

HW 3 Part 2

Q1a)

TF Matrix

	car	auto	insurance	locst
d1	1	5/30	0	15/30
d2	5/35	1	1	0
d3	20/25	0	1	15/25

$$IDF(\text{car}) = \log_2(3/3) = 0$$

$$IDF(\text{auto}) = \log_2(3/2) = -0.585$$

$$IDF(\text{insur}) = \log_2(3/2) = -0.585$$

$$IDF(\text{locst}) = \log_2(3/2) = -0.585$$

	car	auto	insurance	locst
d1	0	0.0975	0	0.2925
d2	0	0.585	0.585	0
d3	0	0	0.585	0.351

Q7(b)

Query vector  $\rightarrow$  "Car Insurance"

$$\frac{1}{2} [1 \times 0, 0 \times 0, 1 \times 0.585, 0 \times 0]$$

$$= [0, 0, 0.293, 0]$$

$$\text{Distance} = \sqrt{\sum_{i=1}^n (q_i - b_i)^2}$$

$$d1 \rightarrow \sqrt{0 + (0.0975)^2 + (0.293)^2 + (0.02925)^2}$$

$$\rightarrow \sqrt{0.1809115}$$

$$\boxed{\rightarrow 0.4253}$$

$$d2 \rightarrow \sqrt{0^2 + (0.585)^2 + (0.293 - 0.585)^2 + 0^2}$$

$$\rightarrow \sqrt{0.428074}$$

$$\boxed{\rightarrow 0.6541}$$

$$d_3 \rightarrow \sqrt{0^2 + 0^2 + (.293 - .513)^2 + (.351)^2}$$
$$\rightarrow \sqrt{0.20909}$$
$$\boxed{\rightarrow 0.4572}$$

Vector space model order

$d_1, d_3, d_2$

Q1c)

Cosine similarity  $\frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|}$

Query magnitude =  $\sqrt{0.213^2 + 0^2 + 0^2 + 0^2}$   
= 0.213

d1  $\rightarrow$  dot product first

$$0 \times 0 + 0.08 \times 0 + 0 \times 0.213 + .584 \times 0 = 0$$

d1 magnitude =  $\sqrt{0^2 + 0.0975^2 + 0^2 + 0.2925^2} = 0.3083$

Cosine similarity = 0

d2  $\rightarrow$   $0 \times 0 + 0.505 \times 0 + 0.885 \times 0.213 + 0 \times 0 = -0.1711$

d2 magnitude =  $\sqrt{0^2 + 0.505^2 + 0.885^2 + 0^2} = 0.8273$

Cosine similarity  $\frac{0.171105}{0.8273 \times 0.213} = 0.706$

$$d3 \rightarrow -0.885 - 0.293 + 0^2 + 0^2 \rightarrow \text{dot product} = 0.171105$$

$$\begin{aligned} d3 \text{ magnitude} &= \sqrt{0^2 + 0^2 + 0.885^2 + 0.381^2} \\ &= 0.6823 \end{aligned}$$

$$\text{cosine similarity} = \frac{0.171105}{0.6823 \times 0.293}$$

$$= 0.856$$

order would be d3, d2, d1

Q2a) 32 retrieved documents

Q2b) 18 total relevant documents

Q2c)

$$\text{Precision} = \frac{TP}{TP + FP} \rightarrow \frac{10}{32} \rightarrow 0.3125$$

$$\text{Recall} = \frac{TP}{TP + FN} \rightarrow \frac{10}{18} \rightarrow 0.5556$$

$$\text{Accuracy} = \frac{TP + TN}{\text{Total Docs}} \rightarrow \frac{70}{100} \rightarrow .7$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$= 2 \times \frac{0.3125 \times 0.5556}{0.3125 + 0.5556} = 0.4$$

Q3)

## Reciprocal Ranks

$$\text{Query 1: } 1/2 \rightarrow 0.5$$

$$\text{Query 2: } 1/3 \rightarrow 0.333$$

$$\text{Query 3: } 1/4 \rightarrow 0.25$$

$$\text{Query 4: } 1/1 \rightarrow 1$$

$$MRR = \frac{1}{N} \sum_{i=1}^N \frac{1}{\text{rank}_i}$$

$$= \frac{1}{4} (0.5 + 0.3333 + 0.25 + 1)$$

$$= \frac{1}{4} (2.0838)$$

$$= 0.5208$$

Q4a)

$$\text{Precision } \text{AR} = \frac{\text{Num relevant docs in top } k}{k}$$

$$k=1 \rightarrow 1$$

$$k=2 \rightarrow 7/2$$

$$k=3 \rightarrow 2/3$$

$$k=4 \rightarrow 2/9 \rightarrow 1/2$$

$$k=5 \rightarrow 2/5$$

$$k=6 \rightarrow 2/6$$

$$k=7 \rightarrow 2/7$$

$$k=8 \rightarrow 3/8$$

$$k=9 \rightarrow 4/9$$

$$k=10 \rightarrow 4/10$$

$$k=11 \rightarrow 4/11$$

$$k=12 \rightarrow 9/12$$

$$k=13 \rightarrow 9/13$$

$$k=14 \rightarrow 9/14$$

$$k=15 \rightarrow 5/15 \rightarrow 1/3$$

$$k=16 \rightarrow 5/16$$

$$k=17 \rightarrow 5/17$$

$$k=18 \rightarrow 5/18$$

$$k=19 \rightarrow 5/19$$

$$k=20 \rightarrow 6/20$$

Q9b)

$$\text{Recall } @ k = \frac{\text{num relevant docs } @ k}{\text{total # relevant docs}}$$

$$k=1 \rightarrow \frac{1}{8}$$

$$k=2 \rightarrow \frac{1}{8}$$

$$k=3 \rightarrow \frac{2}{8} \rightarrow \frac{1}{4}$$

$$k=4 \rightarrow \frac{1}{4}$$

$$k=5 \rightarrow \frac{1}{4}$$

$$k=6 \rightarrow \frac{1}{4}$$

$$k=7 \rightarrow \frac{1}{4}$$

$$k=8 \rightarrow \frac{3}{8}$$

$$k=9 \rightarrow \frac{4}{8} \rightarrow \frac{1}{2}$$

$$k=10 \rightarrow \frac{1}{2}$$

$$k=11 \rightarrow \frac{1}{2}$$

$$l_z=12 \rightarrow 1/2$$

$$l_z=13 \rightarrow 1/2$$

$$l_z=14 \rightarrow 1/2$$

$$l_z=15 \rightarrow 5/8$$

$$l_z=16 \rightarrow 5/8$$

$$l_z=17 \rightarrow 5/8$$

$$l_z=18 \rightarrow 5/8$$

$$l_z=19 \rightarrow 5/8$$

$$l_z=20 \rightarrow 6/8$$

Q9c)

$$F1@k = 2 \times \frac{\text{(Precision} \times \text{Recall})@k}{(\text{Precision} + \text{Recall})@k}$$

$$F1@1 = 2 \times \frac{1 \times .125}{1 + .125} \rightarrow 0.222$$

$$F1@2 = 2 \times \frac{0.5 \times .125}{.5 + .125} \rightarrow 0.200$$

$$F1@3 = 2 \times \frac{0.667 \times .250}{.667 + .250} \Rightarrow 0.364$$

$$F1@4 = 2 \times \frac{0.8 \times .250}{.8 + .250} \rightarrow 0.333$$

$$F1@5 = 2 \times \frac{0.9 \times .250}{0.9 + .250} \rightarrow 0.384$$

$$F1@6 = 2 \times \frac{0.333 \times .250}{.333 + .250} \rightarrow 0.286$$

$$F1@7 = 2 \times \frac{0.286 \times .250}{.286 + .250} \Rightarrow 0.267$$

$$F1 @ 8 = 2x \frac{.375 \times .375}{.375 + .375} \rightarrow 0.375$$

$$F1 @ 9 = 2x \frac{-444 \times .5}{.444 + .5} \rightarrow 0.471$$

$$F1 @ 10 = 2x \frac{.4 \times .5}{.9 + .5} \rightarrow 0.444$$

$$F1 @ 11 = 2x \frac{-364 \times .5}{-364 + .5} \rightarrow 0.421$$

$$F1 @ 12 = 2x \frac{-333 \times .5}{.333 + .5} \rightarrow 0.400$$

$$F1 @ 13 = 2x \frac{.308 \times .5}{.308 + .5} \rightarrow 0.381$$

$$F1 @ 14 = 2x \frac{.286 \times .5}{-.286 + .5} \rightarrow 0.364$$

$$F1 @ 15 = 2x \frac{-337 \times .625}{.837 + .625} \rightarrow 0.435$$

$$F1 \textcircled{Q} 16 = 2x \frac{-312x \cdot 0.625}{.312 + 0.625} \rightarrow 0.417$$

$$F1 \textcircled{Q} 17 = 2x \frac{0.299x - 625}{.299 + 0.625} \rightarrow 0.400$$

$$F1 \textcircled{Q} 18 = 2x \frac{-278x \cdot 625}{-278 + 625} \rightarrow 0.385$$

$$F1 \textcircled{Q} 19 = 2x \frac{.263x \cdot 625}{.263 + 625} \rightarrow 0.370$$

$$F1 \textcircled{Q} 20 = 2x \frac{0.3 * .750}{0.3 + .750} \rightarrow 0.429$$

Q9d)

$$AP = \frac{\sum \text{Precision}@k \text{ of relevant doc positions}}{\text{total number relevant docs}}$$

Sum of precision @ k

$$\rightarrow 2 + .667 + .375 + .949 + .237 + .3 = 3.119$$

thus

$$AP = \frac{3.119}{8} \rightarrow 0.39$$