

Task 1

0.1 Subtask 1

$$w_{t,d} = (1 + \log tf_{t,d}) * \log \frac{N}{df_t}$$

- $w_{pens,d_1} = (1 + \log 1) * \log \frac{3}{1} \approx 1, 1$
- $w_{pens,d_2} = \text{not in the doc}$
- $w_{pens,d_3} = \text{not in the doc}$
- $w_{write,d_1} = (1 + \log 1) * \log \frac{3}{3} = 0$
- $w_{write,d_2} = (1 + \log 1) * \log \frac{3}{3} = 0$
- $w_{write,d_3} = (1 + \log 1) * \log \frac{3}{3} = 0$
- $w_{on,d_1} = (1 + \log 1) * \log \frac{3}{3} = 0$
- $w_{on,d_2} = (1 + \log 1) * \log \frac{3}{3} = 0$
- $w_{on,d_3} = (1 + \log 1) * \log \frac{3}{3} = 0$
- $w_{paper,d_1} = (1 + \log 2) * \log \frac{3}{3} = 0$
- $w_{paper,d_2} = \text{not in the doc}$
- $w_{paper,d_3} = (1 + \log 1) * \log \frac{3}{3} = 0$
- $w_{pencils,d_1} = \text{not in the doc}$
- $w_{pencils,d_2} = (1 + \log 1) * \log \frac{3}{1} \approx 1, 1$
- $w_{pencils,d_3} = \text{not in the doc}$
- $w_{envelope,d_1} = \text{not in the doc}$
- $w_{envelope,d_2} = (1 + \log 1) * \log \frac{3}{1} \approx 1, 1$
- $w_{envelope,d_3} = \text{not in the doc}$
- $w_{ballpens,d_1} = \text{not in the doc}$
- $w_{ballpens,d_2} = \text{not in the doc}$
- $w_{ballpens,d_3} = (1 + \log 1) * \log \frac{3}{1} \approx 1, 1$

Subtask 2

Task 3

- $Precision = \frac{TP_s}{TP_s + FP_s}$
- $Recall = \frac{TP_s}{TP_s + FN_s}$

k	Result Set	Precision	Recall
1	127	1.0	0.2
2	127, 9	0.5	0.2
3	127, 9, 10	0.33	0.2
4	127, 9, 10, 2	0.5	0.4
5	127, 9, 10, 2, 35	0.4	0.4
6	127, 9, 10, 2, 35, 32	0.33	0.4
7	127, 9, 10, 2, 35, 32, 41	0.43	0.6
8	127, 9, 10, 2, 35, 32, 41, 64	0.5	0.8

