*Alberto Saponaro - saponaroalberto97@gmail.com*
*Walter Väth - walter.vaeth@gmail.com*
*Chong Shen - st143575@stud.uni-stuttgart.de*
*Xin Pang - st145113@stud.uni-stuttgart.de*

# Task 1

## Subtask 1

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

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

*Alberto Saponaro - saponaroalberto97@gmail.com*
*Walter Väth - walter.vaeth@gmail.com*
*Chong Shen - st143575@stud.uni-stuttgart.de*
*Xin Pang - st145113@stud.uni-stuttgart.de*

| **Terms** | $d_1$ | $d_2$ | $d_3$ |
|---|---|---|---|
| pens | 1.1 | 0.0 | 0.0 |
| write | 0.0 | 0.0 | 0.0 |
| on | 0.0 | 0.0 | 0.0 |
| paper | 0.0 | 0.0 | 0.0 |
| pencils | 0.0 | 1.1 | 0.0 |
| envelope | 0.0 | 1.1 | 0.0 |
| ballpens | 0.0 | 0.0 | 1.1 |

$\overrightarrow{d_1} = (1.1, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0)$
$\overrightarrow{d_2} = (0.0, 0.0, 0.0, 0.0, 1.1, 1.1, 0.0)$
$\overrightarrow{d_3} = (0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.1)$

## Subtask 2

$w_{ballpens,q} = (1 + \log 1) * \log \frac{3}{1} \approx 1,1$
$w_{envelope,q} = (1 + \log 1) * \log \frac{3}{1} \approx 1,1$

$\overrightarrow{q} = (0.0, 0.0, 0.0, 0.0, 0.0, 1.1, 1.1)$

$SIM(\overrightarrow{q}, \overrightarrow{d_1}) = \frac{\overrightarrow{q} * \overrightarrow{d_1}}{|\overrightarrow{q}| * |\overrightarrow{d_1}|} = 0$
$SIM(\overrightarrow{q}, \overrightarrow{d_2}) = \frac{\overrightarrow{q} * \overrightarrow{d_2}}{|\overrightarrow{q}| * |\overrightarrow{d_2}|} = \frac{1.21}{49} \approx 0.025$
$SIM(\overrightarrow{q}, \overrightarrow{d_3}) = \frac{\overrightarrow{q} * \overrightarrow{d_3}}{|\overrightarrow{q}| * |\overrightarrow{d_3}|} = \frac{1.21}{49} \approx 0.025$

| **Rank** | **Doc** | **SIM** |
|---|---|---|
| 1 | 2 | 0.025 |
| 1 | 3 | 0.025 |
| 2 | 1 | 0 |

*Alberto Saponaro - saponaroalberto97@gmail.com*
*Walter Väth - walter.vaeth@gmail.com*
*Chong Shen - st143575@stud.uni-stuttgart.de*
*Xin Pang - st145113@stud.uni-stuttgart.de*

## Task 3

- $Precision = \frac{TPs}{TPs+FPs}$
- $Recall = \frac{TPs}{TPs+FNs}$

| 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 |