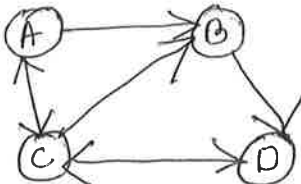


MAY 2022: END SEMESTER ASSESSMENT ,B Tech CSE 6th SEMESTER

UE19CS332: Algorithms for Intelligence Web and Information Retrieval

Time: 3 Hrs		Answer All Questions	Max Marks: 100
1.	a)	Consider these documents Doc1: breakthrough drug for Schizophrenia Doc2: New Schizophrenia drug Doc3: new approach for treatment of Schizophrenia Doc4: new hopes for Schizophrenia patients. Draw the term document incidence matrix for the above document collection Draw the inverted index representation for this collection. What are the results returned for the query i) Schizophrenia AND drug ii) for AND NOT(drug OR Approach)	5
	b)	Give different ways of enumerating “uni*sity”, how can we enumerate all documents containing such terms?	5
	c)	If you wanted to search for s*ng in a permuterm wildcard index , what key(s) one would look up on? write down the entries in the permuterm index dictionary that are generated by the term ‘best’.	5
	d)	What are the different ways of placing skip pointers, How do they help? Why are skip pointers not useful for queries ‘x OR y’?	5
2.	a)	Compute the Jaccards coefficients between the query ‘bord’ and each of the terms shown below that contain the bigram ‘or’ <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid black; padding: 2px 10px;">bo</div> <div>→</div> <div style="border: 1px solid black; padding: 2px 10px;">aboard</div> <div>→</div> <div style="border: 1px solid black; padding: 2px 10px;">about</div> <div>→</div> <div style="border: 1px solid black; padding: 2px 10px;">boardroom</div> <div>→</div> <div style="border: 1px solid black; padding: 2px 10px;">border</div> </div> <div style="display: flex; justify-content: space-around; align-items: center; margin-top: 10px;"> <div style="border: 1px solid black; padding: 2px 10px;">or</div> <div>→</div> <div style="border: 1px solid black; padding: 2px 10px;">border</div> <div>→</div> <div style="border: 1px solid black; padding: 2px 10px;">lord</div> <div>→</div> <div style="border: 1px solid black; padding: 2px 10px;">morbidity</div> <div>→</div> <div style="border: 1px solid black; padding: 2px 10px;">sordid</div> </div> <div style="display: flex; justify-content: space-around; align-items: center; margin-top: 10px;"> <div style="border: 1px solid black; padding: 2px 10px;">rd</div> <div>→</div> <div style="border: 1px solid black; padding: 2px 10px;">aboard</div> <div>→</div> <div style="border: 1px solid black; padding: 2px 10px;">ardent</div> <div>→</div> <div style="border: 1px solid black; padding: 2px 10px;">boardroom</div> <div>→</div> <div style="border: 1px solid black; padding: 2px 10px;">border</div> </div>	5
	b)	Convert the tokens ‘Difficulty’ and ‘Difference’ using Soundex algorithm, write each step.	6
	c)	Give all bi grams for the query “ April is the cruellest month”	4
	d)	What is edit distance algorithm? Explain different operations involved finding the edit distance between two strings s1 and s2.	5

3.	a)	Calculate the page rank for the given graph of web links for 2 iterations. 	4																														
	b)	Let a set of documents, S be S={D1,D2,D3,D4,D5,D6} Where relevance score by user survey is given by Relevance Set, R in the same order, R={3,2,3,0,1,2}. Calculate the Cumulative Gain and Discounted Cumulative Gain(DCG).	6																														
	c)	Solve the problem using Vector Space Model (Using TF IDF weights) as applied to the collection of 4 documents Doc 1 : Information Retrieval Systems Doc 2 : Information Storage Doc 3 : Digital Speech Synthesis Systems Doc 4 : Speech Filtering, Speech Retrieval (i) Compute all non-zero entries in the normalized vector for Doc 1. (ii) Rank all the documents in the collection for the query “Speech Systems”? (iii) Compute the cosine similarities between (a) docs 1 and 2 (b) docs 3 and 4.	10																														
4.	a)	Why do we apply dimensionality reduction techniques in recommender systems? Briefly explain.	4																														
	b)	Briefly discuss the key problems in Recommender system.	4																														
	c)	Consider a matrix which shows four users <i>Alice</i> , <i>U1</i> , <i>U2</i> and <i>U3</i> rating on different news apps. The rating range is from 1 to 5 on the basis of users likability of the news app. The ‘?’ indicates that the app has not been rated by the user. Find the rating for “?” based on user-user collaborative filtering using the formula for similarity as below Similarity for any two users ‘a’ and ‘b’ can be calculated from the given formula, $Sim(a,b) = \frac{\sum_p (r_{ap}-\bar{r}_a)(r_{bp}-\bar{r}_b)}{\sqrt{\sum_p (r_{ap}-\bar{r}_a)^2} \sqrt{\sum_p (r_{bp}-\bar{r}_b)^2}}$ <i>r_{up}</i> : rating of user <i>u</i> against item <i>p</i> <i>p</i> : items <table><tr><td>Name</td><td>Inshorts(I1)</td><td>HT(I2)</td><td>NYT(I3)</td><td>TOI(I4)</td><td>BBC(I5)</td></tr><tr><td>Alice</td><td>5</td><td>4</td><td>1</td><td>4</td><td>?</td></tr><tr><td>U1</td><td>3</td><td>1</td><td>2</td><td>3</td><td>3</td></tr><tr><td>U2</td><td>4</td><td>3</td><td>4</td><td>3</td><td>5</td></tr><tr><td>U3</td><td>3</td><td>3</td><td>1</td><td>5</td><td>4</td></tr></table>	Name	Inshorts(I1)	HT(I2)	NYT(I3)	TOI(I4)	BBC(I5)	Alice	5	4	1	4	?	U1	3	1	2	3	3	U2	4	3	4	3	5	U3	3	3	1	5	4	8
Name	Inshorts(I1)	HT(I2)	NYT(I3)	TOI(I4)	BBC(I5)																												
Alice	5	4	1	4	?																												
U1	3	1	2	3	3																												
U2	4	3	4	3	5																												
U3	3	3	1	5	4																												
	d)	What is the role of SVD in movie recommender system? Briefly explain. What type of recommender system is movie recommendation? Justify your answer.	4																														
5.	a)	Explain the 3 basic elements of intelligent web applications	6																														
	b)	State any 5 fallacies of intelligent applications.	5																														
	c)	What are the different applications of Hystrix?	5																														
	d)	Explain the role of Apache Chukwa in Netflix application.	4																														

VB19CS 332

VB19 CS 332

Soundex Algorithm Chart:

Letters	Assigned Code
a, e, i, o, u, w, h, y	0
b, f, p, v	1
c, g, j, k, q, s, x, z	2
d, t	3
l	4
m, n	5
r	6