



CONVENTION CIFRE¹ N° 2012/0305

ENTREPRISE : SAP Labs France SAS

DOCTORANT : Ahmad ASSAF

RAPPORT D'ACTIVITES A L'ISSUE DE LA 2^{ème} ANNEE DE THESE²

Si ce n'est déjà fait, pensez à joindre
la dernière attestation d'inscription en doctorat (2^{ème} ou 3^{ème} année)

Merci de présenter succinctement les travaux réalisés au cours de l'année écoulée. Préciser l'avancement des recherches par rapport au planning initial, l'évolution par rapport au projet initial et les éventuels problèmes rencontrés (intégration à l'équipe de recherche, adaptation aux contraintes du milieu industriel, ...).

For the past year, I have been actively consolidating my state of the art review that spans different fields ranging from Business Intelligence, Data Analysis and Data Integration in the Semantic Web. This has helped to narrow down the research problems tackled in this thesis as well as clearly define the thesis contribution.

I have participated in an internal project at SAP called **Business Intelligence Graph (BIG)**. The Business Intelligence Graph is a set of foundation services for Business Intelligence (BI) applications such as SAP Lumira that aims to simplify the experience of Decision Makers and Analysts. With the BI Knowledge Graph (BIG), we harvest BI artefacts, BI usage and user profiles and store them as a graph to enable a set of recommendation and ranking services.

Last year, my main research direction was the enrichment of enterprise data with semantics and public open datasets, which in turn, enables better data integration thus resulting in more informed business decisions. It has evolved into a framework that will enable **self-service data provisioning through semantic enrichment of data**. Our goal is to provide a mechanism that annotates and profiles tabular data to provide better dataset description. Furthermore, we aim to aggregate the enhanced datasets descriptions so that people can search and browse through content. We also aim at providing ranking mechanism that leverages a comprehensive data quality metric and license information attached to the dataset description.

Merci de préciser la répartition du temps passé en entreprise et au laboratoire.

Cette répartition vous convient-elle ? Va-t-elle évoluer ?

Rappeler les interruptions longues de travail (ex : congé maternité, accident, ...)

I have spent most of my time in SAP while meeting the research group in EURECOM on a weekly basis. I believe that for the next year, this distribution is expected to remain the same although from the beginning of next year, I will probably be spending most of my time in EURECOM as I will focus on writing the thesis where I will need more direct interaction with my PhD supervisor.

Avez-vous bénéficié de formation(s) au cours de l'année écoulée ? Merci d'en préciser le thème et l'organisateur.

None

Vos travaux ont-ils donné lieu à publication, présentation à un congrès, poster ou encore à dépôt de brevet au cours de l'année écoulée ? Merci de préciser notamment le caractère international des communications.

¹ Rappeler le numéro de convention (année de la demande / numéro de dossier)

² Rayer la mention inutile. Attention, dans certains cas particuliers, les conditions générales d'octroi stipulent qu'un rapport doit être fourni tous les 6 mois

- Ahmad Assaf, Ghislain Auguste Atemezang, Raphaël Troncy, Elena Cabrio: What Are the Important Properties of an Entity? - Comparing Users and Knowledge Graph Point of View. ESWC (Satellite Events) 2014: 190-194

Quelles sont vos perspectives pour l'année à venir ?

I am currently in the process of finalizing the implementation of the framework I proposed in my thesis, and validating it with the enterprise. I will also start writing the manuscript in the beginning of 2015 with the goal to finish my thesis by the summer 2015.

Dans l'état actuel d'avancement de vos travaux, envisagez-vous de soutenir votre thèse dans les délais impartis ?

Yes, I expect to defend my thesis in the summer of 2015.

Commentaires et appréciations du responsable scientifique dans l'entreprise




Dr. Aline Senart

Après une étude bibliographie très poussée, Ahmad a réussi au cours de cette année à focaliser sa recherche sur une des pistes les plus prometteuses de ses investigations. Il a proposé une solution innovante autour d'un canevas logiciel en libre-service pour l'accès et l'enrichissement de données ouvertes ou de données de l'entreprise. L'implémentation de sa solution est déjà bien avancée et devrait faire lieu à une évaluation dans les mois qui arrivent. Au cours de ces deux dernières années, Ahmad a su rester actif au niveau de la communauté en publiant régulièrement ses résultats. Ahmad est très sérieux et appliqué et je n'ai aucun doute sur le fait qu'il devrait finir ses travaux dans les délais impartis.

Commentaires et appréciations du directeur de thèse

Dr. Raphael Troncy

Ahmad a fait une très bonne deuxième année de thèse, s'adaptant aux nombreuses restructurations qui s'opèrent en milieu industriel. Une soutenance à mi-parcours, conformément aux règles de l'école doctorale de l'EDITE s'est déroulée au mois d'avril 2014 où un jury extérieur a confirmé les contributions innovantes de cette thèse. Les travaux d'Ahmad ont donné lieu à d'autres publications et une publication de référence est attendue pour le début de l'année 2015. La thèse d'Ahmad couvre plusieurs domaines (intégration de données, business intelligence, fouille de données, visualisation de données) et Ahmad a poursuivi un travail important de bibliographie de ces différents champs disciplinaires. L'objectif de la troisième année est désormais de finaliser l'implémentation de l'architecture logicielle proposée, et de la valider avec des données d'entreprise.

Nom du doctorant Ahmad Assaf	Nom et fonction du tuteur industriel Dr Aline Senart	Nom et titre du tuteur de thèse Dr Raphael Troncy
Date et signature 9/12/2014 	Date, cachet et signature ³ 9/12/2014 	Date, cachet et signature ³ 09/12/2014 

Ce document est à retourner à : ANRT – Service CIFRE, 41 bd des Capucines, 75002 Paris.

³ Conformément à l'article 7 des conditions générales d'octroi d'une CIFRE, si le rapport n'est pas signé par les 2 tuteurs, il ne sera pas valide. En conséquence, le paiement de la subvention sera interrompu.