# Information Retrivel: Assignment 3

**Shifei Chen**

---

## Exercise 8.8

### a

In the first system we have $|Q| = 1$ and $m_j = 4$. There were 4 relevant results and the 4 $R_{jk}$s are

$$
\begin{aligned}
R_{j1} &= R \\
R_{j2} &= RNR \\
R_{j3} &= RNRNNNNR \\
R_{j4} &= RNRNNNNRR
\end{aligned}
$$

which leads to their precisions

$$
\begin{aligned}
Precison(R_{j1}) &= 1/1 = 1 \\
Precison(R_{j2}) &= 2/3 = 0.667 \\
Precison(R_{j3}) &= 3/9 = 0.333 \\
Precison(R_{j4}) &= 4/10 = 0.4
\end{aligned}
$$

Then plugging the numbers into the formula

$$MAP(Q) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} \frac{1}{m_j} \sum_{k=1} m_j Precision(R_{jk})$$

, we have $1/4 * (1 + 0.667 + 0.333 + 0.4) = 0.6$

The second system we have $1/4 * (1/2 + 2/5 + 3/6 + 4/7) = 0.493$. So system 1 has a higher MAP value.

### b

The result from part a makes sense as system 1 hit the relevant document in the first query. From the formula we can also see that to get the maxium score it's important to hit all relevant docuemnts as early as possible, hence the user can get his information in the first or second lines from the result list.

## 8.10

### a

| | **Judge 2 Yes** | **Judge 2 No** | **Total** |
|---|---|---|---|
| **Judge 1 Yes** | 2 | 4 | 6 |
| **Judge 1 No** | 4 | 2 | 6 |
| **Total** | 6 | 6 | 12 |

$$P(A) = (2+2)/12 = 0.333$$
$$P(relevant) = (6+6)/(12+12) = 0.5$$
$$P(nonrelevan) = (6+6)/(12+12) = 0.5$$
$$P(E) = P(n)^2 + P(r)^2 = 0.5$$
$$Kappa = \frac{P(A) - P(E)}{1 - P(E)} = -0.333$$

**b**

| | **Relevant** | **Nonrevalant** |
|---|---|---|
| **Retrived** | 1 | 4 |
| **Not Retrived** | 1 | 6 |

$$precision = 1/(1+4) = 0.2$$
$$recall = 1/(1+1) = 0.5$$
$$F1 = 2 * 0.2 * 0.5/(0.2+0.5) = 0.286$$

**c**

| | **Relevant** | **Nonrevalant** |
|---|---|---|
| **Retrived** | 5 | 0 |
| **Not Retrived** | 5 | 2 |

$$precision = 5/(5+0) = 1$$
$$recall = 5/(5+5) = 0.5$$
$$F1 = 2 * 1 * 0.5/(1+0.5) = 0.667$$