Exercise 6 Indexing

Given the following term/document table where values in table are actual term frequencies in each document:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Terms** | math | physics | computer | cpu | memory | disk | cache |
|  |  |  |  |  |  |  |  |
| Doc1 | 2 | 1 | 0 | 0 | 4 | 5 | 8 |
| Doc2 | 0 | 0 | 0 | 5 | 5 | 0 | 0 |
| Doc3 | 2 | 2 | 2 | 2 | 4 | 6 | 2 |
| Doc4 | 0 | 3 | 0 | 3 | 4 | 0 | 0 |
| Doc5 | 3 | 12 | 0 | 3 | 6 | 6 | 0 |

1. Rewrite term/document matrix normalizing all documents by dividing by total number of words in each document

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Terms** | math | physics | computer | cpu | memory | disk | cache |
|  |  |  |  |  |  |  |  |
| Doc1 |  |  |  |  |  |  |  |
| Doc2 |  |  |  |  |  |  |  |
| Doc3 |  |  |  |  |  |  |  |
| Doc4 |  |  |  |  |  |  |  |
| Doc5 |  |  |  |  |  |  |  |

1. Using original matrix calculate the IDF for Computer, Memory and cache showing values. Then modify original matrix using those to calculate weights.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Terms** | math | physics | computer | cpu | memory | disk | cache |
|  |  |  |  |  |  |  |  |
| Doc1 |  |  |  |  |  |  |  |
| Doc2 |  |  |  |  |  |  |  |
| Doc3 |  |  |  |  |  |  |  |
| Doc4 |  |  |  |  |  |  |  |
| Doc5 |  |  |  |  |  |  |  |

1. Using original matrix calculate the Signal weight for physics, memory and cache showing each value. Then update matrix using those values to calculate weights.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1. **Terms** | math | physics | computer | cpu | memory | disk | cache |
|  |  |  |  |  |  |  |  |
| Doc1 |  |  |  |  |  |  |  |
| Doc2 |  |  |  |  |  |  |  |
| Doc3 |  |  |  |  |  |  |  |
| Doc4 |  |  |  |  |  |  |  |
| Doc5 |  |  |  |  |  |  |  |