Written Homework for Vector Space Model

-2016358 25 김성원-

Query

- "fake novel"

Documents

- doc-1: "twitter fake"
- doc-2: "concert fake news"
- doc-3: "clancy news"
- doc-4: "fake picaso"

TF

	twitter	fake	concert	news	clancy	picaso
Doc-1	1	1	0	0	0	0
Doc-2	0	1	1	1	0	0
Doc-3	0	0	0	1	1	0
Doc-4	0	1	0	0	0	1

IDF

twitter	$\log_{10}(4/1) = 0.602$
fake	$\log_{10}(4/3) = 0.124$
concert	$\log_{10}(4/1) = 0.602$
news	$\log_{10}(4/2) = 0.301$
clancy	$\log_{10}(4/1) = 0.602$
picaso	$\log_{10}(4/1) = 0.602$

TF × IDF

	twitter	fake	concert	news	clancy	picaso
Doc-1	0.602	0.124	0	0	0	0
Doc-2	0	0.124	0.602	0.301	0	0
Doc-3	0	0	0	0.301	0.602	0
Doc-4	0	0.124	0	0	0	0.602

TF-IDF of the query

	twitter	fake	concert	news	clancy	picaso
q	0	(3/3) × 0.124 = 0.124	0	0	0	0

Length of Documents

length of Doc-1 $= \sqrt{(0.602)^2 + (0.124)^2} = 0.614$

length of Doc-2 $= \sqrt{(0.124)^2 + (0.602)^2 + (0.301)^2} = 0.684$

length of Doc-3 $= \sqrt{(0.301)^2 + (0.602)^2} = 0.673$

length of Doc-4 $= \sqrt{(0.124)^2 + (0.602)^2} = 0.614$

length of $q = \sqrt{(0.124)^2} = 0.124$

Similarity Values

$$\text{CosSim}(\text{Doc-1}, q) = \frac{(0\times0.602) + (0.124\times0.124) + (0\times0) + (0\times0) + (0\times0) + (0\times0)}{(0.614 \times 0.124)}$$

$$= \frac{0.0153}{0.076} = 0.201$$

$$\text{CosSim}(\text{Doc-2}, q) = \frac{(0\times0) + (0.124\times0.124) + (0\times0.602) + (0\times0.301) + (0\times0) + (0\times0)}{(0.684 \times 0.124)}$$

$$= \frac{0.0153}{0.084} = 0.182$$

$$\text{CosSim}(\text{Doc-3}, q) = \frac{(0\times0) + (0\times0.124) + (0\times0) + (0\times0.301) + (0\times0.602) + (0\times0)}{(0.673 \times 0.124)}$$

$$= 0$$

$$\text{CosSim}(\text{Doc-4}, q) = \frac{(0\times0) + (0.124\times0.124) + (0\times0) + (0\times0) + (0\times0) + (0\times0.602)}{(0.614 \times 0.124)}$$

$$= \frac{0.0153}{0.076} = 0.201$$

Rank list

Doc-1, Doc-4

Doc-2

Doc-3