1. A

Define idf as log(N/df)

|  |  |  |
| --- | --- | --- |
| Term | DF | IDF |
| Car | 18.165 | 0.74 |
| auto | 6723 | 1.17 |
| insurance | 19241 | 0.716 |
| best | 25235 | 0.596 |

B

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Term | Doc1 tf-idf | Norm(doc1) | Doc2 tf-idf | Norm(doc2) | Doc3 tf-idf | Norm(doc3) |
| L-2 Norm | 2.8034 |  | 3.655 |  | 2.824 |  |
| Car | 1.7992 | 0.6418 | 1.1855 | 0.32435 | 1.761 | 0.62358 |
| Auto | 1.728 | 0.6164 | 2.95 | 0.8071 | 0 | 0 |
| Insurance | 0 | 0 | 1.803 | 0.49329 | 1.763 | 0.62429 |
| best | 1.279 | 0.5989156 | 0 | 0 | 1.329 | 0.4706 |

C.

Doc1,Doc2 = 0.6418\*0.32435+0.6164\*0.8071 = 0.70566427

Doc1,Doc3 = 0.6418\*0.62358 +0.5989156\*0.4706 = 0.682

Doc2,Doc3 = 0.32435\*0.62358 + 0.49329\*0.62429 = 0.5102

2.

Doc1: <https://en.wikipedia.org/wiki/Database_normalization>

Doc2: <https://www.guru99.com/database-normalization.html>

Doc3: <https://www.studytonight.com/dbms/database-normalization.php>

tf

|  |  |  |  |
| --- | --- | --- | --- |
| Term | Doc1 | Doc2 | Doc3 |
| normal | 46 | 40 | 36 |
| form | 17 | 6 | 21 |

Idf

|  |  |  |
| --- | --- | --- |
| Term | DF | IDF |
| Normal | 4250000000 | 4.37 |
| form | 4660000000 | 4.33 |

Tf-idf

|  |  |  |  |
| --- | --- | --- | --- |
| Term | Doc1 | Doc2 | Doc3 |
| Normal | 11.636 | 11.371 | 11.171 |
| Form | 9.658 | 7.699 | 10.055 |

Vector normalization

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Term | Doc1 | Norm(doc1) | Doc2 | Norm(doc2) | Doc3 | Norm(doc3) |
| L-2Norm | 15.122 |  | 13.732 |  | 15.03 |  |
| Normal | 11.636 | 0.76947 | 11.371 | 0.82806 | 11.171 | 0.74324 |
| Form | 9.658 | 0.63867 | 7.699 | 0.56066 | 10.055 | 0.66899 |

Query

|  |  |  |
| --- | --- | --- |
| Term | Q tf-idf | Norm(Q) |
| Normal | 11.171 | 0.74324 |
| form | 7.699 | 0.51224 |

Consine similarity:

Doc1,Q = 0.76947\*0.74324 + 0.63867\*0.51224 = 0.899

Doc2,Q = 0.82806\*0.74324 + 0.56066\*0.51224 = 0.9026

Doc3,Q = 0.74324\*0.74324 + 0.66899\*0.51224 = 0.895