|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Document | **User A** | **TF** | **IDF** | **tf\*idf** |
| 013 | 4 | 0.0540540540540541 | 2.39794000867204 | 0.129618378847137 |
| 098 | 5 | 0.238095238095238 | 2.39794000867204 | 0.570938097302866 |
| 126 | 2 | 0.0454545454545455 | 2.39794000867204 | 0.108997273121456 |
| 135 | 1 | 0.0263157894736842 | 2.39794000867204 | 0.0631036844387378 |
| 234 | 3 | 0.142857142857143 | 2.39794000867204 | 0.34256285838172 |
| 287 | 4 | 0.153846153846154 | 2.39794000867204 | 0.368913847488006 |
| 301 | 2 | 0.05 | 2.39794000867204 | 0.119897000433602 |
| 322 | 2 | 0.0434782608695652 | 2.39794000867204 | 0.10425826124661 |
| 323 | 4 | 0.142857142857143 | 2.39794000867204 | 0.34256285838172 |
| 324 | 4 | 0.210526315789474 | 2.39794000867204 | 0.504829475509903 |
| 433 | 1 | 0.0256410256410256 | 2.39794000867204 | 0.061485641248001 |

If we were to rank them based upon the tf\*idf, the ranking of the documents for User A, would be:

098, 324, 287, 234, 323, 013, 301, 126, 322, 135, 433.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Document | **User B** | **TF** | **IDF** | **tf\*idf** |
| 013 | 4 | 0.0540540540540541 | 2.49485002168009 | 0.134856757928654 |
| 098 | 3 | 0.142857142857143 | 2.49485002168009 | 0.356407145954299 |
| 126 | 2 | 0.0454545454545455 | 2.49485002168009 | 0.113402273712732 |
| 135 | 1 | 0.0263157894736842 | 2.49485002168009 | 0.0656539479389498 |
| 234 | 3 | 0.142857142857143 | 2.49485002168009 | 0.356407145954299 |
| 287 | 3 | 0.115384615384615 | 2.49485002168009 | 0.287867310193857 |
| 301 | 2 | 0.05 | 2.49485002168009 | 0.124742501084005 |
| 322 | 2 | 0.0434782608695652 | 2.49485002168009 | 0.108471740073048 |
| 323 | 3 | 0.107142857142857 | 2.49485002168009 | 0.267305359465724 |
| 324 | 3 | 0.157894736842105 | 2.49485002168009 | 0.393923687633699 |
| 433 | 1 | 0.0256410256410256 | 2.49485002168009 | 0.0639705133764127 |

If we were to rank them based upon the tf\*idf, the ranking of the documents for User A, would be:

324, 098, 234, 287, 323, 013, 301, 126, 322, 433, 135.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Document | **User C** | **TF** | **IDF** | **tf\*idf** |
| 013 | 4 | 0.0540540540540541 | 2.65757731917779 | 0.143652828063665 |
| 098 | 1 | 0.0476190476190476 | 2.65757731917779 | 0.126551300913228 |
| 126 | 2 | 0.0454545454545455 | 2.65757731917779 | 0.120798969053536 |
| 135 | 1 | 0.0263157894736842 | 2.65757731917779 | 0.0699362452415209 |
| 234 | 2 | 0.0952380952380952 | 2.65757731917779 | 0.253102601826456 |
| 287 | 1 | 0.0384615384615385 | 2.65757731917779 | 0.102214512276069 |
| 301 | 2 | 0.05 | 2.65757731917779 | 0.13287886595889 |
| 322 | 2 | 0.0434782608695652 | 2.65757731917779 | 0.115546839964252 |
| 323 | 2 | 0.0714285714285714 | 2.65757731917779 | 0.189826951369842 |
| 324 | 1 | 0.0526315789473684 | 2.65757731917779 | 0.139872490483042 |
| 433 | 1 | 0.0256410256410256 | 2.65757731917779 | 0.068143008184046 |

If we were to rank them based upon the tf\*idf, the ranking of the documents for User A, would be:

234, 323, 013, 424, 301, 098, 126, 322, 287, 135, 433

**Problem 2**

Precision is the fraction of retrieved documents that are relevant.

Precision A = 2 / 11

Precision B = 1 / 11

Precision C = 5 / 11

Recall is the fraction of the documents that are relevant to the query that are successfully retrieved.

Recall A = 2 / 2 = 1

Recall B = ¼

Recall C = ½

**Problem 3**

The list for user A will be the best list, and they will be most satisfied because all of the relevant documents in the system. While they all might not be near the top of the list, they are there and the user can decide whether or not the returned list is relevant with some ease.

**Problem 4**

Step 1:

Terms: {Tiger(s), North, Carolina}

Documents:

013: (4, 0, 0)

098: (1, 2, 2)

126: (2, 0, 0)

135: (1, 0, 0)

235: (2, 1, 1)

287: (1, 2, 1)

301: (2, 0, 0)

322: (2, 0, 0)

323: (2, 1, 1)

324: (1, 2, 1)

433: (1, 0, 0)

Step 2:

It shows that there is a high importance in the term, “Tiger” and “Carolina” but less so about “North.”

Step 3:

Query: {(2, 0.5, 1.2)}

Documents:

Doc 013:

Doc 098:

Doc 126:

Doc 135:

Doc 235:

Doc 287:

Doc 301:

Doc 322:

Doc 323:

Doc 324:

Doc 433:

Step 4:

126, 287, 234, 098, 013, 135, 301, 322, 433, 235, 323

Step 5:

All of the relevant documents appear within the first three results of User A’s query, same as the original search. This is a significant improvement from the original search, where only one of the relevant documents was within the top 3 results.