



# Séminaire départemental

---

- Sous le thème:

Comparaison d'Approches pour l'Interrogation d'Ontologies  
Formelles via le Langage Naturel



# Séminaire départemental

---

- **PLAN:**

- ➔ Contexte
- ➔ Présentation des approches
- ➔ Discussion et Perspectives



# Séminaire départemental

- Contexte :

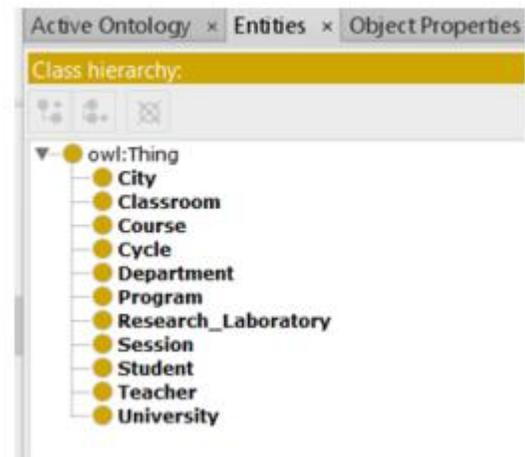
## Qu'est-ce qu'une ontologie ?

Le web sémantique repose sur l'utilisation d'ontologies pour assurer le partage, la réutilisation et l'interopérabilité des données, représentant ainsi des connaissances compréhensibles par les ordinateurs. Les ontologies informatiques servent spécifiquement à formaliser un ensemble de connaissances de manière exploitable par un ordinateur.

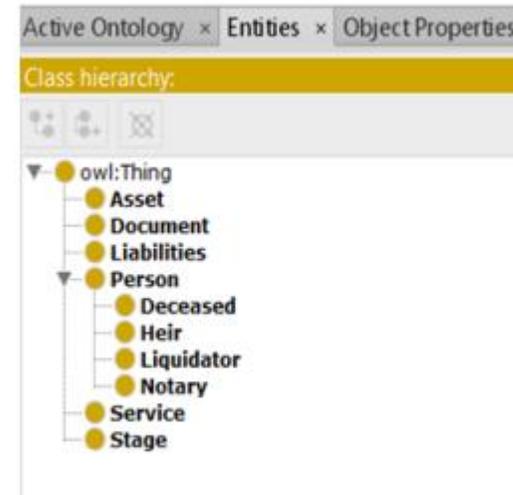
Les figures a et b ci-dessous illustrent les artefacts édités avec l'application Protégé 2000, lesquels ont été utilisés pour évaluer nos approches. Il s'agit de deux ontologies distinctes, l'une axée sur la liquidation de succession et l'autre sur les concepts universitaires.



# Séminaire départemental



(a) university ontology



(b) ontology of estate liquidation



# Séminaire départemental

---

- Contexte :

Quelle est la meilleure approche pour explorer une ontologie ?

L'interrogation des ontologies, souvent réalisée en utilisant le langage SPARQL, se présente comme un défi, en particulier pour les utilisateurs non familiers. Les obstacles incluent des barrières linguistiques dues à la complexité syntaxique, la nécessité de comprendre la structure sous-jacente de l'ontologie, les erreurs potentielles dans la formulation des requêtes et la difficulté à exprimer des requêtes complexes. Pour rendre l'accès à la connaissance plus convivial, cette présentation explore l'interrogation des ontologies en langage naturel.



# Séminaire départemental

- **Approche à base de scenarios: Application d'une métrique de Levenshtein sur une structure de modèle de Ngrammes**

L'approche adoptée repose sur l'extraction de scénarios de triplets d'une ontologie pour former une requête capable de répondre aux besoins des utilisateurs exprimés en langage naturel. Les divers scénarios peuvent inclure des questions unaires, binaires et ternaires selon les besoins de l'utilisateur. Les requêtes générées en fonction des types de scénario suivent la logique suivante :

Soit : **Rg**=requête générée

**To**=Les triplet (s) similaire(s) extrait(s) de l'ontologie

**$\forall t \in T_0, =s, p, o$**

{s=sujet, p=prédictat, o=objet}

Soit **L<sub>o</sub>** une liste d'entités identifiées (Classe, propriétés de données, propriétés objets et instances) dans **To**, avec

**C<sub>o</sub>**=classes identifiées dans To

**P<sub>do</sub>**= propriétés de données dans To,

**P<sub>oo</sub>** =propriétés objet dans To,

**I<sub>o</sub>** = instances dans To

**R<sub>g</sub>=f(L<sub>o</sub>)**

6/14

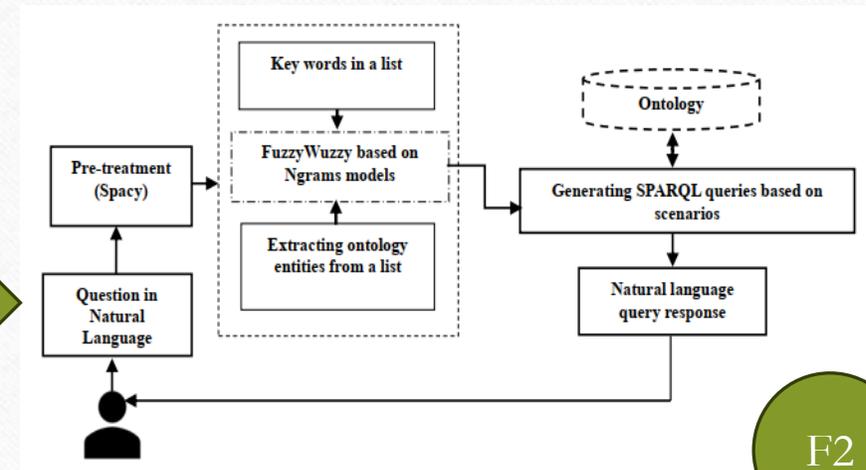
Les figures F1 et F2 illustrent respectivement les types de scénarios et l'architecture générale de l'approche.



# Séminaire départemental

- Approche à base de scenarios: Application d'une métrique de Levenshtein sur une structure de modèle de Ngrammes

Type de question et descriptions	Scenarios(sc) possibles
<b>Questions Unaires :</b> questions impliquant une seule contrainte ou paramètre. Elles demandent généralement la liste des instances d'une classe.	Sc1 : Valeurs instances des classes, cela correspond: $R_g = f(C_0)$
<b>Questions Binaires:</b> Questions impliquant deux contraintes ou paramètres. Elles peuvent inclure des requêtes avec des valeurs de propriétés spécifiques ou des relations entre classes.	Sc2 : La liste d'instances d'une classe avec leurs valeurs respectives dans une propriétés spécifique. Cela correspond à $R_g = f(C_0, P_{do})$
	Sc3 : Une instance spécifique avec une valeur de propriété particulière propriété. Cela correspond à $R_g = f(C_0, P_{do}, I_o)$
	Sc4 : Les instances d'une classe avec la relation spécifique qu'elles entretiennent avec les instances d'une autre classe. $R_g = f(C_{01}, C_{02}, P_{oo})$
<b>Questions Ternaires :</b> Questions impliquant trois contraintes ou paramètres. Elles peuvent inclure des requêtes avec deux valeurs de propriétés spécifiques d'une instance.	Sc5 : L'instances spécifique d'une classe avec la relation spécifique qu'elle entretient avec les instances d'une autre classe. $R_g = f(C_{01}, C_{02}, P_{oo}, I_o)$
	Sc6 : <b>Obtenir</b> les instances d'une classe avec deux valeurs de propriétés spécifiques. Cela correspond à $R_g = f(C_0, P_{do1}, P_{do1})$
	Sc7 : <b>Obtenir</b> les deux valeurs de propriétés spécifiques d'une instance particulière d'une classe. Cela correspond à $R_g = f(C_0, P_{do1}, P_{do1}, I_o)$



Approche basée sur des scénarios





# Séminaire départemental

- Approche basée sur les patrons de requêtes SPARQL: Moyenne pondérée(Jaccard + Cosinus)

---

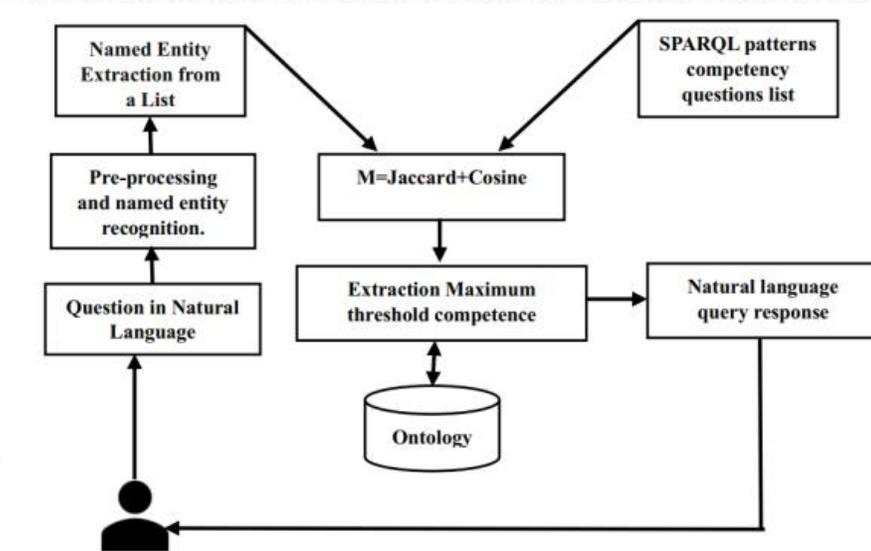
Dans cette approche, nous associons la question de l'utilisateur à sa requête SPARQL correspondante dans le modèle préétabli. Le processus commence par la reconnaissance d'entités nommées, éclaircissant le domaine de recherche. Ces entités préviennent les malentendus entre la phrase de l'utilisateur et le domaine associé. La reconnaissance d'entités a été appliquée à un corpus de 300 à 500 questions via le module Doccano. Après des prétraitements, incluant la normalisation et la reconnaissance d'entités, Nous constituons ensuite une liste d'entités clés présentes dans la phrase de l'utilisateur. La similarité entre les listes de mots-clés(Liste composée des mots-clés extraits de la phrase et d'un ensemble de listes générées à partir des mots-clés présents dans les requêtes existantes dans le patron) est évaluée pour prédire la classe de requête, utilisant Jaccard et le Cosinus avec une moyenne pondérée.

Les figures F3 et F4 présentent respectivement l'architecture générale de l'approche et la structure des patrons de requêtes SPARQL.



# Séminaire départemental

- Approche basée sur les patrons de requêtes SPARQL: Moyenne pondérée (Jaccard + Cosinus)



```
"query1": "SELECT ?comment ?designer WHERE {?comment rdf:type table:Acteurs .  
?comment table:designer ?designer. FILTER ( ?comment = table:liquidateur ) OFFSET 0 LIMIT 10",  
"query2": "SELECT ?document ?role WHERE {?document rdf:type table:Documents .  
?document table:role ?role.} OFFSET 0 LIMIT 10",  
"query3": "SELECT ?document ?role WHERE {?document rdf:type table:Documents .  
?document table:role ?role. FILTER ( ?document = table:testament )} OFFSET 0 LIMIT 10",  
"query4": "SELECT ?ouvrir ?comment WHERE {?ouvrir rdf:type table:compte_bancaires  
. ?ouvrir table:ouvrir ?comment.} OFFSET 0 LIMIT 10",  
"query5": "SELECT ?ouvrir ?quand WHERE {?ouvrir rdf:type table:compte_bancaires .  
?quand table:debuter ?ouvrir.} OFFSET 0 LIMIT 10",  
"query6": "SELECT ?fermer ?comment WHERE {?fermer rdf:type table:compte_bancaires  
. ?fermer table:fermer ?comment .} OFFSET 0 LIMIT 10",  
"query7": "SELECT ?commencer ?fermer WHERE {?commencer rdf:type table:compte_bancaires .  
?commencer table:debuter ?commencer .} OFFSET 0 LIMIT 10",  
"query8": "SELECT ?inscrire ?comment WHERE {?inscrire rdf:type table:Acteurs.  
?inscrire table:inscrire ?comment. FILTER ( ?inscrire = table:liquidateur )} OFFSET 0 LIMIT 10",  
"query9": "SELECT ?appeler ?heritier WHERE {?appeler rdf:type table:Acteurs.  
?appeler table:appeler ?heritier.} OFFSET 0 LIMIT 10",  
"query10": "SELECT ?quand ?heritier WHERE {?quand rdf:type table:Acteurs.  
?quand table:appeler ?heritier.} OFFSET 0 LIMIT 10",  
"query11": "SELECT ?comment ?proceder WHERE {?comment rdf:type table:etape.  
?comment table:proceder ?proceder.} OFFSET 0 LIMIT 10",
```

Approche basée sur les patrons de requêtes SPARQL

F3

F4



# Séminaire départemental

- Approche basée sur la structure d'arbre de décision : Moyenne pondérée (Jaccard + Cosinus)
- 

- Cette approche offre une approche progressive pour présenter le problème à l'utilisateur, passant de la généralité à la spécificité. Il permet de partir de l'intention exprimée par l'utilisateur dans sa requête pour présenter un point central du problème, à partir duquel plusieurs branches émergent, représentant des idées spécifiques qui peuvent être déduites. Son principe est le suivant : nous élaborons un arbre composé de différents nœuds sous forme de liste de jetons et chacun représentant un sous-graphe du domaine représenté.
- Les décisions à prendre sont déterminées par des conditions ou critères préétablis, basés sur les valeurs de similarité maximale entre les entités spécifiques de la phrase et les nœuds de l'arbre. Chaque branche reliant les nœuds à la similarité maximale indiquera la séquence de décisions à présenter à l'utilisateur.

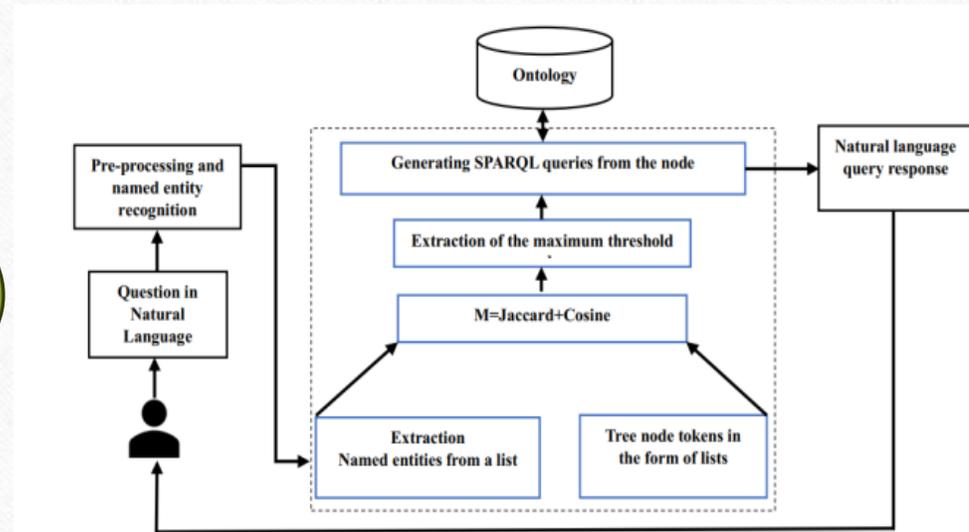
La figure F5 présente l'architecture générale de cette approche.



# Séminaire départemental

- Approche basée sur la structure d'arbre de décision : Moyenne pondérée (Jaccard + Cosinus)

F5



Architecture de l'approche basée sur la structure d'arbre de décision



# Séminaire départemental

---

- Discussion et perspectives:

Avantages :

- L'approche basée sur les scénarios s'est avérée parfaitement adaptée pour répondre aux questions portant sur des ontologies distinctes, en alignant ses réponses avec la logique des scénarios utilisés pour valider notre méthode.
- L'approche basée sur les patrons réussit à cartographier précisément les questions prédéfinies dans les patrons SPARQL.
- De manière analogue à l'approche des patrons, celle axée sur la structure d'arbre de décision, également focalisée sur l'ontologie de la liquidation d'une succession, a habilement attribué chaque question au nœud correspondant dans l'arbre de décision.



# Séminaire départemental

- Discussion et perspectives:

## Limites

- ✓ Les trois approches testées présentent chacune leurs avantages et limites. Toutefois, en raison des mises à jour fréquentes qu'un domaine de connaissances peut subir, l'approche basée sur des patrons de requête SPAQRL et celle basée sur des arbres de décision prendraient plus de temps pour intégrer les nouvelles mises à jour dans l'ontologie. En effet, en raison de leurs structures peu évolutives, chaque nouvelle connaissance ajoutée à l'ontologie nécessiterait un enrichissement supplémentaire des patrons de requête pour l'approche basée sur des patrons, ainsi que l'ajout de plus de nœuds pour répondre aux nouveaux besoins dans le cas de l'approche basée sur des arbres de décision.
- ✓ Il est important de souligner que l'approche basée sur les patrons et celle axée sur la structure d'arbre démontrent une sensibilité marquée à des éléments spécifiques, tels que l'entraînement du modèle pour une gestion précise des entités nommées, ainsi que leur efficacité dans le contexte d'une ontologie particulière.
- ✓ L'efficacité de l'approche basée sur les scénarios est étroitement liée aux types de scénarios, au jargon spécifique utilisé, ainsi qu'à la nomenclature des entités.



# Séminaire départemental

---

- Discussion et perspectives:

## Perspectives

Nous cherchons à améliorer considérablement notre approche basée sur des patrons prédéfinis en automatisant la génération de requêtes après une phase d'apprentissage. Pour atteindre cet objectif, nous envisageons d'intégrer un modèle de transformateur encodeur-décodeur dans notre méthodologie.