



## Offre de stage Thalès / LRI – Université Paris-Sud

### Modèle de filtrage de données

#### Thales Research & Technology (TRT):

Dir./Dépt. : Groupe de recherche Sciences et Techniques de l'Information

Service : Laboratoire Raisonnements et Analyses dans les Systèmes Complexes

**Personnes à contacter** : Gaëlle Lortal, [gaelle.lortal@thalesgroup.com](mailto:gaelle.lortal@thalesgroup.com), Claire Laudy, [claire.laudy@thalesgroup.com](mailto:claire.laudy@thalesgroup.com)

#### Université Paris-Sud :

Laboratoire : Laboratoire de Recherche en Informatique (LRI)

Equipe : LaHDAK, Large-scale Heterogeneous DATA and Knowledge

**Personne à contacter** : Yue Ma ([yue.ma@lri.fr](mailto:yue.ma@lri.fr))

**Durée du stage** : 6 mois

**A pourvoir** : dès que possible

**Lieu** : TRT (Palaiseau) et LRI-Université Paris-Sud (Orsay)

**Description** : Le groupe de recherche en Sciences et Techniques de l'Information de Thales Research & Technology (entité centrale de recherche de l'entreprise Thales) se compose de plusieurs laboratoires dont un spécialisé dans l'analyse et le raisonnement dans les systèmes complexes (LRASC).

Un des thèmes de LRASC est la fusion d'information. Du fait de l'existence de données massives provenant de sources hétérogènes, cette fusion doit être accompagnée d'un processus de filtrage pour obtenir des informations plus pertinentes. Au LRASC, un modèle d'une situation d'intérêt est utilisé pour filtrer les informations venant de différentes sources et ne garder que celles ayant un intérêt opérationnel. Ce modèle de situation a été rajouté en tant qu'entrée du processus de fusion.

Construire un modèle d'une situation d'intérêt requiert de l'expertise dans le domaine d'application concerné. Ainsi, les ingénieurs cogniticiens se basent sur la connaissance des utilisateurs (les opérationnels) et sur des connaissances expertes provenant de documents textes ou collectées à partir d'interviews. Lorsque les connaissances sont nombreuses, ce travail est toutefois laborieux.

Le but du stage est de proposer une solution automatisée pour aider à définir un tel modèle et éviter, ou minimiser, le recours à des ingénieurs cogniticiens. Les opérationnels ont de grosses difficultés à modéliser le domaine d'application sur lequel ils travaillent car ce modèle est implicite dans le processus de traitement des données. Ils ne connaissent que les grandes étapes des traitements (concepts de haut niveau), des parties de modèles plus ou moins génériques. A ces éléments sont toutefois associées des données qui peuvent être utilisées pour générer le modèle de situation requis.

Le stage s'appuiera sur l'étude de cas d'utilisation pour lesquels on dispose du résultat du filtrage à réaliser. Le travail demandé consiste à abstraire un modèle de situation à partir de ces résultats. Il sera nécessaire de définir le langage de représentation du modèle. Les langages de requêtes des systèmes d'accès aux données via des ontologies (Ontology-Based Data Access ou OBDA en anglais), sur lesquels travaille l'équipe LaHDAK (LRI), sont les langages visés dans ce stage. Ils permettent de générer une sous classe bien connue de requêtes SQL, appelées requêtes conjonctives (composées d'une conjonction d'atomes). L'implémentation de l'approche sous cette forme permettra de raisonner sur les connaissances du domaine (représentées dans une ontologie) et sur les données. Les réponses aux requêtes générées correspondront aux données filtrées par le modèle de situation préalablement défini, la présence de données indésirables devant entraîner la révision du modèle. Les opérationnels seront ainsi assistés pour valider le modèle de situation généré automatiquement.

Le travail réalisé durant le stage pourra être poursuivi dans le cadre d'une thèse de doctorat.

**Vos missions :**

L'objectif du stage est de proposer une approche permettant de générer un modèle de situation correspondant à une réelle expertise de filtrage de données.

Les différentes tâches à accomplir sont les suivantes :

- Modélisation ontologique d'une base de données existante ;
- Création d'un algorithme pour générer un modèle de situation ;
- Evaluation du modèle de situation créé sur des données des cas d'utilisation fournis.

**Rémunération :** Ce stage sera rémunéré par la société Thales.

**Votre profil:** Ingénierie informatique, Intelligence Artificielle

**Niveau d'étude :** Master2 Recherche ou équivalent

**Connaissances indispensables :** Bases théorique logique, un langage de programmation (Java, Python, ...)

**Connaissances souhaitables :** Programmation logique inductive, langage de modélisation d'ontologie (OWL), langage de requêtes, mécanismes de raisonnement.

**Connaissances à acquérir pendant le stage :** Construction d'ontologie, Systèmes OBDA, Fusion d'informations

**Langues :** Français/Anglais

**Contacts :**

Claire Laudy : [claire.laudy@thalesgroup.com](mailto:claire.laudy@thalesgroup.com)

Gaëlle Lortal : [gaelle.lortal@thalesgroup.com](mailto:gaelle.lortal@thalesgroup.com)

Yue Ma : [yue.ma@lri.fr](mailto:yue.ma@lri.fr)