

# SUJET DE THESE

## Requêtes Approximatives et Cooperatives d'Objets Complexes

Encadrement : Allel HADJALI et Ludovic LIETARD  
Thèse financée par la Région Bretagne & par le Projet-ANR AOC  
Laboratoire d'accueil : ENSSAT/IRISA – Lannion  
Durée : 03 ans

**Mots clés :** Bases de données, graphes, requête à graphe, appariement, réponses approximatives et coopératives

### **Description :**

Le sujet de thèse se situe dans le contexte de la recherche d'information dans des modèles décrivant des objets complexes tels que les documents structurés, des processus métiers ou des ontologies relatives à des domaines de connaissances. Ces modèles sont hétérogènes et riches du point de vue sémantique. Généralement, ils se basent sur une représentation sous forme de diagrammes ou de graphes dont l'interrogation se décline par un processus d'appariement avec un modèle cible qui s'appuie principalement sur des propriétés de structures (ou topologiques).

Du fait de la complexité des données recherchées, il est fréquent que la structure des documents cibles ne soit pas totalement connue par l'utilisateur. Il est alors nécessaire de pouvoir exprimer des requêtes avec un minimum de connaissances, ce qui amène à un appariement imprécis et les résultats de ce type de recherche sont plus ou moins en adéquation avec les besoins exprimés. L'approche fondée sur un appariement approximatif permet également de prendre en compte des éléments sémantiques en plus de l'appariement structurel et de pallier au problème des réponses vides ou au contraire de traiter des situations où les réponses sont pléthoriques.

L'originalité du travail proposé est de participer à la recherche méthodologique sur les méthodes d'appariement de graphes en y intégrant des notions de préférences et d'imprécision des informations. Cette thèse s'inscrit donc dans le cadre de notre activité au sein du projet ANR-AOC (dont la problématique principale est l'appariement d'objets complexes). Cette thèse pourra bénéficier de l'avancement des différents travaux en cours dans ce projet. Un second aspect majeur dans ce travail est le développement des techniques coopératives permettant au processus d'appariement de retourner les réponses servant le mieux les attentes de l'utilisateur. En cas de réponses insatisfaisantes ou contenant des informations indésirables, le système d'interrogation adopte un comportement coopératif et intelligent en analysant, par exemple, le but de la requête et en retournant des réponses généralisées ou approximatives.

L'objectif de ce travail est de mettre à la disposition des utilisateurs, d'une part, un langage de graphes requêtes souple, facilitant la prise en compte de la sémantique du modèle manipulé, et d'autre part, un système intelligent de recherche permettant de répondre à ce type de requêtes d'une manière coopérative. Les deux axes d'étude qui en découlent sont donc les suivants :

- Le premier axe porte sur les langages d'interrogation de modèles à base de graphes afin d'y introduire des notions de préférences et d'appariements approximatifs. Le problème de la caractérisation des informations est un point essentiel pour établir une correspondance entre un graphe représentant la requête de l'utilisateur et un graphe de la base de modèles. En effet, l'utilisateur n'a généralement qu'une connaissance partielle et assez imprécise des aspects syntaxiques, sémantiques et structurels des graphes qu'il recherche. Aussi, les systèmes d'interrogation doivent prendre en compte

des mécanismes de description imprécis et de filtrage approximatif ou flexible. Nous cherchons à montrer comment la connaissance, même imprécise, des objets peut enrichir la recherche par le contenu.

- Le second axe s'intéresse aux réponses coopératives à des requêtes définies sous forme de graphes et exprimant les besoins d'un utilisateur. L'objectif est de traiter le cas des réponses non satisfaisantes ou contenant des informations non désirées. Le système doit alors montrer un comportement intelligent et coopératif, afin de fournir les réponses qui correspondent le plus aux désirs de l'utilisateur. Cette fonctionnalité évoluée du système ne requiert pas d'effort supplémentaire de la part de l'utilisateur et n'induit aucun changement de la base de données interrogée. Notre démarche comporte deux approches : i) exploiter le contenu de la requête et proposer des mécanismes de relaxation / intensification ; ii) découvrir les connaissances implicites du domaine pour guider les mécanismes de réécriture de requêtes.