TF-IDF :

Q: איזה יום היום

1: היום שימשי מאוד בחוץ.

2: היום יום חמישי.

3: ! איזה יום נעים היום

4: היום יום הולדת ליונתן.

Q	היום	יום	איזה	#words
1.	1	0	0	4
2.	1	1	0	3
3.	1	1	1	4
4.	1	1	0	4
#Doc	4	3	1	

TFIDF

For a term i in document j:

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ = number of occurrences of i in j
df_i = number of documents containing i
N = total number of documents

	Tf Idf score
1.	(1/4)*log(4/4) = 0
2.	(1/3)*log(4/4) + (1/3)*log(4/3) = 0.138
3.	(1/4)*log(4/4) + (1/4)*log(4/3) + (1/4)*log(4) = **0.603**
4.	(1/4)*log(4/4) + (1/4)*log(4/3) = 0.103

the third sentence is the most compatible result according to TF IDF.

Source for the formula : http://www.cnblogs.com/youth0826/archive/2012/08/11/2633688.html