

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Programme | : | **B.Tech** | Semester | : | **Win Sem 21-22** |
| Course | : | **Web Mining** | Code | : | **CSE3024** |
| Faculty | : | **Dr.Bhuvaneswari A** | Slot | : | **L7+L8** |
| Date | : | **08-02-2022** | Marks | : | **10 Points** |

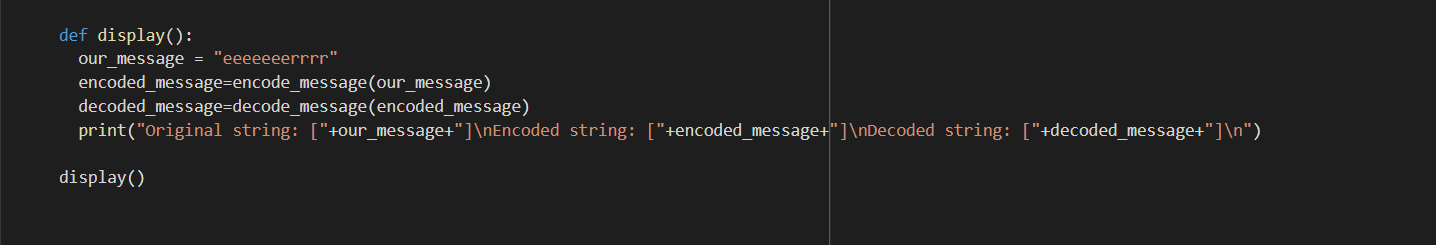
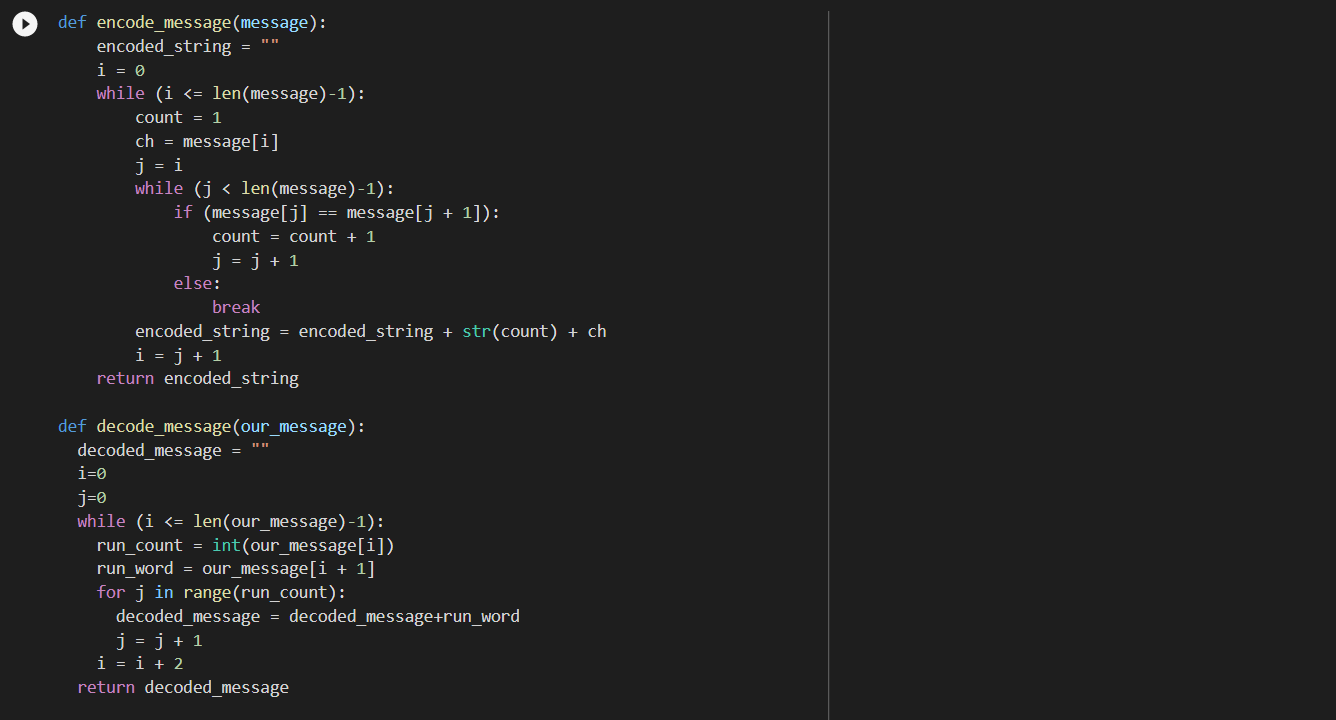
Name: RAHUL GARG  
Register Number: 19BCE1431

COLLAB LINK**: https://colab.research.google.com/drive/1PRXsdn9sKLHL1RRFtQdMH1yA7TRqhalF#scrollTo=P2C7B35LGs7M**

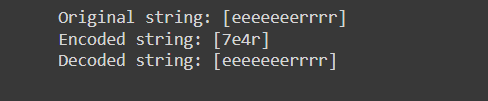
**Exercise 5:Index Compression Techniques**

1. Apply run length encoding for the following string and compress it.

“eeeeeeerrrrrrrrrrrrrrrrttttttttttttttiiiiiiiifffffeft”

**CODE:** 

**OUPUT:**

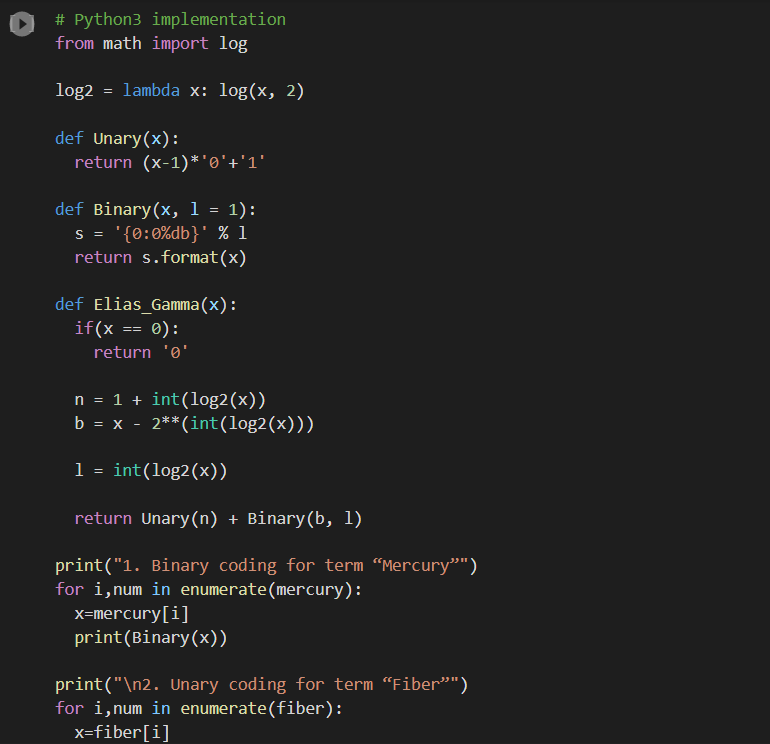
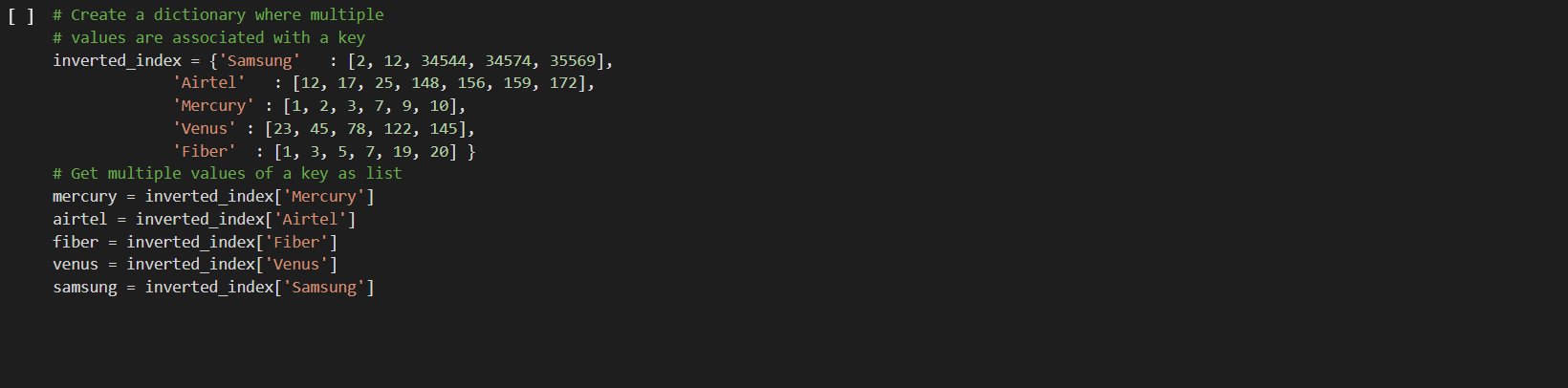


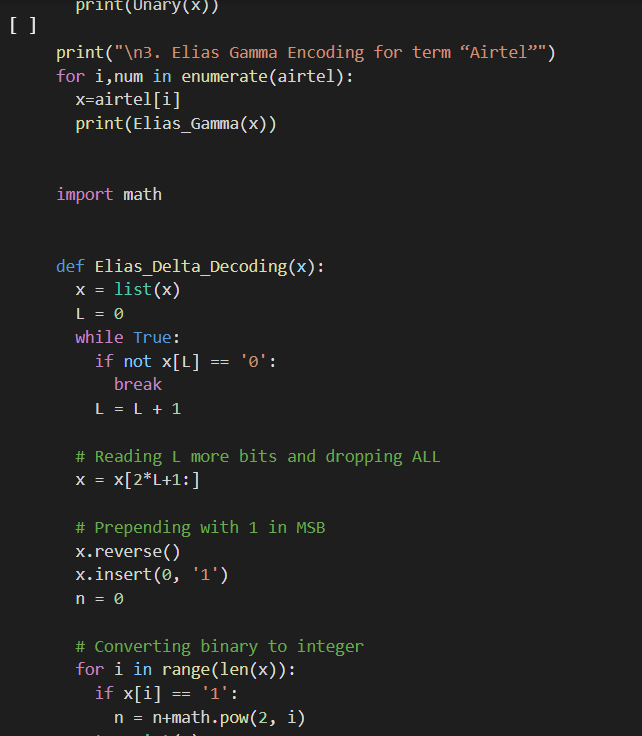
1. Consider the following Inverted Index File with Terms, Occurrences and Posting List

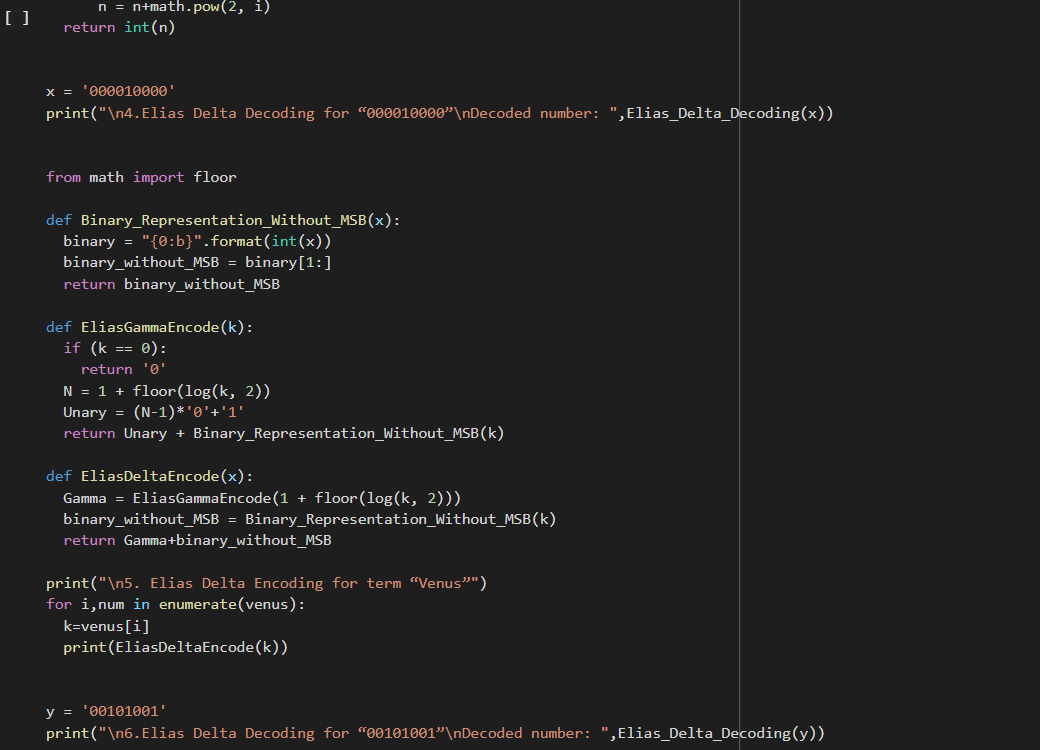
|  |  |  |
| --- | --- | --- |
| **Term** | **Occurrences** | **Posting List (Doc ids)** |
| Samsung | 233 | 2, 12, 34544, 34574, 35569, … |
| Airtel | 12 | 12, 17, 25, 148, 156, 159, 172, … |
| Mercury | 15 | 1, 2, 3, 7, 9, 10, … |
| Venus | 12 | 23, 45, 78, 122, 145, … |
| Fiber | 6 | 1, 3, 5, 7, 19, 20 |

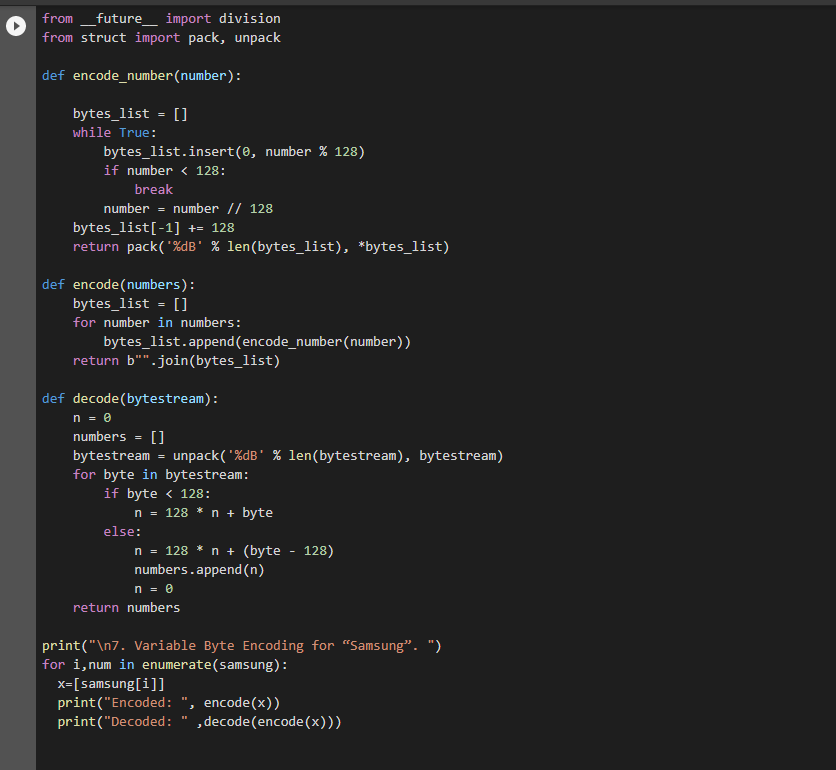
1. Apply Binary coding for term “Mercury” (apply for all doc ids)
2. Apply Unary coding for term “Fiber”
3. Apply Elias Gamma Encoding for term “Airtel”
4. Apply Elias Delta Decoding for “000010000”
5. Apply Elias Delta Encoding for term “Venus”
6. Apply Elias Delta Decoding for “00101001”
7. Apply Variable Byte Encoding for “Samsung”. (Use doc ids gap)

**CODE:**









**OUTPUT:**

