

Q1. F measure, P@5, Avg Precision

relevant
total/docs = 15

R N R N R N N N N N R N R N N
1 3 5 10 12

Q3. What is weakness in laplace smoothing? why interpolation methods are better for documents?

Q2. Don't want to miss any citation.

Answers:

Q1. OF measure: precision = $\frac{\text{no. of relevant docs retrieved}}{\text{total docs}} = \frac{5}{14} = 0.357$

recall = $\frac{\text{no. of rel docs retrieved}}{\text{total relevant docs}} = \frac{5}{15} = \frac{1}{3} = 0.333$

$$F_{\text{measure}} = \frac{2 \times P \times R}{P + R} = \frac{2 \left(\frac{\frac{5}{14} \times \frac{5}{15}}{\frac{5}{14} + \frac{5}{15}} \right)}{29} = \frac{10}{29} = 0.3448$$

① = 0.345

② $P@5 = \frac{3}{5} = \boxed{0.6}$ — ①

③ Avg precision:

$$P@1 = \frac{1}{1} = 1; \quad P@3 = \frac{2}{3} = 0.67; \quad P@5 = \frac{3}{5} = 0.6$$

$$P@10 = \frac{4}{10} = 0.4; \quad P@12 = \frac{5}{12} = 0.417$$

$$\text{Avg precision} = \frac{\text{precisions at relevant docs}}{\text{total relevant docs}}$$

$$= \frac{1 + 0.67 + 0.6 + 0.4 + 0.417}{5}$$

$$= \boxed{3.08}$$

②

Date:

~~retrieval~~

Q2. To get all citations, recall must be 1 as that search engine will retrieve all the relevant documents. In the graph, both search engines are giving the same recalls at every precision cut off point and both are reaching recall of 1. Hence both the search engines will give all citations to the scientist. So they can use any of the search engines.

②

Q3. Laplace smoothing deals with each of the missing words in the same way, regardless of how significant or insignificant the word might be. Interpolation methods perform smoothing while taking relevance and lengths of documents into account and so they are a more descriptive method, and deal with all the words in our query with regards to the background probability as well.

②