

CONCEPTION D'UN FRAMEWORK POUR LE TRAITEMENT COOPERATIF DES REQUETES SEMANTIQUES

Présenté par :

Géraud FOKOU

Sous l'encadrement de :

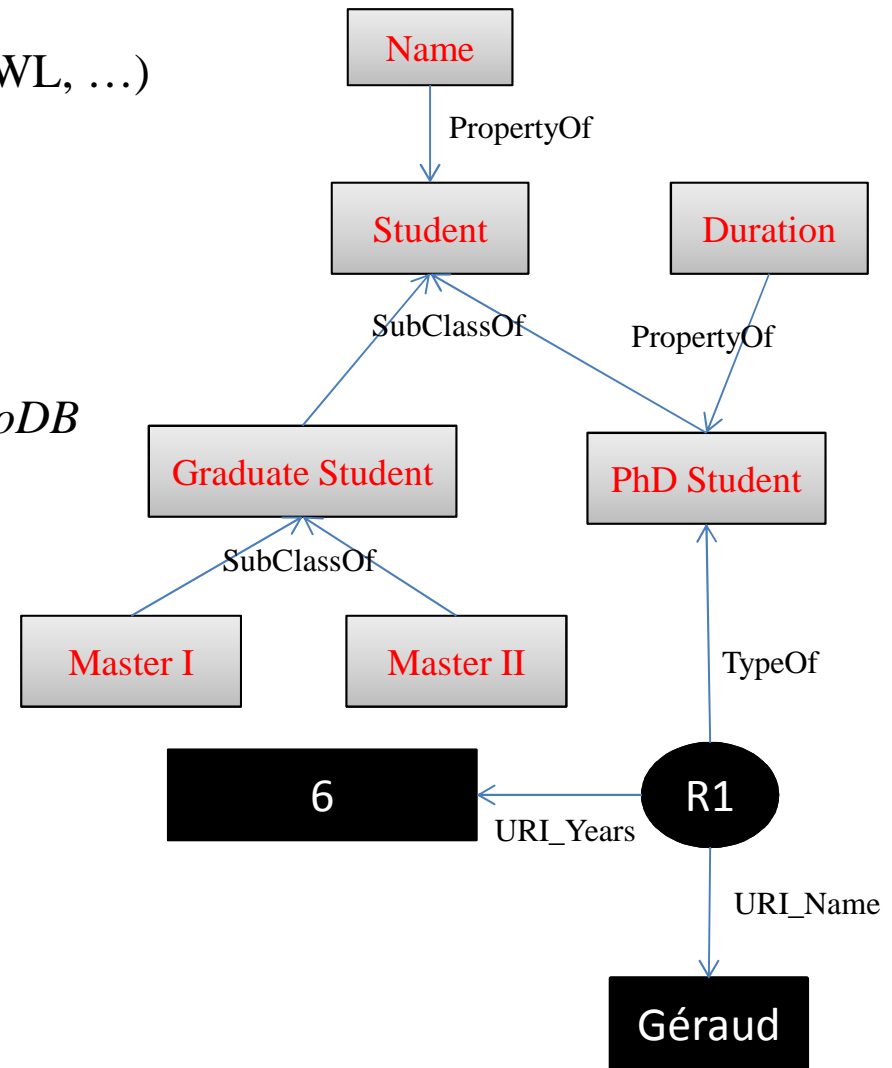
Pr. Allel HADJALI

Dr. Stéphane JEAN

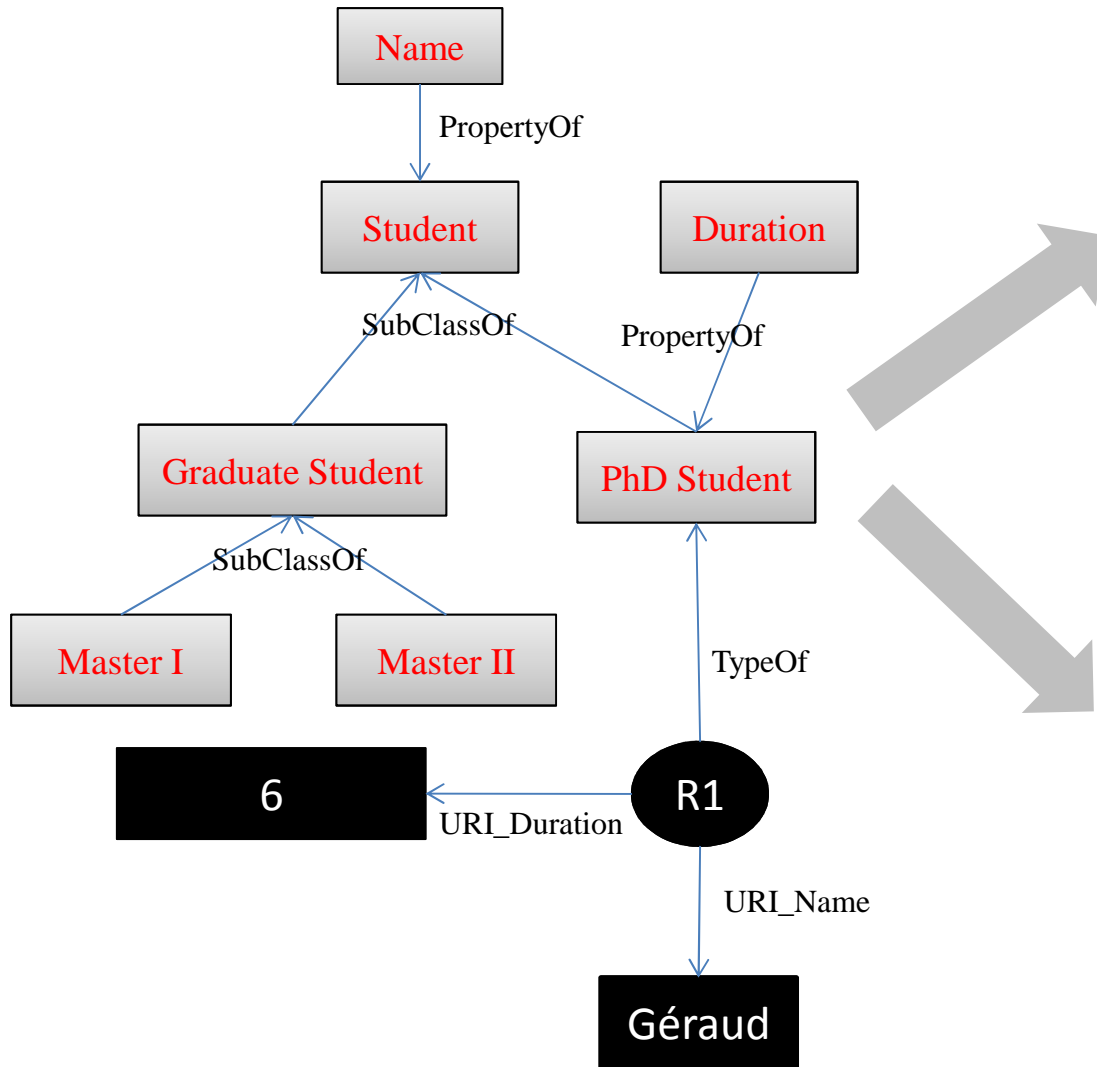
le 10 /04/2014

□ Bases de Données Sémantiques (BDSs)

- Stockage des données sémantiques (RDFS, OWL, ...)
- Représentation
 - Verticale (Triplets), *ex: Jena*
 - Horizontale (Classes, Attributs), *ex: OntoDB*
 - Binaire (Propriétés-valeurs)
- Interrogation
 - SPARQL, (*?X, Years, I_{st}*)
 - SQL, (*Select ... From ...*)
 - Hybride (*OntoQL*)



□ Bases de Données Sémantiques (BDSs)



Représentation Verticale

Subject	Predicat	Object
Student	RDF_TypeOf	RDF_Class
PhDStudent	RDF_SubClassOf	Student
GraduateStudent	RDF_SubClassOf	Student
Name	RDF_PropertyOf	Student
Years	RDF_PropertyOf	PhDStudent
R ₁	RDF_TypeOf	PhDStudent
R ₁	URI_name	Géraud
R ₁	URI_Duration	6

Représentation Horizontale

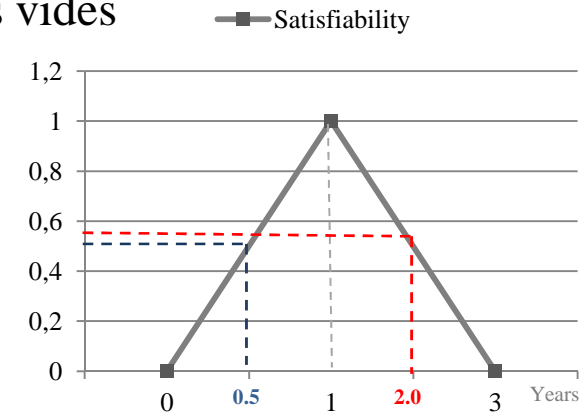
MASTER I		MASTER II	
ID	Name	ID	Name

Student		Graduate Student	
ID	Name	ID	Name

PhD Student		
ID	Name	Years
R1	Geraud	6

□ Relaxation/Approximation

- **Objectif:** Proposer des solutions aux requêtes à réponses vides
 - Approximation
 - BDR: Estimation à l'aide d'une relation d'ordre
 - BDS: Définition des **mesures de similarité**
 - Mesure de similarité
 - Evaluation de la **distance (pertinence)**: *Distance (PhD Student, Student), Distance (PhD Student, Master I)*
 - **Distance:** Hiérarchique (*chemins du graphe*), Information (*instances ou valeur*), Utilisateur
 - Relaxation de requête
 - Approximation des résultats d'une requête: *Select Name From PhD Student Where duration = 12*
 - BDR utilisation de la clause « where » et la logique floue sur les ensembles de valeur
- Evolution des BDR vers les BDS → **Relaxation dans les BDS??**



I. ETAT DE L'ART

II. PROBLEMATIQUES ET OBJECTIFS

III. FORMALISATION D'OPERATEURS DE RELAXATION

IV. VERS DES PROCESSUS DE RELAXATION GUIDES ET OPTIMISES

□ Techniques de Relaxation

- Automatique : logique floue et Raisonnement

➤ Ex: *Select Name From PhD Student Where duration = 12*

- \approx *Select Name From Student Where duration = 12*
- \approx *Select Name From PhD Student Where duration IN [6,18]*

- Orientées Utilisateurs

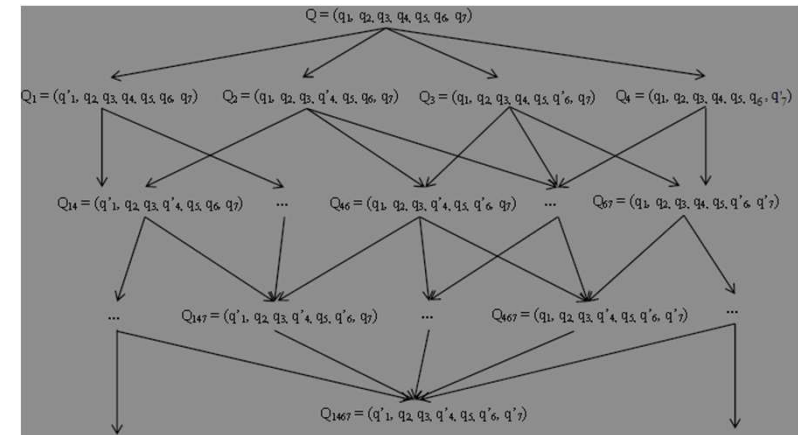
- Etudes des feedbacks utilisateurs
- Modélisation des Profils utilisateurs

➤ Invariants

- Réécriture de la requête initiale Q en requête Q'
- $Answer(Q) \subset Answer(Q')$
- Processus itératif, si Q'' est une réécriture de Q alors :
 - $Answer(Q) \subset Answer(Q') \subset Answer(Q'')$
 - $Dist(Q, Q') < Dist(Q, Q'')$

Table 1. Inference Rules

Group A (Sub-Property)	(1) $\frac{(a, sp, b)(b, sp, c)}{(a, sp, c)}$	(2) $\frac{(a, sp, b)(x, a, y)}{(x, b, y)}$
Group B (Sub-Class)	(3) $\frac{(a, sc, b)(b, sc, c)}{(a, sc, c)}$	(4) $\frac{(a, sc, b)(x, type, a)}{(x, type, b)}$
Group C (Typing)	(5) $\frac{(a, dom, c)(x, a, y)}{(x, type, c)}$	(6) $\frac{(a, range, d)(x, a, y)}{(x, type, d)}$



□ Mesures de similarités

- Approximation s'évalue via la similarité entre les requêtes

- $Sim(Q, Q') = \frac{2 * depth(LCA(C_1, C_2))}{depth(C_1) + depth(C_2)}$ (Distant-Based) [Huang08]

- $Sim(Q, Q') = \frac{IC(LCA(C_1, C_2))}{IC(C_1) + IC(C_2) - IC(LCA(C_1, C_2))}$ (Content-Based) [Jean13] où C_2 dans

Q' relaxe C_1 dans Q , IC est la quantité d'information dans une classe et LCA est le plus proche ancêtre commun entre C_1 et C_2 (*Less common Ancestor*)

- $Sat(h_i)$ donne la satisfiabilité d'une instance h_i . [Jean13]

□ Les Processus de Relaxation

- Processus automatique et itératif [Hurtado08]
- Processus de substitution par des préférences utilisateurs [Dolog09]
- Processus avec Feedback utilisateurs [Islam12]

□ Synthèse

	Problématiques			
	Operateurs	Ranking	Contrôle	Implémentation et intégration
Dolog et al.	n/a	Based on users preferences	Trigger rules	Sesame
Hurtado et al.	Relax Clause	Distance-Based	Top-k	n/a
Huang et al.	n/a	Distance-Based Content-Based	Top-k	Jena (LUBM)
Elbassuoni et al.	n/a	Content-Based	Domain Target	Tests on real data set
Hogan et al.	n/a	Content-Based	Domain Target	Generic Framework for EADS

□ Objectifs

O₁ : Définition d'opérateurs de relaxation:

combinaison et paramétrage

O₂ : Fonction de Ranking des résultats

O₃ : Contrôle et guide des processus de relaxation

O₄ : Implémentation et intégration de la relaxation dans les BDSs

□ Définition des opérateurs de relaxation dans les BDS

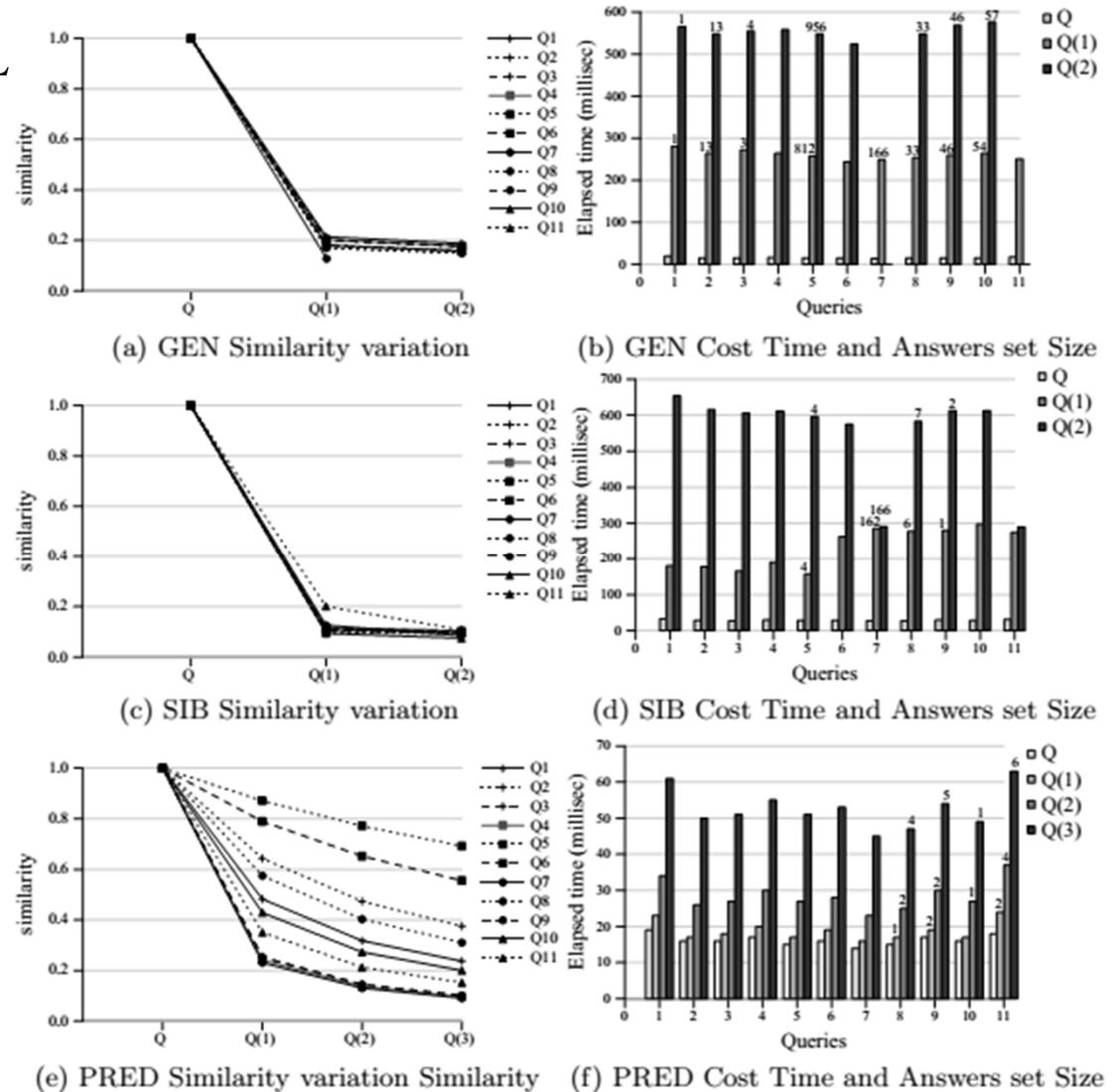
- Définition des opérateurs :
 - Clause de Relaxation: $APPROX(OP, TopK)$
 - Relaxation de prédicat : $PRED(Q, Prop, \epsilon)$
 - Généralisation: $GEN(Q, C, level)$
 - Substitution: $SIB(Q, C, [C_1, C_2, \dots, C_n])$
 - Agrégation des opérateurs : AND
 - **EX:** *Select Name From PhD Student Where duration = 12 APPROX (GEN (PhD Student, 1) AND (PRED (duration, 6))*

- Mesure de similarité :
 - **PRED** : Distance de Hausdorff
 - **GEN:** $Sim(C, C')$ entre deux classes
 - **SIB:** $Sim(C, C')$ entre deux classes
 - **AND:** min pour l'agrégation entre similarité de deux opérateurs
 - $Sat(h_i)$: évaluation de l'incidence de la relaxation sur chaque instance

FORMALISATION D'OPERATEURS DE RELAXATION

Intégration et expérimentation des opérateurs de relaxation dans les BDS

- Intégration dans OntoDB/OntoQL
- Expérimentation:
 - Relaxation de 11 requêtes par chaque opérateur
- Résultats
 - La répartition des données impact les performances de la relaxation
 - Processus itératif coûteux



□ Minimal Failing Subqueries (MFS)

- Requête formalisée sous forme de **conjonction de conditions (CC)** : $C_1 \wedge C_2 \wedge \dots C_3$
- Recherche de la plus petite sous liste de conditions qui annule la requête
- Utilisation de SPARQL pour la représentation de requête sous formes de (CC)
 - *Select ?name where {(?x, typeOf, PhD Student)(?x, uri_duration, '12 ')(?x, uri_name, 'Geraud')}*
- **Techniques de recherche de MFS**
 - MEL et ISHMAEL
 - JANNACH
- **Proposition d'algorithme de recherche de MFS**
- **Processus de relaxation avec utilisation des MFS**
 - Détection des MFS
 - Application des operateurs de relaxation adéquat :
 - ✓ Généralisation ou Substitution: $(?x, typeOf, Graduate Student)$
 - ✓ Relâchement de prédicat : $(?x, uri_years, '2st')$

□ Optimisation

- Relaxation \equiv Processus itératif
 - Utilisation des invariants du processus
 - $Answer(Q) \subset Answer(Q') \subset Answer(Q'')$
 - $Dist(Q, Q') < Dist(Q, Q'')$
- Estimation du **pas de relaxation** nécessaire pour avoir le Top-K des résultats
- Utilisation de la **sélectivité** des requêtes
- Construction d'un graphe de relaxation étiquette par la **similarité et la sélectivité**

□ Profils Utilisateurs

- Formalisation des profils utilisateurs pour guider la relaxation
 - Utilisation des Profils utilisateurs des systèmes de recommandation
- Définition d'un modèle de profil utilisateurs
- Possibilité d'adapter les Quadri-concepts ($\langle Produits, Tag, User, Profil \rangle$)

- Corby O., Dieng-Kuntz R., Faron-Zucker C., Gandon F. L., Searching the Semantic Web : Approximate Query Processing Based on Ontologies , IEEE Intelligent Systems, 2006.
- Dolog P., Stuckenschmidt H., Wache H., Diederich J., Relaxing RDF queries based on user and domain preferences, Journal of Intelligent Information System, 2009.
- Godfrey P., Minimization in cooperative response to failing database queries, IJCIS, 1997.
- Hogan A., Mellotte M., Powell G., Stampouli D., Towards Fuzzy Query-Relaxation for RDF, ESWC'12, 2012.
- Huang H., Liu C., Zhou X., Approximating query answering on RDF databases, Journal of World Wide Web, 2012.
- Hurtado C. A., Poulouvasilis A., Wood P. T., Query Relaxation in RDF, JODS, 2008.
- Poulouvasilis A., Wood P. T., Combining Approximation and Relaxation in Semantic Web Path Queries, Proceedings of the 9th International Semantic Web Conference (ISWC'10), 2010.
- Hai Huang, Chengfei Liu, and Xiaofang Zhou. Approximating query answering on rdf databases. World Wide Web, January 2012.
- Islam M. S., Liu C., Zhou R., On Modeling Query Refinement by Capturing User Intent Through Feedback, Proceedings of the Twenty-Third Australasian Database Conference - Volume 124, ADC '12, Australian Computer Society, Inc., Darlinghurst, Australia, Australia, 2012.
- Jannach D., ((Finding Preferred Query Relaxations in Content-Based Recommenders)), Intelligent Techniques and Tools for Novel System Architectures, vol. 109, Springer Berlin, Heidelberg, p. 81-97, September.
- Stéphane J., Allel H., Ammar M., Towards a Cooperative Query Language for Semantic Web Database Queries, On the Move to Meaningful Internet Systems : OTM 2013 Conferences, Springer Berlin Heidelberg, September

