

**Examen du cours Recherche d'information / web sémantique**  
**ENSEEIH 3A Info – année 2013-2014**  
**Mercredi 5 mars 2014**

Les documents sont autorisés pour les 2 parties.

**Partie « recherche d'information »**

(durée indicative : 1h – barème 10 pts)

**Exercice 1 (4pts)**

Les mots très fréquents dans une collection de documents sont souvent des mots outils (stoplist) qui ont peu d'intérêt dans le processus de sélection de documents pertinents pour une requête donnée.

**Question :**

Donner, en fonction de paramètres longueur du document et/ou la taille de la collection (somme des longueurs des documents), une estimation du nombre de termes apparaissant au moins 2000 fois dans une collection comportant N documents.

**Exercice 2 (6pts)**

Considérons la collection ci dessous composée de 3 documents :

$D_1 = \{2 t_1, 3 t_2, 4 t_5, 1 t_7\}$

$D_2 = \{1 t_2, 2 t_3, 2 t_5, 2 t_6, 2 t_8\}$

$D_3 = \{1 t_1, 1 t_3, 5 t_4, 2 t_8\}$

Et la requête suivante :  $Q = \{2 t_1, 1 t_2, 1 t_4\}$

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 :**

- 1- Donner graphiquement le fichier inverse (dictionnaire + *Posting*) représentant cette collection de 3 documents
- 2- Donner pour chacun des modèles ci-dessous l'ordre dans lequel les 3 documents seront renvoyés en réponse à la requête Q :
  - Le modèle vectoriel utilisant la mesure du cosinus (*Cosine Measure*) en prenant une pondération de type *Inc.Inc*
  - Le modèle probabiliste (Probabilistic Ranking Principle) (BIR Model)
  - Le modèle de langue mixte basé sur un lissage de Dirichlet ( $\mu=1$ )

## **Partie « web sémantique »**

(durée : 1h00 – barème : 10 points)

### **Exercice 3 (Questions de cours – 2 pts)**

3.1 Lorsque l'on construit une ontologie à partir de textes, en dépouillant les documents pour identifier les concepts d'un domaine, le développeur accède-t-il directement :

- a) aux objets réalisant les concepts ?
- b) aux notions des concepts ?
- c) aux termes exprimant les concepts ?

(justifier la réponse choisie)

3.2 Qu'appelle-t-on le lexique d'une ontologie ?

### **Exercice 4 (4 pts)**

Voici un « fichier » qui décrit un vocabulaire et des ressources.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/#"
  xmlns:ex3="http://www.irit.fr/MELODI/ontologie/ex3#"
  xml:base="http://www.irit.fr/ MELODI/ontologie/ex3">

  <rdfs:Class rdf:ID="Stagiaire">
    <rdfs:subClassOf rdf:resource="foaf:person"/>
  </rdfs:Class>

  <rdfs:Class rdf:ID="Tuteur">
    <rdfs:subClassOf rdf:resource="foaf:person"/>
  </rdfs:Class>

  <rdf:Property rdf:ID="encadre">
    <rdfs:domain rdf:resource="#Tuteur"/>
    <rdfs:range rdf:resource="#Stagiaire"/>
    <rdfs:subPropertyOf rdf:resource="http://xmlns.com/foaf/0.1/knows"/>
  </rdf:Property>

  <rdf:Description rdf:ID="Mouna">
    <ex3:encadre rdf:resource="#Jean-Philippe"/>
  </rdf:Description>
```

4.1- Quel est la notation utilisée ici ? quels sont les langages utilisés ? (1 pt en tout)

4.2- Donner sous la forme d'un graphe les 9 triplets décrits dans ce fichier. (1,5pt)

4.3- Quels sont les 5 triplets supplémentaires que l'on peut inférer d'après la sémantique de `rdfs:subClassOf` et `rdfs:subPropertyOf` ? (1,5 pt)

### Exercice 5 (4 pts)

Considérons la représentation OWL et RDF dans le « fichier » ci-dessous :

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:foaf="http://xmlns.com/foaf/0.1/#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:ex3="http://www.irit.fr/MELODI/ontologie/ex3#"
xml:base="http://www.irit.fr/ MELODI/ontologie/ex3">

<rdfs:Class rdf:ID="RoleSocial">

<rdfs:Class rdf:ID="Stagiaire">
<rdfs:subClassOf rdf:resource="#RoleSocial"/>
</rdfs:Class>

<rdfs:Class rdf:ID="Tuteur">
<rdfs:subClassOf rdf:resource="#RoleSocial"/>
</rdfs:Class>

<rdf:Property rdf:ID="exerce">
<rdfs:domain rdf:resource="foaf:Person "/>
<rdfs:range rdf:resource="#RoleSocial"/>
</rdf:Property>

<rdf:Property rdf:ID="encadre">
<rdfs:domain rdf:resource="#Tuteur"/>
<rdfs:range rdf:resource="#Stagiaire"/>
<rdfs:subPropertyOf rdf:resource="http://xmlns.com/foaf/0.1/knows"/>
</rdf:Property>

<owl:Class rdf:ID="PersonneTuteur">
  <rdfs:subClassOf rdf:resource="foaf:Person" />
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#exerce" />
      <owl:someValuesFrom rdf:resource="#Tuteur" />
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

<owl:Class rdf:ID="PersonneStagiaire">
  <rdfs:subClassOf rdf:resource="foaf:Person" />
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#exerce" />
      <owl:someValuesFrom rdf:resource="#Stagiaire" />
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

<rdf:Description rdf:ID="Mouna">
<ex3:encadre rdf:resource="#Jean-Philippe"/>
</rdf:Description>

<rdf:Description rdf:ID="Nathalie">
<ex3:encadre rdf:resource="#Julien"/>
<ex3:encadre rdf:resource="#Laurent"/>
</rdf:Description>
```

5.1- Que signifie la définition de `#personneStagiaire` ? (0,5 pt)

5.2- Ecrire une requête SPARQL qui retourne toutes les personnes qui sont des stagiaires. (1pt)

5.3- Que retourne cette requête à partir du fichier ci-dessus ? (0,5pt)

5.4- Soient les 2 requêtes SPARQL suivantes portant sur les connaissances précédentes

```
SELECT ?p1 ?p2
WHERE
{ ?p1 ex3:encadre ?p2 } GROUP BY ?p1
```

```
SELECT ?p1 ?p2
WHERE
{ ?p1 rdf:type foaf:Person .
  ?p2 rdf:type foaf:Person .
  ?p1 ex3:encadre ?p2 } GROUP BY ?p1
```

Que retournent les deux requêtes à partir des connaissances de l'énoncé ? (2pt)