

Documents autorisés (Parties RI et Web sémantique)

Exercice 1 (4pts)

Considérons la collection suivante composée de 4 documents :

$D1 = \{ t1\ 2, t2\ 3, t5\ 4, t7\ 1 \}$

$D2 = \{ t2\ 1, t3\ 2, t5\ 2, t6\ 2, t8\ 2 \}$

$D3 = \{ t1\ 1, t3\ 1, t4\ 1, t8\ 2 \}$

$D4 = \{ t3\ 1, t4\ 3, t8\ 4 \}$

Et la requête suivante : $Q = \{ 2\ t1, 1\ t4 \}$

Un document (resp. requête) est représenté par une liste de termes pondérés ayant la forme suivante : $D_j \{ tf(t_i) t_i \}$, $i=1..8$. Où, $tf(t_i)$ signifie la fréquence du terme t_i dans D_j .

Questions :

Donner pour chacun des modèles ci-dessous l'ordre dans lequel les 4 documents seront renvoyés en réponse à la requête :

- Le modèle vectoriel utilisant la mesure produit (Inner Product) en prenant une pondération de type (qqq.ddd) nnn.ltn
- Le modèle de langue basé sur un lissage de Dirichlet ($\mu=1$)

Exercice 2 (6 pts)

Soit $q = q_1, \dots, q_m$ une requête, d un document et $P(q_i/d)$ la probabilité du mot q_i dans le modèle de langue de d . On suppose que nous disposons d'une collection de documents comportant au total 8 mots $w1, \dots, w8$.

La Table ci-dessous liste pour chaque mot sa probabilité dans le modèle de langue de référence, $P_{ml}(w|REF)$, estimé sur la collection (2ème colonne), la fréquence du terme $c(w; d)$ dans un document (3ème colonne). Les colonnes 4 et 5 représentent les probabilités du terme dans le modèle langue du document d , estimé respectivement selon le maximum de vraisemblance et Dirichlet avec le paramètre μ .

Mots	$P_{ml}(w REF)$	$c(w, d)$	$P_{ml}(w d)$	$P_{dir}(w d)$
w1	0.3	2		
w2	0.15	1		
w3	0.1	2		0.125
w4	0.1	4		
w5	0.05	1		
w6	0.1	0		
w7	0.1	0		
w8	0.1	0		

- 1- Remplir la colonne 4, ($P_{ml}(w|d)$), le modèle de langue du document.
- 2- La colonne 5 représente la probabilité du terme calculée après un lissage de Dirichlet effectuée sur la collection. Seule la probabilité de w_3 est donnée dans le tableau, déduire la valeur de μ ? (posez l'équation puis déduire cette valeur)
- 3- Sans calculer les probabilités de la colonnes 5, indiquer pour chacun des mots de cette colonne si sa probabilité lissée ($P_{dir,\mu}(w|d)$) est $\{>;=;<\}$ à celle non lissée, calculée selon $P_{ml}(w|d)$, c'est-à-dire celle de la colonne 4.
- 4- Quelle condition doit satisfaire $c(w, d)$ pour que la probabilité lissée du mot w soit toujours la même que la valeur non lissée quelque soit le paramètre μ la probabilité non lissée).

Partie Web Sémantique

Exercice 3 (4pts) (réponses en rouge)

```
SELECT DISTINCT ?album ?label where {  
  ?album rdf:type dbpedia:Album  
  ?x rdfs:label ?label }  
LIMIT 100
```

3.1 Corriger cette requête qui comporte 2 erreurs syntaxiques.

```
SELECT DISTINCT ?album ?label where {  
  ?album rdf:type dbpedia:Album .  
  ?artist rdfs:label ?label }  
LIMIT 100
```

3.2 La modifier pour qu'elle retourne les albums classés par ordre alphabétique de leur étiquette.

```
SELECT DISTINCT ?album ?label where {  
  ?album rdf:type dbpedia:Album .  
  ?artist rdfs:label ?label }  
LIMIT 100  
ORDER BY ?label
```

3.3 Sachant que dans DBpedia la propriété dbpedia:artist relie une entité de type artist et une entité de type album, écrire la requête qui retourne classées par ordre alphabétique et sans doublon les étiquettes des albums de l'artiste « Bowie » (on supposera que l'entité [http://fr.dbpedia.org/resource/Catégorie:David Bowie](http://fr.dbpedia.org/resource/Catégorie:David_Bowie) a un label)

```
SELECT DISTINCT ?labelAlbum ?label ?artist where {  
  ?artist rdf:type dbpedia:Artist .  
  ?artist dbpedia:artist ?album .  
  ?album rdf:type dbpedia:Album . (ligne facultative)  
  ?artist rdfs:label ?label .  
  ?album rdfs:label ?labelAlbum .  
  FILTER (CONTAINS (str(?label), "Bowie")) ou FILTER (regex (?label, "Bowie"))  
}  
ORDER BY ?label  
LIMIT 100
```

3.4 La requête suivante peut-elle retourner le tableau de la Figure 1 (extrait des résultats) ? Expliquer pourquoi et si ce n'est pas le cas, donner la requête modifiée qui permet de l'obtenir.

```
SELECT DISTINCT ?award ?label ?artist where {  
  ?artist rdf:type dbpedia:Artist .  
  ?artist dbpedia:award ?award .  
  ?artist rdfs:label ?label .  
  FILTER (CONTAINS (str(?label), "Williams"))  
}  
ORDER BY ?label
```

LIMIT 100

award	label	artist
Eisner Award	Aaron Williams (cartoonist)	Aaron Williams (cartoonist)
	Al Williamson	Al Williamson
	Alexander Williams (cartoonist)	Alexander Williams (cartoonist)
	Allan Williams	Allan Williams
	Andy Williams	Andy Williams
	Andy Williams (Doves)	Andy Williams (Doves)
	Anna-Lynne Williams	Anna-Lynne Williams
	Annie Williams (singer)	Annie Williams (singer)
	Anthony Williams (comics)	Anthony Williams (comics)
	Astrid Williamson	Astrid Williamson
	Beau Williams	Beau Williams
	Ben Ames Williams	Ben Ames Williams
	Big Joe Williams	Big Joe Williams
	Billy Drease Williams	Billy Drease Williams
	Boris Williams	Boris Williams
	Brandi Williams	Brandi Williams
	Brooks Williams	Brooks Williams
	Buddy Williams (country musician)	Buddy Williams (country musician)
	Buster Williams	Buster Williams
	Cathy Williams	Cathy Williams
	Charles Williams (British writer)	Charles Williams (British writer)
	Charles Williams (artist)	Charles Williams (artist)
	Christopher Williams (Welsh artist)	Christopher Williams (Welsh artist)

Figure 1

Non : les lignes sans awards n'apparaîtraient pas.

Pour autoriser que ?award soit vide, il faut ajouter OPTIONAL

```

SELECT DISTINCT ?award ?label ?artist where {
  ?artist rdf:type dbpedia:Artist .
  OPTIONAL (?artist dbpedia:award ?award .)
  ?artist rdfs:label ?label .
  FILTER (CONTAINS (str(?label), "Williams"))
}
ORDER BY ?label
LIMIT 100

```

Exercice 4 (6pts)

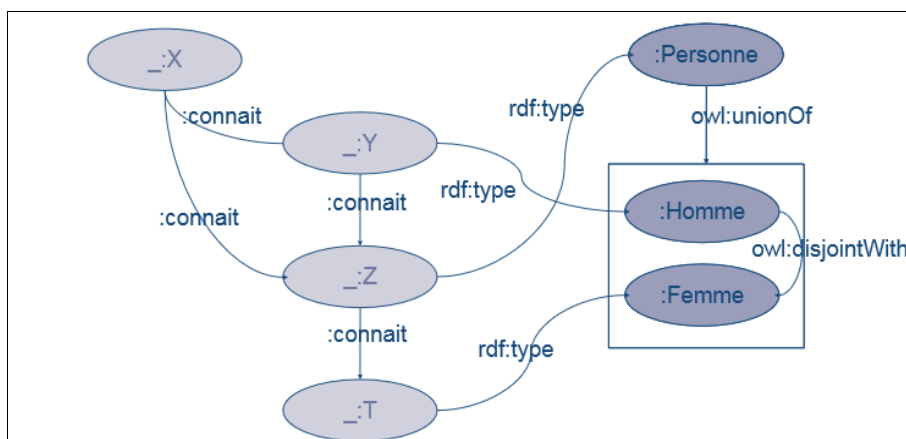


Figure 2 - graphe de connaissances

4.1 sur ce diagramme (figure 2): que représentent `_:x`, `_:y`, `_:t` et `_:z` ?

Ce sont des variables qui représentent des nœuds vides. $_ :T$ est de type $:femme$, $_ :Z$ est de type $:Personnes$ et $_ :Y$ de type $:Homme$. Ce sont des instances si $:Personne$, $:Homme$ et $:Femme$ sont des classes.

4.2 Quelles sont les classes ?

$:Personne$, $:Homme$ et $:Femme$ sont des catégories ontologies et éventuellement des owl :class

4.3 Ce graphe traduit-il « $_ :X$ connaît un homme qui connaît une femme » ?

Non, il traduit « $_ :X$ connaît une personne ($_ :Z$) qui connaît une femme ($_ :T$) »

Ainsi que « $_ :X$ connaît un homme ($_ :Y$) qui connaît une personne ($_ :Z$) qui connaît une femme »

4.4 A partir de ce graphe, peut-on donner le type de $_ :X$?

non

4.5 Supposons que $:connaît$ soit définie ainsi

```
<owl:ObjectProperty rdf:about=":connaît">
  <rdf:type rdf:resource="&owl;SymetricProperty"/>
  <rdf:type rdf:resource="&owl;TransitiveProperty"/>
  <rdfs:range rdf:resource=":Person"/>
  <rdfs:domain rdf:resource=":Person"/>
</owl:ObjectProperty>
```

Quel est alors le type de $_ :X$ avant lancement d'un raisonneur ?

Inconnu

après lancement d'un raisonneur ?

$:Person$

4.6 Quels sont tous les triplets qui peuvent être inférés à partir des connaissances de la figure 1 ?

Comme $:connaît$ est transitive et symétrique

Transitivité : $_ :X :connaît _ :T ; _ :Y :connaît _ :T$

Symétrie : $_ :T :connaît _ :Z, _ :T :connaît _ :Y ; _ :T :connaît _ :X ; _ :Z :connaît _ :Y ; _ :Y connaît _ :X ;$

$_ :Z :connaît _ :X$