Evaluation Plan

Precision will capture the fraction of relevant information out of retrieved information. The recall will calculate the fraction of relevant information that was retrieved. To cover both aspects of the performance, F-Measure can be used as a Harmonic Mean of precision and recall. However, these metrics are single-valued and it is based on the entire document returned by the system. We need to evaluate based on the ranking of the articles suggested. We want to display only the relevant information and hence precision would be a better measure than recall [1].

Precision@K: This metric refers to relevant results from top K documents. However, it does not take into consideration the placement of top K documents.

R-precision: This metric can be used if all documents relevant to a particular query are known. R-precision requires knowing all documents that are relevant to a query. R refers to the number of relevant documents and it is used as the cutoff for calculation [1].

Mean average precision: This metric calculates the mean of the average precision scores for each query, where Q is the number of queries.

$$\mathrm{MAP} = \frac{\sum_{q=1}^{Q} \mathrm{AveP(q)}}{Q}$$

Precision@K:

Query	P@6	P@7	P@8
Fever	4/6=0.66	4/7=0.57	5/8
Common cold	⅚=0.833	6/7=0.85	6/8
covid-19	4/6	5/7=0.71	5/8
cancer	4/6	4/7	4/7
Heart disease	4/6	5/7	6/8
Abdominal pain	3/6=0.5	5/7	5/8

Recall@K:

Query	P@6	P@7	P@8
Fever	3/6 =0.5	4/5	4/5
Common cold	4/6=0.66	5/6	5/6
covid-19	⅚=0.83	6/7	7/7
cancer	4/5	4/5	4/5
Heart disease	4/6	5/7	6/7
Abdominal pain	3/6	4/5	4/5

Mean Average Precision:

Query	Calculation	Mean Average Precision
Fever	1/5(2/3 + 2/5 + 3/6 + 3/4 + 5/10)	0.562
Common cold	1/6(3/4 + 2/3 + 3/5 + 4/7 + 5/7 + 6/10)	0.53
covid-19	1/7(1/2 + 2/3 + 3/4 + 4/6 + 5/7 + 5/8 + 6/10)	1/7(0.5+0.66+0.75+ 0.66+0.71+0.62+0.6) = 0.64
cancer	1/5(1 + 2/4 + 3/6 + 4/8 + 4/9)	⅕(1+0.5+0.5+0.5+0. 44)= 0.58
Heart disease	1/5(2/3 + 3/5 + 3/6 + 4/7 + 2/4)	⅕(0.66+0.6+0.5+0.5 7+0.5)=0.566
Abdominal pain	1/6(1/2 + 3/4 + 3/5 + 5/7 + 4/8+1)	⅙(0.5+0.75+0.6+0.7 1+0.5+1)=0.67

References:

1. Parikshit Sondhi, Jimeng Sun, Hanghang Tong, ChengXiang Zhai, “SympGraph: A Framework for Mining Clinical Notes through Symptom Relation Graphs”, ResearchGate, 2012.