

# Solution des deux exercices du chapitre 4 : modèles de base de RI

## 1. Solution de l'exercice du modèle vectoriel :

### Collection :

D1 : { langage de programmation python est très utilisé pour le traitement de texte }

D2 : { le langage JAVA est basé sur le langage C++ }

D3 : { un langage de programmation est un langage utilisé pour traduire un algorithme en un programme }

Motsvides : { de, est, très, pour, le, un, en }

La requête q : { langage python java }

### 1.1. Le fichier inverse avec la formule :

$$\text{poids}(t_i, d_j) = (\text{freq}(t_i, d_j) / \max(\text{freq}(d_j))) * \log(N/n_i + 1)$$

#### Exemple de calcul :

poids(langage, D1) = (1/1)\*log(3/3 + 1) = 0.301

poids(langage, D2) = (2/2)\*log(3/3 + 1) = 0.301

poids(programmation, D1) = (1/1)\*log(3/2 + 1) = 0.397

..... etc.

Le fichier inverse sera alors :

Terme \ doc	D1	D2	D3
langage	0.301	0.301	0.301
programmation	0.397		0.198
python	0.602		
utilisé	0.397		0.198
traitement	0.602		
texte	0.602		
java		0.301	
basé		0.301	
c++		0.301	
traduire			0.301
algorithme			0.301
programme			0.301

## 1.2. Calcul de similarité par la modèle vectoriel pour la requête :

q : {langage python java}

Terme \ q	q
langage	1
programmation	0
python	1
utilisé	0
traitement	0
texte	0
java	1
basé	0
c++	0
traduire	0
algorithme	0
programme	0

### 1.2.1. Par produit interne :

$$RSV(d_j, q) = \sum w_{ij} * w_{iq}$$

$$RSV(D1, q) = 0.301 + 0.602 + 0 = \mathbf{0.903}$$

$$RSV(D2, q) = 0.301 + 0 + 0.301 = \mathbf{0.602}$$

$$RSV(D3, q) = 0.301 + 0 + 0 = \mathbf{0.301}$$

### 1.2.2. Par coef de Dice :

$$RSV(d_j, q) = \frac{2 * \sum w_{ij} * w_{iq}}{\sum w_{ij}^2 + \sum w_{iq}^2}$$

$$RSV(D1, q) = (2 * 0.903) / (0.301^2 + 0.397^2 + 0.602^2 + 0.397^2 + 0.602^2 + 0.602^2 + 1^2 + 1^2 + 1^2)$$

$$= \mathbf{0.401}$$

$$RSV(D2, q) = (2 * 0.602) / (0.301^2 + 0.301^2 + 0.301^2 + 0.301^2 + 1^2 + 1^2 + 1^2)$$

$$= \mathbf{0.358}$$

$$RSV(D3, q) = (2 * 0.301) / (0.301^2 + 0.301^2 + 0.301^2 + 0.301^2 + 0.198^2 + 0.198^2 + 1^2 + 1^2 + 1^2)$$

$$= \mathbf{0.174}$$

### 1.2.3. Par cosinus :

$$RSV(d_j, q) = \frac{\sum w_{ij} * w_{iq}}{\sqrt{\sum w_{ij}^2 * \sum w_{iq}^2}}$$

$$RSV(D1, q) = 0.903 / \sqrt{(0.301^2 + 0.397^2 + 0.602^2 + 0.397^2 + 0.602^2 + 0.602^2) * (1^2 + 1^2 + 1^2)}$$
$$= 0.426$$

$$RSV(D2, q) = 0.602 / \sqrt{(0.301^2 + 0.301^2 + 0.301^2 + 0.301^2) * (1^2 + 1^2 + 1^2)}$$
$$= 0.577$$

$$RSV(D3, q) = 0.301 / \sqrt{(0.301^2 + 0.301^2 + 0.301^2 + 0.301^2 + 0.198^2 + 0.198^2) * (1^2 + 1^2 + 1^2)}$$
$$= 0.261$$

### 1.2.4. Par Jaccard :

$$RSV(d_j, q) = \frac{\sum w_{ij} * w_{iq}}{\sum w_{ij}^2 + \sum w_{iq}^2 - \sum w_{ij} * w_{iq}}$$

$$RSV(D1, Q) = 0.903 / (0.301^2 + 0.397^2 + 0.602^2 + 0.397^2 + 0.602^2 + 0.602^2 + 1^2 + 1^2 + 1^2 - 0.903)$$
$$= 0.251$$

$$RSV(D2, Q) = 0.602 / (0.301^2 + 0.301^2 + 0.301^2 + 0.301^2 + 1^2 + 1^2 + 1^2 - 0.602)$$
$$= 0.218$$

$$RSV(D3, Q) = 0.301 / (0.301^2 + 0.301^2 + 0.301^2 + 0.301^2 + 0.198^2 + 0.198^2 + 1^2 + 1^2 + 1^2 - 0.301)$$
$$= 0.095$$

## 2. Solution de l'exercice 2 (extension du modèle booléen):

Ensemble des termes d'indexation = (document, web, information, recherche, image, contenu).

d1=(document 1, web 0,5)

q1 = (document OU web)

q2= (web ET document)

q3= ((web OU document) ET image)

Calculer la similarité entre d1 et chaque requête, par le modèle booléen basé sur les ensembles flous, par le modèle booléen étendu, et par le modèle p\_norme avec p=2.

## 2.1. Avec le modèle booléen basé sur les ensembles flous

$$RSV(d_j, t_i) = W_{ij}$$

$$RSV(d_j, t_1 \wedge t_2) = \min(RSV(d_j, t_1), RSV(d_j, t_2))$$

$$RSV(d_j, t_1 \vee t_2) = \max(RSV(d_j, t_1), RSV(d_j, t_2))$$

$$RSV(d_j, \text{not } t_i) = 1 - RSV(d_j, t_i)$$

Pour q1 :  $RSV(d1, q1) = \max(\text{poids}(\text{document}, d1), \text{poids}(\text{web}, d1)) = 1$

Pour q2 :  $RSV(d1, q2) = \min(\text{poids}(\text{web}, d1), \text{poids}(\text{document}, d1)) = 0.5$

Pour q3 :  $RSV(d1, Q3) = \min(\max(\text{poids}(\text{web}, d1), \text{poids}(\text{document}, d1)), \text{poids}(\text{image}, d1)) = \min(\max(0.5, 1), 0) = \min(1, 0) = 0.$

## 2.2. Avec le modèle booléen étendu

$$RSV(d_j, t_1 \vee t_2) = \frac{\sqrt{(w_{1j}^2 + w_{2j}^2)}}{\sqrt{2}}$$

$$RSV(d_j, t_1 \wedge t_2) = 1 - \frac{\sqrt{((1 - w_{1j})^2 + (1 - w_{2j})^2)}}{\sqrt{2}}$$

$$RSV(d_j, \text{qnot}) = 1 - RSV(d_j, q)$$

Pour q1 :  $RSV(d1, q1) = \text{racine}(1^2 + 0.5^2) / (\text{racine}(2)) = 0.79$

Pour q2 :  $RSV(d1, q2) = 1 - \text{racine}((1-1)^2 + (1-0.5)^2) / \text{racine}(2) = 0.64$

Pour q3 :  $RSV(d1, q3) = RSV1(\text{web ou document}) \text{ et image.}$

$$RSV1 = 0.79 \Rightarrow RSV(d1, q3) = 1 - \text{racine}((1 - 0.79)^2 + (1 - 0)^2) / \text{racine}(2) = 0.27$$

## 2.3. Avec le modèle P-norme (avec p=2) :

$$RSV(d_j, qor) = \left( \frac{w_{1j}^p + w_{2j}^p + \dots + w_{mj}^p}{m} \right)^{1/p}$$

$$RSV(d_j, qand) = 1 - \left( \frac{(1 - w_{1j})^p + (1 - w_{2j})^p + \dots + (1 - w_{mj})^p}{m} \right)^{1/p}$$

$$RSV(d_j, \text{qnot}) = 1 - RSV(d_j, q)$$

Pour q1 :  $RSV(d1, q1) = ((1^2 + 0.5^2)/2)^{1/2} = 0.79$

Pour q2 :  $RSV(d1, q2) = 1 - (((1-1)^2 + (1-0.5)^2)/2)^{1/2} = 0.64$

Pour q3 :  $RSV(d1, q3) = RSV(0.79 \text{ et image}) = 1 - (((1-0.79)^2 + (1-0)^2)/2)^{1/2} = 0.27$