

Web Information Retrieval

Assignment 2

Team Name : Gamma

Members

| Name | Matriculation Number |
|-----------------------------|-----------------------------|
| Mohammad Nizam Uddin | 216101140 |
| Md. Shohel Ahamad | 216203438 |
| MD Jakaria Nawaz | 216203442 |
| Shreya Chatterjee | 216100848 |

Calculation of Evaluation Measures

Answer:

Result set - 1

1. Search Result, $Aq = \{d11, d12, d13, d14, d15, d16, d17\}$
Relevant Result, $|G| = 4$
Relevant in Search Result, $Arq = \{d12, d13, d15\}$
 $|Aq \cap G| = 3$
 $Aq \cup G = \{d11, d12, d13, d14, d15, d16, d17\}$

$$\text{Precision, } P = |Aq \cap G| / |Aq|$$

$$= 3/7$$
$$= 0.428$$

2. Recall, $r = |Aq \cap G| / |G|$

$$= 3/4$$
$$= 0.75$$

3. F- measure, $F1 = 2rP/(P+r)$

$$= 2 \times 0.75 \times 0.428 / (0.428 + 0.75)$$

$$= 0.642 / 1.178$$

$$= 0.544$$

Result set - 2

1. Search Result, $Aq = \{d21, d22, d23, d24, d25, d26, d27, d28\}$
Relevant Result, $|G| = 8$
 $|Aq \cap G| = 4$
 $|Aq \cup G| = 12$

$$\text{Precision, } P = |Aq \cap G| / |Aq|$$

$$= 4/8$$

$$= 0.5$$

$$2. \quad \text{Recall, } r = |Aq \cap G| / |G|$$

$$= 4/8$$

$$= 0.5$$

$$3. \quad \text{F- measure, } F1 = 2rP/(P+r)$$

$$= 2 \times 0.5 \times 0.5 / (0.5 + 0.5)$$

$$= 0.5 / 1$$

$$= 0.5$$

Result set - 3

=====

$$1. \quad \text{Search Result, } Aq = \{d31, d32, d33, d34\}$$

$$\text{Relevant Result, } |G| = 1$$

$$|Aq \cap G| = 1$$

$$|Aq \cup G| = 4$$

$$\text{Precision, } P = |Aq \cap G| / |Aq|$$

$$= 1/4$$

$$= 0.25$$

$$2. \quad \text{Recall, } r = |Aq \cap G| / |G|$$

$$= 1/1$$

$$= 1$$

$$3. \quad \text{F- measure, } F1 = 2rP/(P+r)$$

$$= 2 \times 1 \times 0.25 / (0.25 + 1)$$

$$= 0.5 / 1.25$$

$$= 0.4$$

Precision at k ($P@k$)

=====

Result set - 1 :

The *precision at k* ($p@k$) for cutoff values of $k = 1, k = 5$

| K | # Relevant docs | P@K |
|---|-----------------|-----|
| 1 | 0 | 0 |
| 5 | 3 | 0.6 |

Result set - 2 :

The *precision at k* ($p@k$) for cutoff values of $k = 1, k = 5$

| K | # Relevant docs | P@K |
|---|-----------------|-----|
| 1 | 1 | 1 |
| 5 | 2 | 0.4 |

Result set - 3 :

The *precision at k* ($p@k$) for cutoff values of $k = 1, k = 5$

| K | # Relevant docs | P@K |
|---|-----------------|-----|
| 1 | 0 | 0 |
| 5 | 1 | 0.2 |

R-Precision

=====

| Result set | $ G = A $ | # Relevant docs | R-Precision ($p@ G $) |
|------------|-------------|-----------------|-------------------------|
| 1 | 4 | 2 | 0.5 |
| 2 | 8 | 4 | 0.5 |
| 3 | 1 | 0 | 0 |

2. F-Measure

1. The $F\beta$ - score for arbitrary values reflects both the precision and the recall of the test to compute the score. The number of correct definite outcomes divided by the number of all positive outcomes, the number of correct actual outcomes divided by the number of positive outcomes that should have been returned. The $F\beta$ score can be interpreted as a weighted average of the precision and recall, where a $F\beta$ score reaches its best value at 1 and worst at 0.
2. As, here β is a determinant expressing how many times more we value recall than precision. Using, $\beta = 1$ gives the balanced F-score which is corresponding value of $\alpha=1/2$, where both are equally relevant. That is why the choice of $\beta = 1$ is most obvious.
3. Higher values of β , put more weight on precision. Lower values of β which is closer to 0, put more weight on recall. In that case if $F1/3$ ($\beta=1/3$) it is offering a less value of F less than $F0.50$, then the α value will be $1/6$ and the recall will increase compare to precision.