

PES University, Bengaluru (Established under Karnataka Act No. 16 of 2013)

UE19CS332

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

UE19CS332: Algorithms for Intelligence Web and Information Retrieval

| Т | ime: 3 Hrs | Answer All Questions | | | | | | |
|----|---|---|---|--|--|--|--|--|
| 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 | 5 | | | | | |
| | b) | ii) for AND NOT(drug OR Approach) Give different ways of enumerating "uni*sity", how can we enumerate all documents containing such terms? | | | | | | |
| 12 | 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'. | | | | | | |
| | 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'? | | | | | | |
| 2. | a) | Compute the Jaccards coefficients between the query 'bord' and each of the terms shown below that contain the bigram 'or' | 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 cruelest month" | 4 | | | | | |
| | d) What is edit distance algorithm? Explain different operations involved finding the distance between two strings s1 and s2. | | | | | | | |

| | a) | Calculate the page rank for the given graph of web links for 2 iterations. | | | | | | | |
|----|---|---|--|---|--|--|--|----|---|
| | | | | | | | | | |
| | 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. | a) | Why do we apply dimensionality reduction techniques in recommender systems? Briefly explain. | | | | | | | |
| | b) | Briefly discus | ss the key proble | ems in Reco | ommender sy | stem. | | | 4 |
| | | 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, $Sim(a,b) = \frac{\sum_p (r_{ap} - \bar{r}_a)(r_{ab} - \bar{r}_b)}{\sqrt{\sum (r_{ap} - \bar{r}_a)^2} \sqrt{\sum (r_{bp} - \bar{r}_b)^2}}$ $r_{up} : rating \ of \ user \ u \ against \ item \ p$ $p : items$ | | | | | | | |
| | | on user-user of | collaborative filting two users $lpha$ an r_a | ering using $Sim(a, rac{1}{2})$ | the formula calculated from | for similarit in the given fo | ty as below rmula, | | |
| | | on user-user of | collaborative filting two users $lpha$ an r_a | ering using $Sim(a, rac{1}{2})$ | the formula calculated from | for similarit in the given fo | ty as below rmula, | 2 | |
| | | on user-user of Similarity for a | collaborative filt iny two users 'a' ar r_i r_i p | fering using $Sim(a, s_{ip}: rating : items$ | the formula calculated from $b) = rac{\Sigma}{\sqrt{\Sigma}(g \ of \ user \ s)}$ | for similarity the given for $\mathbb{C}_p(r_{ap}-ar{r}_a)(r_{ap}-ar{r}_a)^2\sqrt{u}$ against | ty as below r_{ab} $-ar{r}_b$) $\sum (r_{bp} - ar{r}_b)$ $item~p$ | 2 | |
| | | on user-user of Similarity for a | collaborative filting two users a an r_a p | Sim(a, sp: rating : items | the formula calculated from $b)=rac{\Sigma}{\sqrt{\Sigma}(g\ of\ user\ c}$ | for similarity the given for $C_p(r_{ap}-ar{r}_a)(r_{ap}-ar{r}_a)^2\sqrt{u}$ against | ty as below r_{mula} , r_{ab} – $ar{r}_b$) $\sum (r_{bp}$ – $ar{r}_b)$ $item p$ | 2 | |
| | | On user-user of Similarity for a | collaborative filting two users a an r_a p | ering using and branche Sim(a, sp: rating : items | the formula calculated from $b)=rac{\Sigma}{\sqrt{\Sigma}(g\ of\ user\)}$ | for similarity the given for $C_p(r_{ap}-ar{r}_a)(\sqrt{r_{ap}-ar{r}_a})^2\sqrt{u}$ against | ty as below rmula, $r_{ab} - ar{r}_b)$ $\sum (r_{bp} - ar{r}_b)$ $item\ p$ | 2 | |
| | | On user-user of Similarity for a Name Auce | collaborative filting two users a and r_i p Inshorts(I1) | ering using and branch be a Sim(a, sp.: rating : items HT(12) | the formula calculated from $b)=rac{\Sigma}{\sqrt{\Sigma}(g\ of\ user\)}$ | for similarity the given for $C_p(r_{ap}-\bar{r}_a)(r_{ap}-\bar{r}_a)^2\sqrt{against}$ and $C_p(14)$ | ty as below rmula, $r_{ab} - \bar{r}_b)$ $\sum (r_{bp} - \bar{r}_b)$ $item\ p$ | 2 | |
| | d) | Name Auce U1 U2 U3 | collaborative filting two users a and r_a p Inshorts(I1) 5 | ering using and branche of Sim(a, sp.: rating: items HT(12) | the formula calculated from $b = \frac{\sum}{\sqrt{\sum}(g \ of \ user \)}$ NYT(13) | for similarity the given for $C_p(r_{ap}-\bar{r}_a)(r_{ap}-\bar{r}_a)^2\sqrt{against}$ and $C_p(r_{ap}-\bar{r}_a)^2\sqrt{against}$ | ty as below rmula, $r_{ab} = r_b$) $\sum (r_{bp} = r_b)$ $item p$ BBC(15) 2 3 5 | 2 | 4 |
| 5. | d) a) | Name Alice U1 U2 U3 | collaborative filting two users a and a are collaborative filting two users a and a and a are collaborative filting a and a are | ering using and branche of Sim(a, Sim(a, p): rating : items HT(12) 4 1 3 anovie recomme | the formula calculated from $b = \frac{\Sigma}{\sqrt{\Sigma}} (g \ of \ user \ s)$ NYT(13) 1 2 4 1 2 4 1 2 4 1 2 4 1 | for similarity the given for $C_p(r_{ap}-\bar{r}_a)(\sqrt{r_{ap}-\bar{r}_a})^2\sqrt{u}$ against $C_p(14)$ $C_$ | ty as below rmula, $r_{ab} = r_b$) $\sum (r_{bp} = r_b)$ $item p$ BBC(15) 2 3 5 | 2 | 4 |
| 5. | | Name Alice U1 U2 U3 What is the rerecommender Explain the 3 | collaborative filting two users 'a' and provided the collaborative filting two users 'a' and provided the collaboration for the collaboration filting the collaborative filtin | sering using of b'can be of sim(a, sp: rating: items HT(12) 4 1 3 3 aovie recomme of intelliger | the formula calculated from b = $\frac{\Sigma}{\sqrt{\Sigma}}$ (g of user a) NYT(13) 1 2 4 2 4 4 | for similarity the given for $C_p(r_{ap}-\bar{r}_a)(\sqrt{r_{ap}-\bar{r}_a})^2\sqrt{u}$ against $C_p(14)$ $C_$ | ty as below rmula, $r_{ab} = r_b$) $\sum (r_{bp} = r_b)$ $item p$ BBC(15) 2 3 5 | 2 | |
| 5. | a) | Name Auce U1 U2 U3 What is the recommended Explain the 3 State any 5 fa | ry two users 'a' and pole of SVD in mer system is moved basic elements | sering using of b'can be of Sim(a, sp.: rating : items HT(12) 4 1 3 avovie recomme of intelligent applic | the formula calculated from b) = $\frac{\Sigma}{\sqrt{\Sigma}}$ (g of user a) NYT(13) 1 2 4 2 4 2 4 2 4 2 4 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 4 | for similarity the given for $C_p(r_{ap}-\bar{r}_a)(\sqrt{r_{ap}-\bar{r}_a})^2\sqrt{u}$ against $C_p(14)$ $C_$ | ty as below rmula, $r_{ab} = r_b$) $\sum (r_{bp} = r_b)$ $item p$ BBC(15) 2 3 5 | 2 | 6 |

UB1905332 UB1903332

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 |
| 1 | 4. |
| m, n | 5 |
| r | 6 |