

## TDA : indexation

### exercice 1 :

1) mesure simple :  $f \times idf = f \times \frac{1}{df(t,d)}$

•  $poids("fait", d) = 1 \times \frac{1}{10^3} = 10^{-3}$

•  $poids("beau", d) = 1 \times \frac{1}{10^3} = 10^{-3}$

•  $poids("soin", d) = 1 \times \frac{1}{10^3} = 10^{-3}$

2) mesure basé sur le logarithme :

$$f \times \log_{10} \left( \frac{|D|}{df} \right)$$

$$poids("fait", d) = 1 \times \log_{10} \left( \frac{10000}{1000} \right)$$

$$= 1 \times \log_{10}(10) = \underline{1}$$

$$poids("beau", d) = 1 \times \log_{10} \left( \frac{10000}{1000} \right) = \underline{1}$$

$$poids("soin", d) = 1 \times \log_{10} \left( \frac{10000}{1000} \right) = \underline{1}$$

### exercice 2 :

$$|D| = 64$$

$$df("domaine", D) = 16 ; f("domaine", d) = 3$$

$$df("théorique", D) = 8 ; f("théorique", d) = 2$$

$$df("informatique", D) = 4 ; f("informatique", d) = 4$$

mesure simple:

$$\text{poids}(t, d) = \frac{tf \times 1}{idf(t, d)}$$

t

$$\text{poids}(\text{"domaine"}, d) = 3 \times \frac{1}{16} = \frac{3}{16}$$

forest

$$\text{poids}(\text{"théorique"}, d) = 2 \times \frac{1}{8} = \frac{1}{4}$$

anima

$$\text{poids}(\text{"informatique"}, d) = 4 \times \frac{1}{4} = 1$$

livr

abri

sauva

num

mesure basé sur le logarithme:

$$\text{poids}(t, d) = tf \times \log_{10} \left( \frac{100}{idf(t, d)} \right)$$

$$\text{poids}(\text{"domaine"}, d) = 3 \times \log_{10} \left( \frac{64}{16} \right)$$

$$= 3 \times \log_{10}(4)$$

$$\text{poids}(\text{"théorique"}, d) = 2 \times \log_{10} \left( \frac{64}{8} \right)$$

$$= 2 \times \log_{10}(8)$$

$$\text{poids}(\text{"théorique informatique"}, d) = 4 \times \log_{10} \left( \frac{64}{4} \right)$$

$$= 4 \log_{10}(16)$$



NB: l'extraction sous forme de racines.

le thésaurus: espèce = animal

humain = naturaliste

Sauvage = dangereux

1) matrice de fréquence

	d1	d2	d3	d4	d5	d6	d7	<del>total</del>	rang
forêt	1	0	1	0	0	0	0	2	3
animal	1	1	0	1	0	1	1	5	1
lion	0	0	0	1	1	0	0	2	3
arbre	0	0	1	0	0	0	0	1	4
Sauvage	0	2	0	0	0	0	1	3	2
humain	0	0	0	0	1	0	1	2	3

~~total~~

matrice de présence :

t \ d	d <sub>1</sub>	d <sub>2</sub>	d <sub>3</sub>	d <sub>4</sub>	d <sub>5</sub>	d <sub>6</sub>	d <sub>7</sub>
foret	1	0	1	0	0	0	0
animal	1	1	0	1	0	1	1
lion	0	0	0	1	1	0	0
arbre	0	0	1	0	0	0	0
Sauvage	0	1	0	0	0	0	1
humain	0	0	0	0	1	0	1



# matrice de poids

t \ d	d1	d2	d3	d4	d5	d6	d7
forêt	0,54	0	0,54	0	0	0	0
animal	0,14	0,14	0	0,14	0	0,14	0
lion	0	0	0	0,54	0,54	0	0
arbre	0	0	0,54	0	0	0	0
sauvage	0	1,08	0	0	0	0	0,54
humain	0	0	0	0	0,54	0	0,54

2/

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

=

$$\begin{pmatrix} 0 & 2 & 2 \\ 1 & 2 & 2 \\ 0 & 1 & 0 \end{pmatrix}$$

classement des termes selon le "tf" :

~~forêt~~ ~~lion~~ ~~humain~~ ~~arbre~~

1 - animal

2 - Sauvage

3 - forêt

lion

humain

4 - arbre

Après élagage, les mots extraits sont :

("sauvage", "forêt", "lion", "humain", "arbre")

exercice 5 :

1) matrice de présence  $\Rightarrow$  matrice de présence

$$\begin{matrix} & d_1 & d_2 & d_3 \\ t_1 & \begin{pmatrix} 2 & 2 & 0 \end{pmatrix} \\ t_2 & \begin{pmatrix} 3 & 2 & 4 \end{pmatrix} \\ t_3 & \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \end{matrix} \Rightarrow \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$