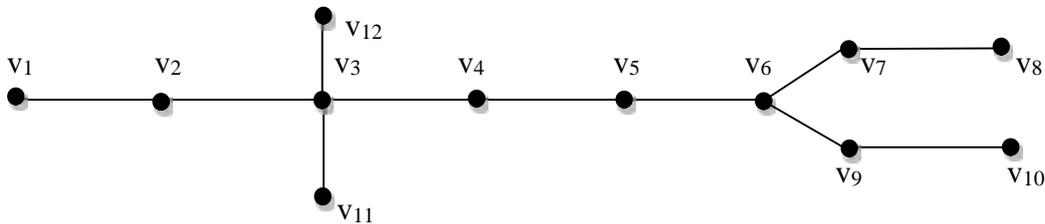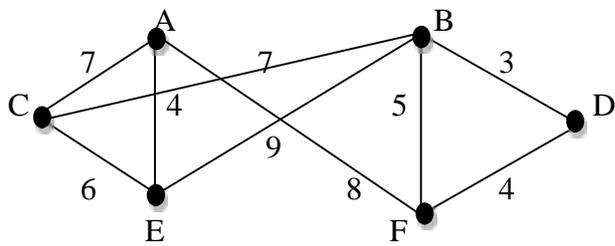| SI. No.	Assignment Question
1.	1. What is the difference between Direct and Arbitrated digital signature? 2. Users A and B use the Diffie-Hellman key exchange technique with a common prime q=11 and a primitive root =2. a) If user A has private key XA = 9, what is A's public key YA? b) If user B has private key XB =4, what is B's public key YB? c) What is the shared secret key?
2.	1. Write the difference between public key, private Key, secret, shared key 2. Discuss the discrete logarithm and explain Diffie-Hellman key Exchange algorithm with its merits and demerits
3.	1. Decrypt the message "ymnsp dtz hfs" using the Caesar cipher with key=5. 2. Encipher the message "notification" Using rail fence cipher of column 3. 3. Encrypt the message "discovery" Using the Vigenere cipher with key "google". 4. Construct a Playfair matrix with the key "security". Using that matrix encrypt the message "this is a secret key"
4.	1. Explain in detail about the elgamal signature algorithm with neat sketch.. 2. Perform encryption and decryption using El Gamal algorithm for a plaintext message 3 and assume p=11,g=2,receipeints private key a=5,and random number chosen by sender is 7 .
5.	1. Solve with the hill cipher. Plain text: "POH" Key: GYBNQKURP. 2. Explain Diffie Hellman key exchange algorithm / key exchange.
6.	1. What is a transpose cipher? Illustrate an example. 2. What is Stegnography? List and explain different techniques. 3. In AES, how the encryption key is expanded to produce keys for the 10 rounds
7.	1. Encrypt the message " meet me at the usual place at ten" using the Hill Cipher with the following key. Show your calculations and the result. (9 4) (5 7) 2. Explain the symmetric key cryptography and asymmetric key cryptography with an example
8.	1. Draw the general structure of AES and explain encryption decryption process 2. Describe the RSA cryptosystem. Given n=100 and e=13, encrypt the message " EXAM" using 00 to 25 for letters A to Z. Decrypt the cipher text to find the original message
9.	1. Explain RSA algorithm and find the private key of a user, if his public key e=21 and n=3599? In RSA scheme is an official standard today? 2. State and explain the different approaches to message authentication functions.

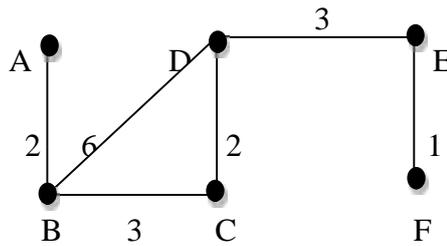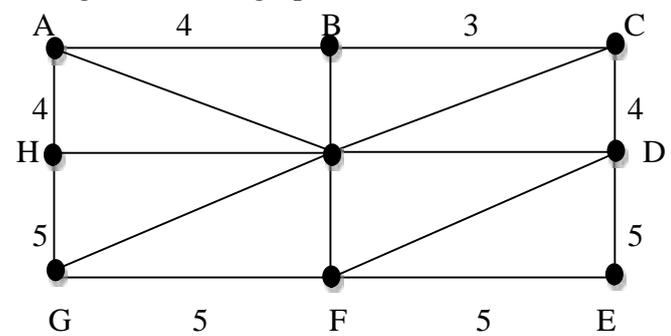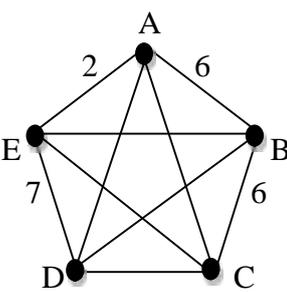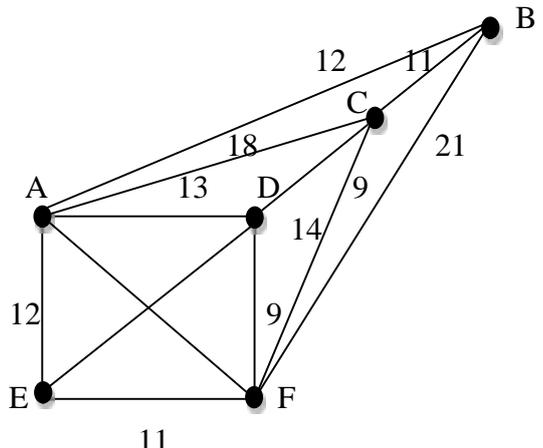|	
10.	1. Explain the computation of hash using SHA-1 OR SECURE HASH ALGORITHM -1. 2. Describe the RSA algorithm and its key generation, encryption and decryption operations.
11.	1. Use Chinese remainder theorem to solve the simultaneous congruence. x 1 mod p for all p {2,3,5,7}. 2. Perform encryption and decryption using RSA algorithm for the following: P=3; q=11; e=7; M=5
12.	1. Explain RSA operations/ RSA key generation/algorithm/RSA encryption and decryption 2. Perform encryption and decryption using RSA algorithms for prime numbers p=3,q=11,e=3,and message = 011101011. 3. Explain RSA applications and performance.
13.	1. What are the typical phases of operation of a virus or worm and how does behaviour blocking software work. 2. Explain the different types of firewall and its configuration in details
14.	1. Explain hill cipher, vigenere cipher and one time pad cipher methods with example. 2. Differentiate between confusion and diffusion 3. Describe the authentication dialogue used by Kerberos for obtaining services from another realm
15.	1. Briefly define a Group, Ring and field. If every element of group G is its own inverse then group G s commutative .Prove 2. Explain Digital Signature Algorithm (DSA) in detail
16.	1. What do you understand by system security ? Describe the way one can prevent virus attack. List the type of virus. ☐3 3 ☐ 2. Use a Hill cipher to encipher the message "Help". Use the following key K= 2 5 . Show the calculations for the corresponding decryption of the cipher text to recover the original plaintext.
17.	1. What is key management? How it is managed. Explain the different methods of key management. 2. With a neat flowchart, describe MD5 processing of a single 512-bit block.
18.	1. Explain Digital signature standard, its approach and algorithm. 2. Explain the types of Intrusion Detection Systems
19.	1. Explain Euclid Algorithm find gcd of 1970 and 1066 by using Euclid Algorithm. Is the two number are relatively prime or not? 2. What are the capabilities, limitations and design goals of firewalls?
20.	1. When a combination of symmetric encryption and an error control code is used for message authentication in what order must the two functions be performed. 2. Explain about Double and Triple DES with an example

21.	1. What are the difference between mono-alphabetic and poly – alphabetic? 2. Explain the concept of password protection system
22.	1. Explain DES algorithm along with round function. 2. Explain Fiestel cipher structure. 3. Explain the requirements of digital signatures and also discuss how problems related to digital signature are taken care by an arbiter?
23.	1. Who are participants in SET system? Describe in brief the sequence of events that are required for transaction 2. Explain in detail about the transformations take place in AES encryption procedure
24.	1. Explain the format of X.509 certificate with a neat figure. 2. Explain in detail about SMIME and its functionalities.
25.	1. Describe the Fermat and Euler's Theorem? What is Euler totient and describe also its limitations. 2. State and explain the different approaches to message authentication or functions.
26.	1. What is AES? Describe some candidates for the AES 2. Discuss about X.509 authentication service in detail.
27.	1. Explain public key infrastructure or functions of PKI. 2. Describe Electronic Mail Security.
28.	1. What is elliptic curve cryptography? Are elliptic curve cryptosystem widely used and how its compared with other cryptosystem? 2. Explain about PGP services.
29.	1. Describe the security of RSA. 2. Differentiate between transport modes Vs. Tunnel mode encryption in IPSec.
30.	1. In SSL and TLS why is there a separate change cipher Spec Protocol, Rather than including a change – cipher-Spec message in the handshake protocol? 2. Describe in detail about the security requirements for hash function and its properties.
31.	1. Discuss conventional encryption model in detail? 2. Sketch the Honey spot Architecture and explain.

32.	1. What is firewall? Describe each type of firewall in detail. 2. Write in detail about SHA algorithm, round function and differentiate between MD5 and SHA
33.	1. Compare the security of Hash function and Message Authentication code against Brute force attack and cryptanalysis. 2. Write in detail about Web Security threats, approaches, protocol stack and Secure Electronic Transaction.
34.	1. What is schedule for the AES? Describe it's all the steps of 128 bit key. 2. Discuss in detail about IPsec Scenario, architecture, benefits and its services.
35.	1. What is public key cryptography? What is the role of the session key in public key scheme? 2. Explain about malicious software in detail.
36.	1. Differentiate….. (A) Hash code and message authentication code (MAC). (B) Weak collision resistance and strong collision resistance 2. List out the participants of SET system, and explain in detail
37.	1. Describe Extended Euclidean Algorithm. 2. Explain PGP message generation and reception.
38.	1. What is difference between a message authentication code and a one – way hash function? Describe the role of a compression function in a Hash function. 2. How does PGP provide confidentiality and authentication service for e-mail and file storage applications? Draw the block diagram and explain its components.
39.	1. What are the essential ingredients of a symmetric cipher? List two basic function used in encryption algorithm. 2. Draw the general structural of DES and explain the encryption decryption process
40.	1. In a Diffie- Hellman key exchange algorithm, let the prime number be 353 and one of its primitive root be 3 and let A and B select their secret key. XA=97and XB=233.Compute the public key of A and B and common secret key. 2. What are the capabilities, limitations and design goals of firewalls?
41.	1. Bring out the difference between secret key cryptography and public key cryptography 2. Explain definition, phases, types of virus structures and types of viruses
42.	1. What basic arithmetical and logical functions are used in MD-5? Explain SHA-1 2. Explain briefly about trusted system.

43.	1. What do you understand by network security attack? Describe active attack and passive security attack. 2. Explain man in the middle attack on Diffie Hellman key exchange algorithm.
44.	1. Why is the segmentation and reassembly function in PGP needed? How does PGP use the concept of trust? 2. Explain SSL handshake protocol. /how a client and a server communicate using SSL handshake protocol.
45.	1. Give the format of X.509 certificate showing his important element of the certificate Explain the format. 2. Explain SSL record layer protocol with a neat figure.
46.	1. What is S/MIME? Why it is used? Explain main function of S/MIME provide? 2. Define the following terms: 1. Certifying Authority b) Addressee c) Digital signature d) Public key.
47.	1. Demonstrate the working of a Kerberos protocol with a neat figure. 2. Explain in detail, the key generation in AES algorithm and its expansion format.
48.	1. Describe in brief fallowing security services Confidentiality. (b) Traffic confidentiality. (c) Access control. (d) Data Integrity. (e) Non-Repudiation. 2. What is the major security aspect in the secure electronic mail system?
49.	1. What Requirement was defined for Kerberos? New the various servers of Kerberos and explain the duties of each server. Write the sequence of message exchanges that happens. 2. Explain the Key Generation, Encryption and Decryption of SDES algorithm in detail
50.	1. What is the SSL and SET? What is the difference between SSL connection and SSL Session? 2. Explain the placement of encryption function for connection oriented and connectionless services With suitable block diagrams
51.	1. Describe how Authentication Header (AH) is used in transport and tunnel modes in IPSec Protocol. 2. Write in detail about how the Diffie-Hellman Key exchange is used for key management schemes?
52.	1. Compose and contrast substitution technique with transposition technique under classical encryption. 2. Describe the block cipher modes of operation in detail 3. Draw the general structure of DES and explain the encryption-decryption process

Sl. No.	Questions
1.	Show that an infinite graph with a finite number of vertices will have at least one pair of vertices joined by an infinite number of parallel edges.
2.	Find the number of leaves in a complete binary tree if it has 29 vertices. A binary tree has 20 leaves. How many vertices does it have?
3.	In a tree with 14 pendant vertices, of the degree of every non-pendant vertex is either 4 or 5. Show that the tree has 3 vertices of degree 4 and 2 vertices of degree 5?
4.	Give an example of a graph which contains (a) A Hamiltonian circuit but not an Euler circuit.　　(b) An Euler circuit but not an Euler circuit. (c) An Euler circuit that is also a Hamiltonian circuit.　(d)Neither an Euler circuit nor a Hamiltonian circuit. (e) An Euler circuit and a Hamiltonian circuit are distinct.
5.	Give an example of a graph which contains (a) A Hamiltonian circuit but not an Euler circuit.　　(b) An Euler circuit but not an Euler circuit. (c) An Euler circuit that is also a Hamiltonian circuit.　(d)Neither an Euler circuit nor a Hamiltonian circuit. (e) An Euler circuit and a Hamiltonian circuit are distinct.
6.	Briefly explain operation on graphs with examples.
7.	Find center, radius and distance of a given tree explain briefly.
8.	Find center, radius and distance of a given tree explain briefly.
9.	Draw the 13 – vertex binary trees with minimum and maximum heights. Find also the path length of both trees.
10.	Explain briefly Travelling Salesman Problem.
11.	Find a minimum spanning tree of the following weighted connected graph.

12.	Find all the spanning trees of the graph G and which is the minimal spanning tree of G A D 3 E 2 6 2 1 B 3 C F
13.	Find all the spanning trees of the graph G and which is the minimal spanning tree of G A 4 B 3 C 4 4 H D 5 5 G 5 F 5 E
14.	Find the minimum spanning tree for the following weighted graph by using Kruskal's algorthim. A 2 6 E B 7 6 D C
15.	Find the minimum spanning tree for the following weighted graph by using Kruskal's algorthim. B 12 11 C 18 21 A 13 D 9 14 12 9 E F 11
16.	Find the minimum spanning tree for the following weighted graph by using Prim's algorthim. A 5 B 4 10 C E 6 7
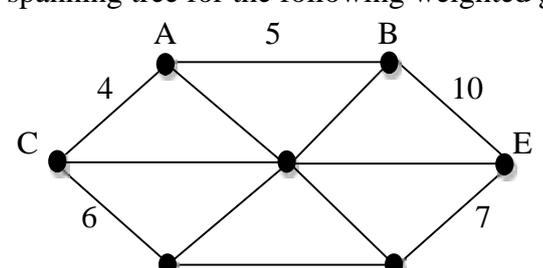
	F　　　2　　　G
17.	Find the minimum spanning tree for the following weighted graph by using Prim's algorthim.
18.	Find the value of the maximal flow in the following network.
19.	Prove that $K \leq \lambda \leq \delta$ and give an example and then explain briefly.
20.	Explain the application for Euler graph.
21.	Find the chromatic polynomial of the following graph G.
22.	Find the chromatic polynomial of the following graph G.
23.	Explain briefly Five – Colour theorem.

24.	Explain briefly applications of Five – Colour theorem.
25.	Show that the vertices of a planar graph with less than 30 edges are 4 – Colourable.

26.	Find the coefficient of :
	    (i) xyz^5 in the expansion of $(x+y+z)^7$.
	    (ii) $x^2y^2z^3$ in the expansion of $(x+y+z)^7$.
	    (iii) $x^3y^3z^2$ in the expansion of $(2x-3y+5z)^8$.
	    (iv) $w^3x^2yz^2$ in the expansion of $(2w-x+3y-2z)^8$.
	    (v) $x_1^2x_3x_4^3x_5^4$ in the expansion of $(x_1+x_2+x_3+x_4+x_5)^{10}$.

27.	Find the number of solutions of the equations $x_1+x_2+x_3+x_4 = 15$ in (a) non-negative integers, (b) positive integers, (c) integers satisfying $x_1 > 2$, $x_2 > -2$, $x_3 > 0$, $x_4 > -3$.

28.	(a) Find the number of non-negative integer solutions of each of the following inequalities:
	    $x_1+x_2+x_3+x_4 \le 6$ (ii) $x_1+x_2+x_3+x_4+x_5 \le 19$.
	(b) How many integer solutions are there to the equation $x_1+x_2+x_3+x_4+x_5 = 20$, where $x_1 \ge 3$,
	    $x_2 \ge 2$, $x_3 \ge 4$,
	    $x_4 \ge 6$, $x_5 \ge 0$?

29.	Using the principle of inclusion and exclusion find the number of prime numbers not exceeding 100.
30.	Find the number of integers between 1 and 2000 inclusive that are not divisible by 2, 3, 5 or 7.
31.	Using the principle of inclusion and exclusion, how many solutions does $x_1+x_2+x_3 = 11$ have where x_1, x_2 and x_3 are non-negative integers with $x_1 \le 3$, $x_2 \le 4$ and $x_3 \le 6$?
32.	Students S_1, S_2, S_3, S_4 are to be accommodated in desks D_1, D_2, D_3, D_4, D_5, D_6 such that no two of these students sit in the same desk. S_1 does not want D_1 and D_2; S_2 does not want D_3; S_3 does not want D_3 and D_4, and S_4 does not want D_4, D_5 and D_6. In how many ways can they be accommodated?
33.	A pair of dice, one red and one white is rolled six times. What is the probability that we obtain all six values on both the red die and the white die if we know that the ordered pairs (1,2), (2,1), (2,5), (3,4), (4,1), (4,5) and (6,6) did not occur ?[Here, an ordered pair (a,b) indicates a on the red die and b on the white].
34.	How many integers between 1 to 100 that are (i) not divisible by 7, 11 or 13 (ii) divisible by 3 but not 7.
35.	How many positive integers not exceeding 1000 that is divisible by 7 or 11.
36.	How many integers between 1 to 300 that are divisible by (i) atleast one of 3,5,7 (ii) 3 and 5 but not by 7 (iii) 5 but not by 3 and 7.
37.	A machine that inserts letters into envelopes goes haywire and inserts letters randomly into envelopes. What is the probability that in a group of 100 letters. (a) no letters is put into the correct envelope. (b) exactly 1 letter is put into the correct envelope. (c) exactly 99 letters put into correct envelope. (d) all letters put into correct envelope.
38.	Using generating f unction, find $p_d(n)$ for $n = 5$ and $n = 7$, and $p_0(n)$ for $n = 4$ and $n = 7$.
39.	Find the formula to express $0^3+1^3+2^3+3^3+\ldots+n^3$ as a function of n.
40.	Solve the following recurrence relations by the method of generating functions. (i) $a_n = 10a_{n-1}+29a_{n-2}$, $n \ge 3$, $a_1 = 10$, $a_2 = 100$ (ii) $a_n+5a_{n-1}+5a_{n-2} = 0$, $n \ge 2$, $a_0 = 0$, $a_1 = 2\sqrt{5}$.
41.	Solve the following recurrence relations by the method of generating functions (i) $a_n - a_{n-1} = 2$, $n \ge 1$, $a_0 = 6$ (ii) $a_n - 3a_{n-1} = 5 \times 3^n$, $n \ge 1$, $a_0 = 2$.

42.	Solve the recurrence relations by the method of generating functions $a_{n+1} = a_n + 3n^2 - n$, $n \geq 0$, $a_0 = 3$.
43.	Solve the recurrence relations by the method of generating functions $a_n - 5a_{n-1} + 6a_{n-2} = n(n-1)$, $n \geq 2$, $a_0 = 1$, $a_1 = 5$.
44.	Using generating functions, solve the recurrence relation $a_k - a_{k-1} - 2a_{k-2} = 2^k$, gives $a_0 = 4$ and $a_1 = 12$.
45.	Solve the recurrence relation $a_{n+2} - 5a_{n+1} + 6a_n = 2$, $n \geq 0$, $a_0 = 3$, $a_1 = 7$.
46.	Solve the recurrence relation $a_{n+2} - 6a_{n+1} + 9a_n = 3(2^n) + 7(3^n)$, $n \geq 0$, $a_0 = 1$, $a_1 = 4$.
47.	Solve the recurrence relation $a_{n+1} - 2a_n = 4^n$, $n \geq 0$, $a_0 = 1$.
48.	Use the method of generating function to solve the recurrence relation $a_n = 4a_{n-1} + 3n.2^n$, $n \geq 1$, given that $a_0 = 4$.
49.	Solve $a_n + 5a_{n-1} + 6a_{n-2} = 3n^2 - 2n + 1$.
50.	Solve $a_n - 7a_{n-1} + 10a_{n-2} = 8k + 6$ with $a_0 = 1$, $a_1 = 2$.
51.	The population of Mumbai city is 60,00,000 at the end of the year 2015. The number of immigrants is 2,00,000 n at the end of year n. The population of the city increases at the rate of 5 % per year. Use recurrence relation to determine the population of the city at the end of 2025 ?

52.	(i) A bank pays 6% (annual) interest on savings, compounding the interest monthly. If priya deposits Rs.1000 on the first day of may, how much will this deposit be worth a year later ?
	
	 (ii) A person invests Rs.1,00,000 at 12 % interest compounded annually.
	     (a) Find the amount at the end of 1st, 2nd, 3rd year.
	     (b) Write the general explicit formula.
	     (c) How long will it take to double the investments.

CS6703 – Grid and Cloud Computing

Sl.No.	Assignment Topics
1.	Compare the similarities and differences between traditional computing clusters/ grids and thecomputing clouds launched in recent years. Consider all technical and economic aspects as listed below. Answer the following questions against real example systems or platforms built in recent years.Also discuss the possible convergence of the two computing paradigms in the future.
2.	a. Hardware, software, and networking support b. Resource allocation and provisioning methods.
3.	Briefly define the following basic techniques and technologies that represent recent related advances in computer architecture, parallel processing, distributed computing, Internet technology, and information services: a. High-performance computing (HPC) system b. High-throughput computing (HTC) system
4.	c. Peer-to-peer (P2P) network d. Computer cluster versus computational grid.
5.	"Is big data analysis relates to cloud computing" Justify your answer.
6.	
7.	Briefly define the following basic techniques and technologies that represent recent related advances in computer architecture, parallel processing, distributed computing, Internet technology, and information services: a.. Virtual machine versus virtual infrastructure
8.	b. Public cloud versus private cloud.
9.	Explain the differences between hypervisor and para-virtualization and give one example VMM (virtual machine monitor), that was built in each of the two categories.
10.	
11.	Circle only one correct answer in each of the following two questions: 1. In the 2009 Top 500 list of the fastest computer systems, which architecture dominates? a. Symmetric shared-memory multiprocessor systems. b. Centralized massively parallel processor (MPP) systems. c. Clusters of cooperative computers. 2. In a cloud formed by a cluster of servers, all servers must be selected as follows: a. All cloud machines must be built on physical servers.

12.	b. All cloud machines must be built with virtual servers. c. The cloud machines can be either physical or virtual servers.Explain each of the terms.
13.	An increasing number of organizations in industry and business sectors adopt cloud systems. Answer the following questions regarding cloud computing: a. List and describe the main characteristics of cloud computing systems. b. Discuss key enabling technologies in cloud computing systems. c. Discuss different ways for cloud service providers to maximize their revenues.
14.	
15.	Briefly answer the following questions on virtualization levels. Highlight the key points and identify the distinctions in different approaches. Discuss their relative advantages, shortcomings and limitations. Also identify example systems implemented at each level.
16.	
17.	Explain about parallel and distributed programming models.
18.	
19.	Install the VMware Workstation on a Windows XP or Vista personal computer or laptop, and then install Red Hat Linux and Windows XP in the VMware Workstation. Configure the network settings of Red Hat Linux and Windows XP to get on the Internet. Write an installation and configuration guide for the VMware Workstation, Red Hat Linux, and Windows XP systems. Include any troubleshooting tips in the guide.
20.	
21.	Compile a table to compare public clouds and private clouds in each of the following four aspects. Also identify their differences, advantages, and shortcomings in terms of design technologies and application flexibility. Give several example platforms that you know of under each cloud class. a. Technology leveraging and IT resource ownership. b. Provisioning methods of resources including data and VMs, and their

22.	management. c. Workload distribution methods and loading policies. d. Security precautions and data privacy enforcement.
23. 24.	Describe the following techniques or terminologies used in cloud computing and cloud services. Use a concrete example cloud or case study to explain the addressed technology. a. Virtualized data center b. Green information technology c. Multitenant technique
25. 26.	Consider two cloud service systems: Google File System and Amazon S3. Explain how they achieve their design goals to secure data integrity and to maintain data consistency while facing the programs of hardware failure, especially concurrent hardware failures.
27. 28.	Explain the differences in the following two machine recovery schemes. Comment on their implementation requirements, advantages and shortcomings, and application potential. a. Recovery of a physical machine failure by another physical machine. b. Recovery of a VM failure by another VM.
29. 30.	Consider the following list of resources allocation, code files, data transfers, and computation tasks performed in typical grid experiments. You are asked to execute a typical job on a computational grid that is available to you. Discuss the resource items that are controlled by the grid resource management system on your platform. Either the Globus GT4 or the CGSP are candidate systems to use. a. Resources allocated to network transfers b. Resources with advance reservation c. Network characteristics and load d. Relevant load information: available space and bandwidth utilization.
31. 32.	Briefly specify the functionality of the software modules GRAM, Nexus, GSI, MDS, and GASS implemented in the Globus GT-4 middleware library. Try to illustrate an example application in using each of the five functional modules..
33.	Circle only one correct answer in the following question related to the use of GT4 or CGSP middleware tools for grid computing. Justify why you chose your answer. Repeat the question for those features that are not supported by CGSP. a. Grid workflow with WSRF support. b. JSDL job submission and legacy job execution using General Running Service. c. Heterogeneous integration using virtual database and virtual table.

34.	d. Multi-Local Scheduler support including PBS, LSF, and Condor.
35.	This is a survey report about all the P2P grids or volunteer computing grids that can manage to join as contributing hosts. Compare the relative performance of all the volunteer grids you have used. Discuss their strengths and weakness in hardware capabilities, software support, and middleware services provided.
36.	
37.	Visit the Google App Engine (GAE) web site, download the SDK, and read the Python Guide or the Java Guide to get started. Note that GAE only accepts Python, Ruby, and Java programming languages. The platform does not provide any IaaS services. a. Develop a specific cloud application using available software services such as Gmail, Docs, or CRM on the GAE platform. Test your applications on the GAE platform. b. Report your application development experience and experimental results in terms of some selected performance measures such as job queuing time, execution time, resource utilization rate, or some QoS attributes such as goal achievement, success rate, fault tolerance, and cost effectiveness. c. Change the program size or data set size and platform configuration to study the scalability and efficiency issues in your GAE experiments.
38.	
39.	Breif note on grid middleware tools.
40.	
41.	Twister K-means extends the Map Reduce programming model iteratively. Many data analysis techniques require iterative computations. For example, K-means clustering is the application where multiple iterations of Map Reduce computations are necessary for overall computation. Twister is an enhanced Map Reduce runtime that supports iterative Map Reduce computations efficiently. For this, study the iterative Map Reduce programming model and learn how to implement the K-means algorithm with Twister.
42.	
43.	Briefly define the following basic techniques and technologies that represent recent related advances in computer architecture, parallel processing, distributed computing, Internet technology, and information services: a. Radio-frequency identifier (RFID) b. Global positioning system (GPS)

44.	c. Sensor network d. Internet of Things (IoT) e. Cyber-physical system (CPS)
45.	Describe data privacy and data life cycle.
46.	
47.	Visit the Microsoft Windows Azure Developer Centre. You can download the Azure development kit to run a local version of Azure. Design an application experiment and test it on a local computer, such as your desktop, or a notebook computer, or a university workstation or server. Report your experiment experiences in using the Azure platform.
48.	
49.	Consider the following list of resources allocation, code files, data transfers, and computation tasks performed in typical grid experiments. You are asked to execute a typical job on a computational grid that is available to you. Discuss the resource items that are controlled by the grid resource management system on your platform. Either the Globus GT4 or the CGSP are candidate systems to use. a. Starting, monitoring, and controlling the execution of the resultant processes b. Putting and getting files
50.	c. Relevant load information: current load and queue state
51.	Briefly define the following basic techniques and technologies that represent recent related advances in computer architecture, parallel processing, distributed computing, Internet technology, and information services:
52.	a. Service-oriented architecture (SOA) b. Pervasive computing versus Internet computing.

Sl. No.	Questions
53	Enter the following LP into TORA, and select the graphic solution mode to reveal the LP graphical screen. Minimixe $Z=3x_1 + 8x_2$ subject to $x_1 + x_2 \geq 8$; $2x_1 - 3x_2 \leq 0$; $x_1 + 2x_2 \leq 30$; $3x_1 - x_2 \geq 0$; $x_1 \leq 10$; $x_2 \geq 9$ a $x_1, x_2 \geq 0$.
54	Solve the following LP problem graphically : Maximize $Z=60x_1 + 90x_2$ subject to $x_1 + 2x_2 \leq 40$; $2x_1 + 3x_2 \leq 90$; $x_1 - x_2 \geq 10$ a $x_1, x_2 \geq 0$.
55	Solve the following LP problem using Simplex method: Maximize $Z=3x_1 + 2x_2 + 5x_3$ subject to $x_1 + x_2 + x_3 \leq 9$; $2x_1 + 3x_2 + 5x_3 \leq 30$; $2x_1 - x_2 - x_3 \leq 8$ a $x_1, x_2, x_3 \geq 0$.
56	Solve the following LP problem using Simplex method: Maximize $Z=5x_1 + 3x_2 + 7x_3$ subject to $x_1 + x_2 + 2x_3 \leq 22$; $3x_1 + 2x_2 + x_3 \leq 26$; $x_1 + x_2 + x_3 \leq 18$ a $x_1, x_2, x_3 \geq 0$.
57	Solve the following LP problem using dual Simplex method: Minimize $Z=x_1 + 2x_2 + 3x_3$ subject to $2x_1 - x_2 + x_3 \geq 4$; $x_1 + x_2 + 2x_3 \leq 8$; $x_2 - x_3 \geq 2$ a $x_1, x_2, x_3 \geq 0$.
58	Consider the following LP problem: Maximize $Z= 3x_1 + 2x_2 - 5x_3$ subject to $x_1 + x_2 \leq 2$; $2x_1 + x_2 + 6x_3 \leq 6$; $x_1 - x_2 + 3x_3 = 0$ and $x_1, x_2, x_3 \geq 0$. Solve the above LP problem. If the right hand side of the primal is changed (2, 6, 0) to (2, 10, 5), find the optimal solution.
59	Explain the productivity improvement tool in RMT.
60	List out the scope of the Operation Research.
61	Consider the assignment problem as shown below. In this problem, five different jobs are to be assignd to five operators such that the total processing time is minimized. The matrix entries represent processing times in hours. Operators |	1	2	3	4	5
---	---	---	---	---	---
1	5	6	8	6	4
2	4	8	7	7	5
3	7	7	4	5	4
4	6	5	6	7	5
5	4	7	8	6	8	 Find the optimal allocation of Jobs to different Operator.
62	Consider the problem of assigning four sales persons to four different sales regions as shown below such that the total sales is maximized. sales regions 		1	2	3	4
---	---	---	---	---
1	5	11	8	9
2	5	7	9	7
3	7	8	9	9
4	6	8	11	12	 The cell entries represent annual sales figures in cores o rupees. Find the optimal allocation of

	the sales persons to different regions.
63	Solve the following LP problem using two phase method: Minimize Z=$10x_1 + 6x_2 + 2x_3$ subject to $-x_1 + x_2 + x_3 \geq 1$; $3x_1 + x_2 - x_3 \geq 2$ *a* $x_1, x_2, x_3 \geq 0$.
64	Use duality to Solve the following LP problem: Maximize Z=$2x_1 + x_2$ subject to $x_1 + 2x_2 \leq 10$; $x_1 + x_2 \leq 6$; $x_1 - x_2 \leq 2$; $x_1 - 2x_2 \leq 1$ *a* $x_1, x_2 \geq 0$.
65	Solve the follown LP problem: Maximize Z=$x_1 + 5x_2 + 3x_3$ subject to $x_1 + 2x_2 + x_3 = 3$; $2x_1 - x_2 = 4$ *a* $x_1, x_2, x_3 \geq 0$. If the objective functin is changed to (2, 5, 2), ind the new optimum solution.
66	Solve the following LP Problem: Maximize Z=$3x_1 + 5x_2$ subject to $x_1 + x_2 \leq 4$; $3x_1 + 2x_2 \leq 18$ *a* $x_1, x_2 \geq 0$. If a new variable is inclided in above LP problem with aprofit of 7, and 2 and 4 as the coefficients of the constraints, respectively, find the solution to the new problem.
67	Solve the following LP problem using Simplex method: Maximize Z=$20x_1 + 80x_2$ subject to $4x_1 + 6x_2 \leq 90$; $8x_1 + 6x_2 \leq 100$ *a* $x_1, x_2 \geq 0$.If the following new constraint is added to this model, find the solution to the new problem. $5x_1 + 4x_2 \leq 80$.
68	Write the procedure for each of the following : (a) Northwest corner cell method (b) Least cost cell method (c) Vogel's approximation method (d) U-V method
69	Write the practical applications of assignment problem.
70	Explain the similarity between transportation problem and assignment problem.
71	A manufacturing company has three factories F_1, F_2 and F_3 with monthly manufacturing capacities of 7000, 4000 and 10,000 units o a product. The product is to be supplied to seven stores. The manufacturing costs in these factories are slightly diferent but the important factor is the shipping cost from each factory to a particular store. The following able represents the facotry capacities, store requirements and unit cost (in rupees) of shipping from each factory to each store. Here, slack is the difference between the total factory capacity and the total requirement. |	S_1	S_1	S_1	S_1	S_1	S_1	S_1	Factory Capacity
---	---	---	---	---	---	---	---	---
F_1	5	6	4	3	7	5	4	7000
F_2	9	4	3	4	3	2	1	4000
F_3	8	4	2	5	4	8	3	10,000
Store	1	2	4	4	2	3	3	
Dema	5	0	5	0	5	5	0	
nd	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0		 Find the optimal transportation plan so as to minimize the transportation cost.
72	Determine an initial basic feasible solution to the following transportation problem using northwest corner cell method.

	To					
	1	2	3	4	5	Availability
1	3	4	6	8	9	20
2	2	10	1	5	8	30
3	7	11	20	40	3	15
4	2	1	9	14	16	13
Demand	40	6	8	18	6	

73

Find the initial basuc feasible solution to the following transportation problem by
 (a) Northwest corner cell method and
 (b) Least cost cell method.

	To			
	1	2	3	Supply
1	2	7	4	5
2	3	3	1	8
3	5	4	7	7
4	1	6	2	14
Demand	2	9	18	

State which of the methods is better.

74

Find the initial basic feasible solution of the following transpotation problem by Vogel's approximation method:

	Warehouses				
	W_1	W_2	W_3	W_4	Capacity
F_1	10	30	50	10	7
F_2	70	30	40	60	9
F_3	40	80	70	20	18
Demand	5	8	7	14	34

75

Determine an initial basic feasible solution to the ollowing transportation problem using:
(a) Northwest corner cell method and
(b) Vogel's approximation method.

	A_1	B_1	C_1	D_1	E_1	Supply
A	2	1	11	3	7	4
			0			
B	1	4	7	2	1	8
C	3	9	4	8	12	9
Demand	3	3	4	5	6	

76 Solve the following LP problem graphically :
Maximize $Z=60x_1 + 90x_2$
subject to $x_1 + 2x_2 \le 40; 2x_1 + 3x_2 \le 90; x_1 - x_2 \ge 10$ a $x_1, x_2 \ge 0.$

77 Solve the following LP problem using Simplex method:
Maximize $Z=5x_1 + 3x_2 + 7x_3$
subject to $x_1 + x_2 + 2x_3 \le 22; 3x_1 + 2x_2 + x_3 \le 26; x_1 + x_2 + x_3 \le 18$ a $x_1, x_2, x_3 \ge 0.$

78 Consider the following LP problem:
Maximize $Z= 3x_1 + 2x_2 - 5x_3$
subject to $x_1 + x_2 \le 2; 2x_1 + x_2 + 6x_3 \le 6; x_1 - x_2 + 3x_3 = 0$ and $x_1, x_2, x_3 \ge 0.$

	Solve the above LP problem. If the right hand side of the primal is changed (2, 6, 0) to (2, 10, 5), find the optimal solution.
79	Consider the problem of assigning four sales persons to four different sales regions as shown below such that the total sales is maximized. sales regions 		1	2	3	4
---	---	---	---	---
1	5	1 1	8	9
2	5	7	9	7
3	7	8	9	9
4	6	8	1 1	1 2	 The cell entries represent annual sales figures in cores o rupees. Find the optimal allocation of the sales persons to different regions.
80	Use duality to Solve the following LP problem: Maximize Z=$2x_1 + x_2$ subject to $x_1 + 2x_2 \le 10$; $x_1 + x_2 \le 6$; $x_1 - x_2 \le 2$; $x_1 - 2x_2 \le 1$ a $x_1, x_2 \ge 0$.
81	Solve the following LP Problem: Maximize Z=$3x_1 + 5x_2$ subject to $x_1 + x_2 \le 4$; $3x_1 + 2x_2 \le 18$ a $x_1, x_2 \ge 0$. If a new variable is inclided in above LP problem with aprofit of 7, and 2 and 4 as the coefficients of the constraints, respectively, find the solution to the new problem.
82	Write the procedure for each of the following : (a) Northwest corner cell method (b) Least cost cell method (c) Vogel's approximation method (d) U-V method
83	Explain the similarity between transportation problem and assignment problem.
84	Determine an initial basic feasible solution to the following transportation problem using northwest corner cell method. To |	1	2	3	4	5	Availability
---	---	---	---	---	---	---
1	3	4	6	8	9	20
2	2	10	1	5	8	30
3	7	11	20	40	3	15
4	2	1	9	14	16	13
Demand	40	6	8	18	6	
85	Find the initial basic feasible solution of the following transpotation problem by Vogel's approximation method: Warehouses 		W_1	W_2	W_3	W_4	Capacity
---	---	---	---	---	---
F_1	10	30	50	10	7
F_2	70	30	40	60	9
F_3	40	80	70	20	18
Demand	5	8	7	14	34	
86	Enter the following LP into TORA, and select the graphic solution mode to reveal the LP graphical screen. Minimixe Z=$3x_1 + 8x_2$

	subject to $x_1 + x_2 \geq 8$; $2x_1 - 3x_2 \leq 0$; $x_1 + 2x_2 \leq 30$; $3x_1 - x_2 \geq 0$; $x_1 \leq 10$; $x_2 \geq 9$ a $x_1, x_2 \geq 0$.
87	Solve the following LP problem using Simplex method: Maximize Z=$3x_1 + 2x_2 + 5x_3$ subject to $x_1 + x_2 + x_3 \leq 9$; $2x_1 + 3x_2 + 5x_3 \leq 30$; $2x_1 - x_2 - x_3 \leq 8$ a $x_1, x_2, x_3 \geq 0$.
88	Solve the following LP problem using dual Simplex method: Minimize Z=$x_1 + 2x_2 + 3x_3$ subject to $2x_1 - x_2 + x_3 \geq 4$; $x_1 + x_2 + 2x_3 \leq 8$; $x_2 - x_3 \geq 2$ a $x_1, x_2, x_3 \geq 0$.
89	Explain the productivity improvement tool in RMT.
90	Consider the assignment problem as shown below. In this problem, five different jobs are to be assignd to five operators such that the total processing time is minimized. The matrix entries represent processing times in hours. Operators 		1	2	3	4	5
---	---	---	---	---	---
1	5	6	8	6	4
2	4	8	7	7	5
3	7	7	4	5	4
4	6	5	6	7	5
5	4	7	8	6	8	 Find the optimal allocation of Jobs to different Operator.
91	Solve the following LP problem using two phase method: Minimize Z=$10x_1 + 6x_2 + 2x_3$ subject to $-x_1 + x_2 + x_3 \geq 1$; $3x_1 + x_2 - x_3 \geq 2$ a $x_1, x_2, x_3 \geq 0$.
92	Solve the follown LP problem: Maximize Z=$x_1 + 5x_2 + 3x_3$ subject to $x_1 + 2x_2 + x_3 = 3$; $2x_1 - x_2 = 4$ a $x_1, x_2, x_3 \geq 0$. If the objective functin is changed to (2, 5, 2), and the new optimum solution.
93	Solve the following LP problem using Simplex method: Maximize Z=$20x_1 + 80x_2$ subject to $4x_1 + 6x_2 \leq 90$; $8x_1 + 6x_2 \leq 100$ a $x_1, x_2 \geq 0$.If the following new constraint is added to this model, find the solution to the new problem. $5x_1 + 4x_2 \leq 80$.
94	Write the practical applications of assignment problem.
95	A manufacturing company has three factories F_1, F_2 and F_3 with monthly manufacturing capacities of 7000, 4000 and 10,000 units o a product. The product is to be supplied to seven stores. The manufacturing costs in these factories are slightly diferent but the important factor is the shipping cost from each factory to a particular store. The following able represents the facotry capacities, store requirements and unit cost (in rupees) of shipping from each factory to each store. Here, slack is the difference between the total factory capacity and the total requirement. |	S_1	S_1	S_1	S_1	S_1	S_1	S_1	Factory Capacity
---	---	---	---	---	---	---	---	---
F_1	5	6	4	3	7	5	4	7000
F_2	9	4	3	4	3	2	1	4000
F_3	8	4	2	5	4	8	3	10,000
Store Demand	1500	200	4500	400	2500	3550	300		 Find the optimal transportation plan so as to minimize the transportation cost.

		Find the initial basuc feasible solution to the following transportation problem by

| 96 | Find the initial basuc feasible solution to the following transportation problem by
(a) Northwest corner cell method and
(b) Least cost cell method.

To
<table><tr><td></td><td>1</td><td>2</td><td>3</td><td>Supply</td></tr><tr><td>1</td><td>2</td><td>7</td><td>4</td><td>5</td></tr><tr><td>2</td><td>3</td><td>3</td><td>1</td><td>8</td></tr><tr><td>3</td><td>5</td><td>4</td><td>7</td><td>7</td></tr><tr><td>4</td><td>1</td><td>6</td><td>2</td><td>14</td></tr><tr><td>Demand</td><td>2</td><td>9</td><td>18</td><td></td></tr></table>
State which of the methods is better. |

96 Find the initial basuc feasible solution to the following transportation problem by
(a) Northwest corner cell method and
(b) Least cost cell method.

To

	1	2	3	Supply
1	2	7	4	5
2	3	3	1	8
3	5	4	7	7
4	1	6	2	14
Demand	2	9	18	

State which of the methods is better.

97 Determine an initial basic feasible solution to the ollowing transportation problem using:
(a) Northwest corner cell method and
(b) Vogel's approximation method.

	A_1	B_1	C_1	D_1	E_1	Supply
A	2	11	10	3	7	4
B	1	4	7	2	1	8
C	3	9	4	8	12	9
Demand	3	3	4	5	6	

98 Find the initial basic feasible solution of the following transpotation problem by Vogel's approximation method:

Warehouses

	W_1	W_2	W_3	W_4	Capacity
F_1	10	30	50	10	7
F_2	70	30	40	60	9
F_3	40	80	70	20	18
Demand	5	8	7	14	34

99 Explain the similarity between transportation problem and assignment problem.

10 Write the practical applications of assignment problem.

10 Solve the following LP Problem:
Maximize $Z=3x_1 + 5x_2$
subject to $x_1 + x_2 \le 4$; $3x_1 + 2x_2 \le 18$ a $x_1, x_2 \ge 0$. If a new variable is inclided in above LP problem with aprofit of 7, and 2 and 4 as the coefficients of the constraints, respectively, find the solution to the new problem.

10 Consider the problem of assigning four sales persons to four different sales regions as shown below such that the total sales is maximized.

sales regions

	1	2	3	4
1	5	11	8	9
2	5	7	9	7
3	7	8	9	9
4	6	8	11	12

	The cell entries represent annual sales figures in cores o rupees. Find the optimal allocation of the sales persons to different regions.
10	Solve the following LP problem using two phase method: Minimize $Z=10x_1 + 6x_2 + 2x_3$ subject to $-x_1 + x_2 + x_3 \geq 1$; $3x_1 + x_2 - x_3 \geq 2$ a $x_1, x_2, x_3 \geq 0$.
10	Solve the following LP problem using Simplex method: Maximize $Z=5x_1 + 3x_2 + 7x_3$ subject to $x_1 + x_2 + 2x_3 \leq 22$; $3x_1 + 2x_2 + x_3 \leq 26$; $x_1 + x_2 + x_3 \leq 18$ a $x_1, x_2, x_3 \geq 0$.

IT6801 - Service Oriented Architecture

SI. No.	Assignment Question
1	SAX Parsing for invoicing a customer's order in XML file.
2	Online railway reservation system using XML
3	Employee details using XML - DTD
4	Hospital management system using XML-Schema
5	Library management systems using X-Path
6	E-mail system using DTD
7	Purchase order using DOM
8	Student details using style-sheets
9	Online purchasing system using SAX
10	Online ticket booking systems using –XML
11	Online payment system using DOM
12	Bookstore Program using style-sheet
13	Department details using DTD
14	DOM Parsing for invoicing a customer's order in XML file.
15	Online railway reservation system using XML DTD
16	Employee details using XML - Schema
17	Hospital management system using XML DTD
18	Library management systems using X-Link
19	E-mail system using X-Path
20	Purchase order using SAX
21	Student details using –XML Schema
22	Online purchasing system using DTD
23	Online ticket booking systems using –SAX
24	Online payment system using XML
25	Bookstore Program using style – XML path
26	Department details using XML
27	SAX Parsing for invoicing a customer's order in XML file.
28	Online railway reservation system using XML
29	Employee details using XML - DTD
30	Hospital management system using XML-Schema
31	Library management systems using X-Path
32	E-mail system using DTD
33	Purchase order using DOM
34	Student details using style-sheets
35	Online purchasing system using SAX
36	Online ticket booking systems using –XML
37	Online payment system using DOM
38	Bookstore Program using style-sheet
39	Department details using DTD
40	DOM Parsing for invoicing a customer's order in XML file.
41	Online railway reservation system using XML DTD
42	Employee details using XML - Schema
43	Hospital management system using XML DTD
44	Library management systems using X-Link
45	E-mail system using X-Path
46	Purchase order using SAX
47	Student details using –XML Schema

48	Online purchasing system using DTD
49	Online ticket booking systems using –SAX
50	Online payment system using XML
51	Program using style – XML path
52	Department details using XML

Sl. No.	Assignment Topic
1	Draw the inverted index that would be built for the following document collection. Doc 1 new home sales top forecasts Doc 2 home sales rise in july Doc 3 increase in home sales in july Doc 4 july new home sales rise
2	For a conjunctive query, is processing postings lists in order of size guaranteed to be optimal? Explain why it is, or give an example where it isn't.
3	When using weighted zone scoring, is it necessary for all zones to use the same Boolean match function?
4	If we were to stem jealous and jealousy to a common stem before setting up the vector space, detail how the definitions of tf and idf should be modified.
5	Consider the case of a query term that is not in the set of M indexed terms; thus our standard construction of the query vector results in V~ (q) not being in the vector space created from the collection. How would one adapt the vector space representation to handle this case?
6	An IR system returns 8 relevant documents, and 10 non relevant documents. There are a total of 20 relevant documents in the collection. What is the precision of the system on this search, and what is its recall?
7	The balanced F measure (a.k.a. F1) is defined as the harmonic mean of precision and recall. What is the advantage of using the harmonic mean rather than "averaging" (using the arithmetic mean)?
8	In Rocchio's algorithm, what weight setting for / / does a "Find pages like this one" search correspond to?
9	Why is positive feedback likely to be more useful than negative feedback to an IR system? Why might only using one nonrelevant document be more effective than using several?
10	What are the differences between standard vector space tf-idf weighting and the BIM probabilistic retrieval model (in the case where no document relevance information is available)?
11	Omar has implemented a relevance feedback web search system, where he is going to do relevance feedback based only on words in the title text returned for a page (for efficiency). The user is going to rank 3 results. The first user, Jinxing, queries for: banana slug and the top three titles returned are: banana slug Ariolimax columbianus Santa Cruz mountains banana slug Santa Cruz Campus Mascot Jinxing judges the first two documents relevant, and the third nonrelevant. Assume that Omar's search engine uses term frequency but no length normalization nor IDF. Assume that he is using the Rocchio relevance feedback mechanism, with = = = 1. Show the final revised query that would be run. (Please list the vector elements in alphabetical order.)
12	Let Xt be a random variable indicating whether the term t appears in a document. Suppose we have |R| relevant documents in the document collection and that Xt = 1 in s of the documents. Take the observed data to be just these observations of Xt for each document in R. Show that the MLE for the parameter pt = P(Xt = 1|R = 1,~q), that is, the value for pt which maximizes the probability of the observed data, is pt = s/|R|.
13	Describe the differences between vector space relevance feedback and probabilistic relevance feedback.
14	How might a language model be used in a spelling correction system? In particular, consider the case of context-sensitive spelling correction, and correcting incorrect usages of words, such as their in Are you their?

15	If the stop probability is omitted from calculations, what will the sum of the scores assigned to strings in the language of length 1 be?
16	Consider making a language model from the following training text: the martian has landed on the latin pop sensation ricky martin a. Under a MLE-estimated unigram probability model, what are P(the) and P(martian)? b. Under a MLE-estimated bigram model, what are P(sensation|pop) and P(pop|the)?
17	The rationale for the positional independence assumption is that there is no useful information in the fact that a term occurs in position k of a document. Find exceptions. Consider formulaic documents with a fixed document structure.
18	Create a training set of 300 documents, 100 each from three different languages (e.g., English, French, Spanish). Create a test set by the same procedure, but also add 100 documents from a fourth language. Train (i) a one-of classifier (ii) an any-of classifier on this training set and evaluate it on the test set. (iii) Are there any interesting differences in how the two classifiers behave on this task?
19	Show that the decision boundaries in Rocchio classification are, as in kNN, given by the Voronoi tessellation.
20	Computing the distance between a dense centroid and a sparse vector is (M) for a naive implementation that iterates over all M dimensions. Based on the equality $(x_i - \mu_i)^2 = 1.0 + \mu_i^2 - 2 x_i \mu_i$ and assuming that μ_i^2 has been pre computed, write down an algorithm that is (M_a) instead, where M_a is the number of distinct terms in the test document.
21	Prove that the region of the plane consisting of all points with the same k nearest neighbors is a convex polygon.
22	Design an algorithm that performs an efficient 1NN search in 1 dimension (where efficiency is with respect to the number of documents N). What is the time complexity of the algorithm?
23	Design an algorithm that performs an efficient 1NN search in 2 dimensions with at most polynomial (in N) preprocessing time.

24	Spam email often makes use of various cloaking techniques to try to get through. One method is to pad or substitute characters so as to defeat word-based text classifiers. For example, you see terms like the following in spam email:
	    Rep1icaRolex bonmus Viiiaaaagra pi11z
	    PHARlbdMACY [LEV]i[IT]l[RA] se∧xual ClAfLlS
	Discuss how you could engineer features that would largely defeat this strategy.

25	Another strategy often used by purveyors of email spam is to follow the message they wish to send (such as buying a cheap stock or whatever) with a paragraph of text from another innocuous source (such as a news article). Why might this strategy be effective? How might it be addressed by a text classifier?
26	What other kinds of features appear as if they would be useful in an email spam classifier?
27	Design an algorithm that performs an efficient 1NN search in 2 dimensions with at most polynomial (in N) preprocessing time.
28	Another strategy often used by purveyors of email spam is to follow the message they wish to send (such as buying a cheap stock or whatever) with a paragraph of text from another innocuous source (such as a news article). Why might this strategy be effective? How might it be addressed by a text classifier?
29	What other kinds of features appear as if they would be useful in an email spam classifier?
30	Define two documents as similar if they have at least two proper names like Clinton or Sarkozy in common. Give an example of an information need and two documents, for which the cluster hypothesis does not hold for this notion of similarity.

31	Make up a simple one-dimensional example (i.e. points on a line) with two clusters where the inexactness of cluster-based retrieval shows up. In your example, retrieving clusters close to the query direct should do worse than nearest neighbor search.
32	(i) Give an example of a set of points and three initial centroids (which need not be members of the set of points) for which 3-means converges to a clustering with an empty cluster. (ii) Can a clustering with an empty cluster be the global optimum with respect
33	Prove that RSSmin(K) is monotonically decreasing in K.
34	There is a soft version of K-means that computes the fractional membership of a document in a cluster as a monotonically decreasing function of the distance from its centroid, e.g., as $e-$. Modify reassignment and recomputation steps of hard K-means for this soft version.
35	Show that K-means can be viewed as the limiting case of EM for Gaussian mixtures if variance is very small and all covariances are 0.
36	The within-point scatter of a clustering is defined as $k\ 1\ 2\ \ \sim xi \in\ k\ \ \sim xj \in\ k\ \|\sim xi - \sim xj\|\ 2$ Show that minimizing RSS and minimizing within-point scatter are equivalent.
37	A single-link clustering can also be computed from the minimum spanning tree of a TREE graph. The minimum spanning tree connects the vertices of a graph at the smallest possible cost, where cost is defined as the sum over all edges of the graph. In our case the cost of an edge is the distance between two documents. Show that if $k-1 >\ k > \ldots >\ 1$ are the costs of the edges of a minimum spanning tree, then these edges correspond to the $k-1$ merges in constructing a single-link clustering.
38	Show that single-link clustering is best-merge persistent and that GAAC and centroid clustering are not best-merge persistent.
39	a. Consider running 2-means clustering on a collection with documents from two different languages. What result would you expect? b. Would you expect the same result when running an HAC algorithm?
40	Download Reuters-21578. Keep only documents that are in the classes crude, interest, and grain. Discard documents that are members of more than one of these three classes. Compute a (i) single-link, (ii) complete-link, (iii) GAAC, (iv) centroid clustering of the documents. (v) Cut each dendrogram at the second branch from the top to obtain K = 3 clusters. Compute the Rand index for each of the 4 clusterings. Which clustering method performs best?
41	Suppose a run of HAC finds the clustering with K = 7 to have the highest value on some prechosen goodness measure of clustering. Have we found the highest-value clustering among all clusterings with K = 7?
42	If the number of pages with in-degree i is proportional to 1/i 2.1, what is the probability that a randomly chosen web page has in-degree 1?
43	If the number of pages with in-degree i is proportional to 1/i 2.1, then as the largest in-degree goes to infinity, does the fraction of pages with in-degree i grow, stay the same, or diminish? How would your answer change for values of the exponent other than 2.1?
44	The average in-degree of all nodes in a snapshot of the web graph is 9. What can we say about the average out-degree of all nodes in this snapshot?
45	The Goto method ranked advertisements matching a query by bid: the highest-bidding advertiser got the top position, the second-highest the next, and so on. What can go wrong with this when the highest-bidding advertiser places an advertisement that is irrelevant to the query? Why might an advertiser with an irrelevant advertisement bid high in this manner?
46	Two web search engines A and B each generate a large number of pages uniformly at random from their indexes. 30% of A's pages are present in B's index, while 50% of B's pages are present in A's index. What is the number of pages in A's index relative to B's?
47	Explain why this estimator would be very difficult to use in practice
48	Why is it better to partition hosts (rather than individual URLs) between the nodes of a distributed crawl system?
49	Why should the host splitter precede the Duplicate URL Eliminator?

50	In the preceding discussion we encountered two recommended "hard constants" – the increment on to being ten times the last fetch time, and the number of back queues being three times the number of crawl threads. How are these two constants related?
51	We noted that expressing a row in terms of one of seven preceding rows allowed us to use no more than three bits to specify which of the preceding rows we are using as prototype. Why seven and not eight preceding rows? (Hint: consider the case when none of the preceding seven rows is a good prototype.)
52	Decoding the links incident on a URL could result in many levels of indirection. Construct an example in which the number of levels of indirection grows linearly with the number of URLs.