TD 2

Exo 1 :

- Soit la collection (C) suivante :

D_1 = "été soleil brille température augmente"

D_2 = "hiver neige tombe température diminue"

D_3 = " soleil fait fondre neige".

D_4 = " montagne neige tombe été".

- Considérant le modèle booléen, répondre aux requête suiv.

R_{t_i} = "ensbl de documents, qui contient le terme t_i"

$R_{été} = \{D_1, D_4\}$

$R_{soleil} = \{D_1, D_3\}$

$R_{température} = \{D_1, D_2\}$

$R_{neige} = \{D_2, D_3, D_4\}$

$R_{montagne} = \{D_4\}$

q_1 = soleil $\wedge$ neige

$R_{q_1} = R_{soleil} \cap R_{neige} = \{D_1, D_3\} \cap \{D_2, D_3, D_4\} = \{D_3\}$

q_2 = température $\vee$ été

$R_{q_2} = R_{température} \cup R_{été} = \{D_1, D_2\} \cup \{D_1, D_4\} = \{D_1, D_2, D_4\}$

q_3 = neige $\wedge \neg$ (montagne $\wedge$ été)

$R_{q_3} = R_{neige} \wedge \neg (R_{montagne} \cap R_{été}) = \{D_2, D_3, D_4\} \wedge \neg (\{D_4\} \cap \{D_1, D_4\})$

$= \{D_2, D_3, D_4\} \cap (\{D_4\}) = \{D_2, D_3, D_4\} \cap \{D_1, D_2, D_3\}$

$= \{D_2, D_3\}$

EX02 :

-Avec la même collection. Sans considérant le modèle Victoriel

1) Construire la matrice des Poids doc-termes. N=4

	été	soleil	brille	temp	augmente	hiver	neige	tombe	clim	fait	fond	Montag
n_i	2	2	1	2	1	1	3	2	1	1	1	1
idf_i	1	1	2	1	2	2	0,141	1	2	2	2	2
tf_{i1}	1	1	1	1	1	0	0	0	0	0	0	0
tf_{i2}	0	0	0	1	0	1	1	1	1	0	0	0
tf_{i3}	0	1	0	0	0	0	1	0	0	1	1	0
tf_{i4}	0	0	0	0	0	0	1	1	0	0	0	1
W_{i1}	1	1	2	1	2	0	0	0	0	0	0	0
W_{i2}	0	0	0	1	0	2	0,141	1	2	0	0	0
W_{i3}	0	1	0	0	0	0	0,141	0	0	2	2	0
W_{i4}	1	0	0	0	0	0	0,141	1	0	0	0	2
W_q		1					1				1	1

2) Réponde à la requête en Etilisant les mesures de similarité :

- Le produit scalaire : $Score_{PS}(d_j, q) = \sum_{i=1}^{n} W_{ij} \times q$

$Score_{PS}(d_1, q) = (1 \times 1) + (1 \times 0) + (1 \times 0) + (1 \times 0) = 1$

$Score_{PS}(d_2, q) = (1 \times 0) + (1 \times 0,141) + (1 \times 0) + (1 \times 0) = 0,141$

$Score_{PS}(d_3, q) = (1 \times 1) + (1 \times 0,141) + (1 \times 2) + (1 \times 0) = 3,141$

$Score_{PS}(d_4, q) = (1 \times 0) + (1 \times 0,141) + (1 \times 0) + (1 \times 2) = 2,141$

* Classement :

$$D_3 > D_4 > D_1 > D_2$$

- la mesure de Cosinus : $\frac{Score_{PS}(d_j, q)}{||d_j|| \times ||q||}$

- $Score_{cos}(d_4, q) = \frac{1}{2 \times \sqrt{11}} \simeq 0,15$

- $Score_{cos}(d_2, q) = \frac{0,14}{2 \times \sqrt{10,16}} \simeq 0,064$

- $Score_{cos}(d_3, q) = \frac{3,14}{2 \times \sqrt{9,16}} \simeq 0,56$

- $Score_{cos}(d_4, q) = \frac{2,14}{2 \times \sqrt{6,16}} \simeq 0,48$

* le classement : $D_3 > D_4 > D_1 > D_2$

- Exemple de cours :

- Soit la requête suivante : q = "Système recherche information"

- Soit les documents : D_1 = "la conception de système d'information"
D_2 = "recherche dans la recherche d'information"
D_3 = "la connaissance au système de gestion de BDD"

- Modèle de pondération locale utilisé : Modèle simple.

Déterminer le classement des documents par rapport à la requête donnée en utilisant un modèle Vectoriel avec les mesure de similarité suivantes : - produit scalaire
- Cosinus.

N = 3

- classement J-M : $D_3 > D_1 > D_5 > D_4 > D_2$

* Dirichlet : avec $\mu = 40$ $\quad \underset{Dir}{score}(D,q) = \prod \left[\lambda_{Dir} P(q_i|M_D) + (1-\lambda_{Dir}) P(q_i|M_D)\right]$

$\lambda_{Dir} = \frac{|D|}{\mu + |D|}$ $\qquad 1 - \lambda_{Dir} = \frac{\mu}{|D| + \mu}$

$\underset{Dir}{score}(D_1, q) = \left[0{,}44 \times \frac{3}{7} + 0{,}56 \times \frac{8}{33}\right] \times \left[0{,}44 \times \frac{2}{7} + 0{,}56 \times \frac{4}{33}\right] \times \left[0{,}44 \times 0 + 0{,}56 \times \frac{3}{33}\right]$

$= 0{,}0032$

$\underset{Dir}{score}(D_2, q) = \left[0{,}33 \times 0 + 0{,}66 \times \frac{8}{33}\right] \times \left[0{,}33 \times 0 + 0{,}66 \times \frac{4}{33}\right] \times \left[0{,}33 \times 0 + 0{,}66 \times \frac{3}{33}\right]$

$= 0{,}000077$

$\underset{Dir}{score}(D_3, q) = 0{,}0038$ $\qquad \underset{Dir}{score}(D_5, q) = 0{,}0018$

$\underset{Dir}{score}(D_4, q) = 0{,}0024$

- classement Dir : $D_3 > D_1 > D_4 > D_5 > D_2$

$Q_0 = \{t_1, t_2\} \xrightarrow{\text{1ère recherche}} \overbrace{D_3 > D_1 > D_5 > D_4 > D_2}^{R}$ → Modèle de langue de R

$$P(t_i|R) = \sum_{i=1}^{|R|} P(t_i|D_i) \times \frac{score(D_i)}{\sum_{D \in R} score(D_i)}$$

$score(D_3) = 0{,}0038$

$score(D_1) = 0{,}0032$

t_i = Recherche.

$P(\text{Recherche}|R) = P(\text{recherche}|D_3) \times \frac{score(D_3)}{score(D_3) + score(D_1)} + P(\text{recherche}|D_1) \times$

$\frac{score(D_1)}{score(D_3) + score(D_1)} = \frac{4}{7} \times \frac{38}{38+32} + \frac{2}{7} \times \frac{38}{38+32} = 0{,}23$

TD 3

Exo 1 :

on a : C = 100 doc

$R = \{D_2, D_{13}, D_{16}, D_{10}, D_{29}, D_{42}, D_{58}, D_{77}, D_{83}, D_{85}, D_{94}, D_{98}\}$ { list de doc retournés pour la requête q.

$P = \{D_2, D_{15}, D_{29}, D_{47}, D_{67}, D_{77}, D_{93}, D_{97}\}$ { doc pertinents pour la q dans la collection.

1) calculer : $RP = \{D_2, D_{29}, D_{77}\} = 3$; $|R| = 12$; $|P| = 8$

rappel : $\frac{|RP|}{|P|} = \frac{3}{8} = 0,375$

Précision $= \frac{|RP|}{|R|} = \frac{3}{12} = 0,25$

2) Calculer la F-mesure

$\alpha = 0,5$ $F = \frac{2PR}{P+R} = \frac{2(0,25)(0,375)}{0,25+0,375} = 0,3$

$\alpha = 0,8$ $F = \frac{1}{0,8 \cdot \frac{1}{0,25} + (1-0,8)\cdot\frac{1}{0,375}} = 0,267$

Exo 2 :

1) calculons le rappel et la précision pour chaque requête.

$|P_{q_1}| = 8$ $|R_{q_1}| = 10$ $|RP_{q_1}| = 4$

Précision$(q_1) = \frac{4}{10} = 0,4$; Rappel$(q_1) = \frac{4}{8} = 0,5$

$|P_{q_2}| = 8$ $|R_{q_2}| = 10$ $|RP_{q_2}| = 5$

Précision$(q_2) = \frac{5}{10} = 0,5$; Rappel$(q_2) = \frac{5}{8} = 0,625$

2) Calculons :

$P@k = \frac{\text{nbr de doc pertinents au rang } k}{|k|}$

q_1 $\quad P@5 = \frac{2}{5}$ $\quad P@10 = \frac{4}{10}$

q_2 $\quad P@5 = \frac{3}{5}$ $\quad P@10 = \frac{5}{10}$

3) Calculer (MAP) :

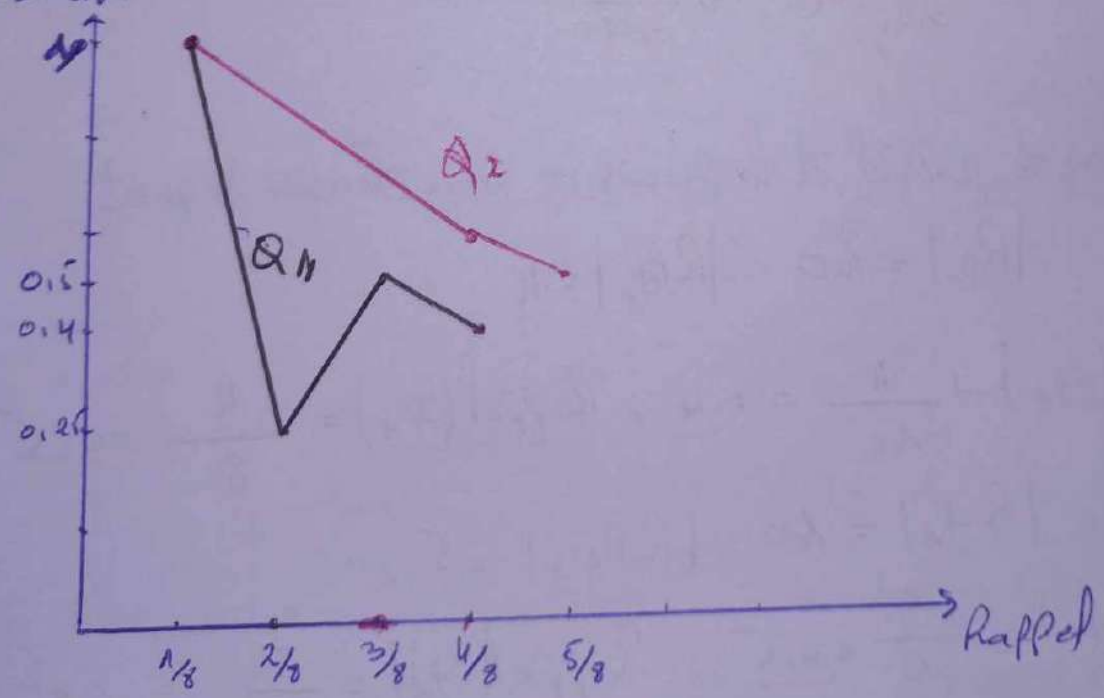
$AP_1 = \frac{1}{8}\left[1 + \frac{2}{5} + \frac{3}{6} + \frac{4}{10}\right] = 0,2875$

$AP_2 = \frac{1}{8}\left[3 + \frac{4}{7} + \frac{5}{10}\right] = 0,5$

$MAP = \frac{1}{2}\left[0,2875 + 0,5\right] = 0,39$

$R_{1j} = 1/8 ; \ldots ; 2/8 ; 3/8 ; \ldots ; 4/8$

$R_{2j} = 1/8 ; 1/8 ; 1/8 ; \ldots ; 4/8 ; \ldots ; 5/8$



Exo 3 : (Modèle probabiliste).

	recherche	base	donnée	information	logique	matrice	pertinence	flou
D_1	1	0	0	1	0	0	1	0
D_2	0	1	1	0	0	1	0	0
D_3	1	0	0	1	1	1	1	1
D_4	0	0	0	0	1	1	0	1
D_5	1	1	1	1	0	0	1	0
n_i	3	2	2	3	2	3	3	2
N	5	5	5	5	5	5	5	5
S	3	3	3	3	3	3	3	3
s_i	2	1	1	2	1	2	2	1
P_i	2/3	1/3	1/3	2/3	1/3	2/3	2/3	1/3
u_i	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2

$P_i = \frac{s_i}{S}$

$u_i = \frac{n_i - s_i}{N - S}$

$$\text{Score}(D, q) = \sum_{\substack{i \in q \\ i \in D}} \log_2 \left[\frac{P_i}{u_i} \times \frac{1 - u_i}{1 - P_i} \right]$$

$$\text{Score}(D_1, q) = \log_2 \left[\frac{P_{rech}}{u_{rech}} \times \frac{1 - u_{rech}}{1 - P_{rech}} \right] + \log_2 \left[\frac{P_{info}}{u_{info}} \times \frac{1 - u_{info}}{1 - P_{info}} \right]$$

$$+ \log_2 \left[\frac{P_{pert}}{u_{pert}} \times \frac{1 - u_{pert}}{1 - P_{pert}} \right] = \log_2 \left[\frac{2/3}{1/2} \times \frac{1 - 1/2}{1 - 2/3} \right] + \log_2 \left[\frac{2/3}{1/2} \times \frac{1 - 1/2}{1 - 2/3} \right] +$$

$$\log_2 \left[\frac{2/3}{1/2} \times \frac{1 - 1/2}{1 - 2/3} \right] = 3 \log_2 \left(\frac{2/3}{1/2} \times \frac{1 - 1/2}{1 - 2/3} \right) = 3 \log_2 \left(\frac{4}{3} \times \frac{3}{2} \right)$$

$$= 3 \log_2 \left(\frac{4}{2} \right) = \boxed{3}$$

+ log c

EXO4 : (Modèle de langue).

1) construction de Modèle de langue des (5) documents.

	recherche	info	Pertinence	base	donnée	math	logique	fbse	
MD_1	2/7	3/7	2/7	0	0	0	0	0	7
MD_2	0	0	0	2/5	2/5	1/5	0	0	5
MD_3	1/7	2/7	1/7	0	0	1/7	1/7	1/7	7
MD_4	0	0	0	0	0	2/6	2/6	2/6	6
MD_5	2/8	3/8	1/8	1/8	1/8	0	0	0	8
MC	5/33	8/33	4/33	3/33	3/33	4/33	3/33	3/33	33

2) calculons le score des doc pour la requête : "info - pertinence - fbse" en utilisant les méthodes de lissage suivantes :

* Jelinek Mercer : Pour $\lambda = 0,8$ $\quad score(D,q) = \prod_{q_i \in q} [\lambda . P(q_i/M_D) + (1-\lambda) P(q_i/M_C)]$

$score(D_1,q) = [0,8 \times \frac{3}{7} + 0,2 \times \frac{8}{33}] * [0,8 \times \frac{2}{7} + 0,2 \times \frac{4}{33}] * [0,8 \times 0 + 0,2 \frac{3}{33}]$

$= 0,0017$

$score(D_2,q) = [0,8 \times 0 + 0,2 \times \frac{8}{33}] \times [0,8 \times 0 + 0,2 \times \frac{4}{33}] \times [0,8 \times 0 + 0,2 \times \frac{3}{33}]$

$= 0,000021$

$score(D_3,q) = [0,8 \times \frac{2}{7} + 0,2 \times \frac{8}{33}] * [0,8 \times \frac{1}{7} + 0,2 \times \frac{4}{33}] \times [0,8 \times \frac{1}{4} + 0,2 \times \frac{3}{33}]$

$= 0,0051$

$score(D_4,q) = [0,8 \times 0 + 0,2 \times \frac{8}{33}] \times [0,8 \times 0 + 0,2 \times \frac{4}{33}] \times [0,8 \times \frac{2}{6} + 0,2 \times \frac{3}{33}]$

$= 0,00033$

$score(D_5,q) = [0,8 \times \frac{3}{8} + 0,2 \times \frac{8}{33}] \times [0,8 \times \frac{1}{8} + 0,2 \times \frac{4}{33}] \times [0,8 \times 0 + 0,2 \times \frac{3}{33}]$

$= 0,00078$

$$\text{score}(D_2, q) = \log_2\left(\frac{1/3}{1/2} \times \frac{1 - 1/2}{1 - 1/3}\right) + \log_2\left(\frac{1/3}{1/2} \times \frac{1 - 1/2}{1 - 1/3}\right) + \log_2\left(\frac{2/3}{1/2} \times \frac{1 - 1/2}{1 - 1/3}\right) = 2\log_2\left(2/3 \times 3/4\right) + \log_2\left(4/3 \times 3/2\right)$$

$$= 2\log_2\left(\frac{2}{4}\right) + \log_2\left(\frac{4}{2}\right) = -1$$

$$\text{score}(D_3, q) = 4\log_2\left(\frac{4}{2}\right) + 2\log_2\left(\frac{2}{4}\right) = 2$$

$$\text{score}(D_4, q) = \log_2\left(\frac{4}{2}\right) + 2\log_2\left(\frac{2}{4}\right) = -1$$

$$\text{score}(D_5, q) = 3\log_2\left(\frac{4}{2}\right) + 2\log_2\left(\frac{2}{4}\right) = 1$$

- suite exo 4 :

$P(\text{info} \mid A) = 0,38$ V
$P(\text{Part} \mid A) = 0,23$ V
$P(\text{Base} \mid A) = 0.$

$P(\text{donné} \mid A) = 0$
$P(\text{log} \mid A) = 0,07$
$P(\text{math} \mid A) = 0,07$
$P(\text{stat} \mid A) = 0,07$ V

EXO 4
1) const
	nc
MD_1	
MD_2	
MD_3	
MD_4	
MD_5	
MC	

e) calcul
en uti
* Teh
score(
score(
score(
score(
score

	Conception	de	système	d	info	recherche	obus	la	Connai	sur	gestion	BDD
ni	1	1	2	2	2	1	1	3	1	1	1	1
idfi	1,58	0	0,58	0,58	0,58	0,58	1,58	0	1,58	1,58	1,58	1,58
tfi1	1	1	1	1	1	0	0	1	0	0	0	0
tfi2	0	0	0	1	1	2	1	1	0	0	0	0
tfi3	0	1	1	0	0	0	0	1	1	1	1	1
Wi1	1,58	0	0,58	0,58	0,58	0	0	0	0	0	0	0
Wi2	0	0	0	0,58	0,58	3,46	0,58	0	0	0	0	0
Wi3	0	0	1,58	0	0	0	0	0	1,58	1,58	1,58	1,58
Wi4	0	0	1	0	1	1	0	0	0	0	0	0

Score P-S $(d_1, q) = (1 \times 0,58) + (1 \times 0,58) + (1 \times 0) = \boxed{1,16}$

Score P-S $(d_2, q) = (1 \times 0) + (1 \times 0,58) + (1 \times 1,16) = \boxed{1,74}$

Score P-S $(d_3, q) = (1 \times 0,58) + (1 \times 0) + (1 \times 0) = \boxed{0,58}$

$$D_2 > D_1 > D_3$$

$$Score_{Cos}(d_1, q) = \frac{1,16}{\sqrt{3} \times \sqrt{3,5}}$$