Stopwords: a, as, had, of, was

Document 1: Mary had a little lamb, little lamb, little lamb.

Document 2: Mary had a little lamb.

Document 3: Whose fleece was white as snow.

Query: white lamb of mary

INDEXING

		mary	little	lamb	whose	fleece	white	snow
idf_i		0.58	0.58	0.58	1.58	1.58	1.58	1.58
$tf_{i,j}$	D1	0.33	1.0	1.0	0	0	0	0
$tf_{i,j}$	D2	1.0	1.0	1.0	0	0	0	0
$tf_{i,j}$	D3	0	0	0	1.0	1.0	1.0	1.0

VECTORS

$w_{i,j}$	D1	0.19	0.58	0.58	0	0	0	0
$w_{i,j}$	D2	0.58	0.58	0.58	0	0	0	0
$w_{i,j}$	D3	0	0	0	1.58	1.58	1.58	1.58

VECTOR LENGTHS

$$|\vec{d}_{D1}| = \sqrt{0.19^2 + 0.58^2 + 0.58^2} = 0.84$$

$$|\vec{d}_{D2}| = \sqrt{0.58^2 + 0.58^2 + 0.58^2} = 1.0$$

$$|\vec{d}_{D3}| = \sqrt{1.58^2 + 1.58^2 + 1.58^2 + 1.58^2} = 3.16$$

QUERY VECTOR

	mary	little	lamb	whose	fleece	white	snow
$tf_{i,q}$	1	0	1	0	0	1	0
$w_{i,q}$	0.58	0	0.58	0	0	1.58	0

$$|\vec{q}| = \sqrt{0.58^2 + 0.58^2 + 1.58^2} = 1.78$$

$$\frac{\vec{d}_{D1} \cdot \vec{q}}{|\vec{d}_j| \times |\vec{q}|} = \frac{(0.19 \times 0.58) + (0.58 \times 0) + (0.58 \times 0.58) + (0 \times 0) + (0 \times 0) + (0 \times 1.58) + (0 \times 0)}{0.84 \times 1.78} = 0.3$$

or ... iterate though the query's non-zero dimensions.

$$\frac{\vec{d}_{D2} \cdot \vec{q}}{|\vec{d}_{D2}| \times |\vec{q}|} = \frac{(0.58 \times 0.58) + (0.58 \times 0.58) + (1.58 \times 0)}{1 \times 1.78} = 0.38$$

$$\frac{\vec{d}_{D3} \cdot \vec{q}}{|\vec{d}_{D3}| \times |\vec{q}|} = 0.44$$

$sim(\vec{d}_{D3}, \vec{q})$	0.46	final ranking.
$sim(\vec{d}_{D2}, \vec{q})$	0.38	
$sim(\vec{d}_{D1}, \vec{q})$	0.30	

Workings

N: no. of documents = 3

$$idf_i = \log_2\left(\frac{N}{n_i}\right)$$

$$idf_{mary} = \log_2\left(\frac{3}{2}\right) = 0.58$$

$$idf_{whose} = \log_2\left(\frac{3}{1}\right) = 1.58$$

$$tf_{i,j} = \frac{freq_{i,j}}{maxfreq_j}$$

$$tf_{mary,D1} = \frac{1}{3} = 0.33$$

$$tf_{little,D1} = \frac{3}{3} = 1.0$$

$$tf_{mary,D2} = \frac{1}{1} = 1.0$$

$$|\vec{d_j}| = \sqrt{\sum_{i=1}^{T} w_{i,j}^2}$$

$$sim(\vec{d_j}, \vec{q}) = \frac{\vec{d_j} \cdot \vec{q}}{|\vec{d_j}| \times |\vec{q}|}$$

$$\vec{d_j} \cdot \vec{q} = \sum_{i=1}^{T} w_{i,j} \times w_{i,q}$$