|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **National University of Computer and Emerging Sciences, Lahore Campus** | | | | |
| C:\Users\saif\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.Word\final design.jpg | **Course:** | **Information Retrieval** | **Course Code:** | **CS317** |
| **Program:** | **BS(Computer Science)** | **Semester:** | **Fall 2019** |
| **Duration:** | **20 Minutes** | **Total Marks:** | **10** |
| **Paper Date:** | **3-Dec-19** | **Weight** | **4%** |
| **Section:** | **A** | **Page(s):** | **2** |
| **Exam:** | **Quiz 5** | **Roll No:** |  |

**Question 1:**

Q3) a) Show the different steps of HAC algorithm using the distance matrix below.

Give partial results after each step. Calculate Similarity of Clusters using Complete Link. [4 Marks]

****

**b)** Show final results using dendogram [2 Marks]

**Solution:**

**First Iteration: Doc 4 and 5 have smallest distance so that are closest clusters. Merge 4 and 5.**

**Since it is complete link similarity so similarity of doc 1 with new merged cluster 4-5 is max of the distance of (8 and 10).**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4-5 |
| **1** | **0** |  |  |  |
| **2** | **2** | **0** |  |  |
| **3** | **4** | **3** | **0** |  |
| **4-5** | **10** | **7** | **9** | **0** |

**Second Iteration**

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1-2 | 3 | 4-5 |
| **1-2** | **0** |  |  |
| **3** | **4** | **0** |  |
| **4-5** | **10** | **9** | **0** |

**Third Iteration**

|  |  |  |
| --- | --- | --- |
|  | 1-2-3 | 4-5 |
| **1-2-3** | **0** |  |
| **4-5** | **10** | **0** |

**Question 2:**

Query = <0, 2, 0, 1, 0, 0>

+ D1 = < 3, 5, 0 , 1, 3, 2 >

- D1 = < 1, 2, 5 , 5, 0, 0 >

- D1 = < 0, 1, 1 , 4, 1, 0 >

- D1 = < 2, 1, 2 , 2, 3, 1 >

+ D1 = < 2, 4, 1 , 2, 2, 1 >

Find modified query vector using Rocchio feedback algorithm. + represents relevant and – represents nonrelevant document. α = 1.5, β = 0.5, ϒ = 0.1 [4 Marks]

**Solution:**

**Relevant doc centroid = (1 / 2) < 3+2 , 5+4, 0+1, 1+2, 3+2, 2+1 > = <2.5, 4.5, 0.5, 1.5, 2.5, 1.5>**

**NonRel doc centroid = (1 / 3)** < 1+0+2, 2+1+1, 5+1+2 , 5+4+2, 0+1+3, 0+0+1 > = <1, 4/3, 8/3, 11/3, 4/3, 1/3>

**Modified query vector = 1.5 \* (**<0, 2, 0, 1, 0, 0>**) + 0.5 \* (<2.5, 4.5, 0.5, 1.5, 2.5, 1.5>) – 0.1\*(**<1, 4/3, 8/3, 11/3, 4/3, 1/3>**)**