

LABORATOIRE



INFORMATIQUE, SIGNAUX ET SYSTEMES
DE SOPHIA ANTIPOLIS
UMR 6070

SELECTION SEMANTIQUE DE SERVICES POUR DISPOSITIF:
GESTION DYNAMIQUE DE LA BASE DE CONNAISSANCE

DESCRIPTION OF WORK

Gérald Rocher, master 2 IFI/IAM

Stage de fin d'étude (recherche)
Mars/Septembre 2015



Encadreur du stage pour le laboratoire I3S

Jean-Yves Tigli

Tuteur enseignant pour Polytech'Nice Sophia

Isabelle Mirbel

RESUME :

Dans les systèmes informatiques ambiants, les applications logicielles sont composées à partir d'une sélection de services logiciels intégrés à des dispositifs et des objets de notre vie quotidienne qui, devenant communicants (Internet of Things), peuvent dès lors, être observés et contrôlés. Ces objets et ces dispositifs, fixes ou mobiles, sont soumis aux phénomènes physiques de l'environnement réel dans lequel ils sont placés ce qui impliquent une évolution de leur disponibilité dans le temps et dans l'espace. Il est donc primordial que le mécanisme de sélection de services, au-delà des fonctionnalités offertes par les services, tienne compte également de leur dynamique physique inhérente à leur intégration dans le monde réel. Lors de premiers travaux [9], l'utilisation des standards du web sémantique (Web of Things) a été étudiée pour obtenir, à partir d'annotations sémantiques formelles sur les dispositifs et les services, une représentation dynamique et incrémentale des connaissances terminologiques et contextuelles qui leur est associée et permettre la gestion de leur évolution dans le temps et dans l'espace.

Dans le cadre de ce stage de recherche, nous allons, dans un premier temps, nous concentrer sur l'aspect incrémental des connaissances terminologiques dans la base de connaissance et sur son intérêt dans le but d'améliorer la pertinence des services sélectionnés. A partir d'une démarche scientifique rigoureuse, l'objectif va être la publication d'un papier exposant notre contribution qui sera alors confrontée avec les chercheurs du domaine. Dans un second temps, nous allons tenter d'améliorer le mécanisme de sélection de services en lui permettant de maîtriser l'évolution de la connaissance terminologique dans le temps sans dégrader la pertinence des services sélectionnés.

MOTS CLES :

Web sémantique, représentation de la connaissance, Contexte, Informatique ambiante, Web des objets

Table des matières

I.	DESCRIPTION DU PROJET.....	3
A.	Contexte	3
B.	Motivations et objectifs.....	5
II.	ELEMENTS DE L'ETAT DE L'ART	6
A.	Sur les approches d'utilisation des ontologies	6
B.	Sur la maîtrise du volume des connaissances terminologiques.....	7
III.	EVALUATION.....	7
IV.	DECOUPAGE EN LOTS ET PLANNING PREVISIONNEL	8
C.	Article relatif à l'apport de l'enrichissement de la connaissance terminologique	8
i.	Définition des tâches.....	8
ii.	Définition des livrables	9
D.	Travail relatif à la maîtrise du volume des connaissances terminologiques	9
i.	Définition des tâches.....	9
ii.	Définition des livrables	9
E.	Activités annexes	9
i.	Définition des tâches.....	9
ii.	Définition des livrables	10
iii.	Planification des congés.....	10
V.	REFERENCES	10

I. DESCRIPTION DU PROJET

A. Contexte

L'informatique ambiante [1] est un terme qui désigne l'intégration du monde physique réel dans le monde numérique d'internet. Cela concerne tous les *objets* de notre vie quotidienne (chaise, table, lampe, etc...) ou environnements physiques (ville, immeuble, véhicule, espace médicalisé,...), statiques ou dynamiques qui deviennent, à partir d'applications logicielles, observables et contrôlables [2] (*Internet des objets, IoT*). Cela est rendu possible grâce à des *dispositifs* communicants embarqués dans les *objets* ou placés dans l'*environnement* [3]. Ces *dispositifs* mettent en œuvre des *ressources* qui permettent d'interagir avec les *objets* (actionneur) et d'obtenir des informations (ID, description, observations de capteurs,...) sur eux-mêmes, sur les *objets* ou l'environnement auxquels ils se rattachent. L'accès à ces *ressources* est réalisé par l'intermédiaire de *services* logiciels qui exposent leur interface et permettent la communication avec le monde numérique. Ces services sont alors en mesure de participer à l'élaboration d'applications logicielles complexes qui doivent, parmi tous les services disponibles, *sélectionner* et *orchestrer* ceux qui vont permettre **d'obtenir une fonctionnalité qui réponde aux besoins des utilisateurs**.

Le caractère dynamique des dispositifs, des objets et des environnements physiques implique que les applications logicielles, pour conserver leur fonctionnalité et **assurer une continuité de service, doivent dynamiquement s'adapter aux variations du monde réel** (la disponibilité et l'état des services évoluant dans le temps et dans l'espace). Les applications sont alors dites sensibles au contexte. Ainsi, l'élaboration d'applications logicielles sensibles au contexte est le fruit d'une composition de services au sein d'une boucle d'auto-adaptation dynamique (*Figure 1*).

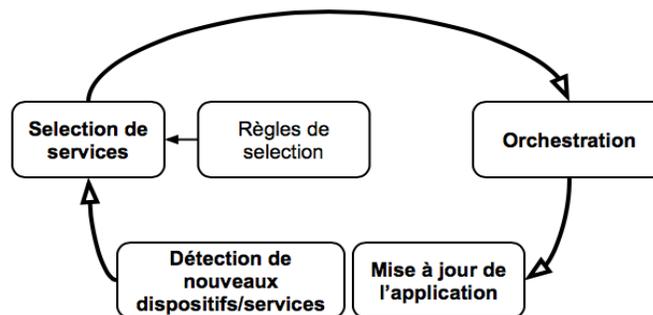


Figure 1 : Boucle d'adaptation dynamique d'une application

Au sein de cette boucle d'auto-adaptation dynamique, le mécanisme de sélection des services est donc primordial : il doit sélectionner, parmi tous les dispositifs disponibles, ceux les plus *pertinents* qui vont permettre la réalisation d'un objectif précis, en *cohérence* avec le contexte d'exécution.

Dans ce cadre, un premier travail au sein du laboratoire I3S [9] a permis, à partir d'annotations sémantiques formelles (*Semantic Web Of Things, SWoT*) sur les dispositifs et les services, basées sur le standard OWL (Web Ontology Language) [4], de construire une représentation dynamique de la connaissance fonctionnelle et contextuelle des dispositifs et des services.

Cette représentation dynamique de la connaissance se décompose en trois niveaux distincts (*Figure 2*):

1. Les annotations sémantiques formelles qui:

- a. Fournissent d'une part, une description formelle de la connaissance sur les concepts du domaine de chacun des dispositifs et des services, et, d'autre part, une description formelle des observations contextuelles sur l'environnement (à partir de capteurs extéroceptifs) ou sur les objets (à partir de capteurs proprioceptifs),

- b. Fournissent dynamiquement, sous forme de propriétés contextuelles, les valeurs observées à partir des capteurs extéroceptifs et proprioceptifs. Les propriétés extéroceptives sont alors propagées à tous les dispositifs placés dans le même environnement (graphe de couplage des dispositifs).
2. **Les éléments terminologiques *TBox*** de chaque domaine dans la base de connaissance (BdC), qui sont ajoutés lors de la découverte de nouveaux dispositifs,
3. **Les assertions *ABox*** (faits et individus) sur tous les domaines dans la BdC, qui sont ajoutés lors de la découverte de nouveaux dispositifs et supprimés lors de leur disparition.

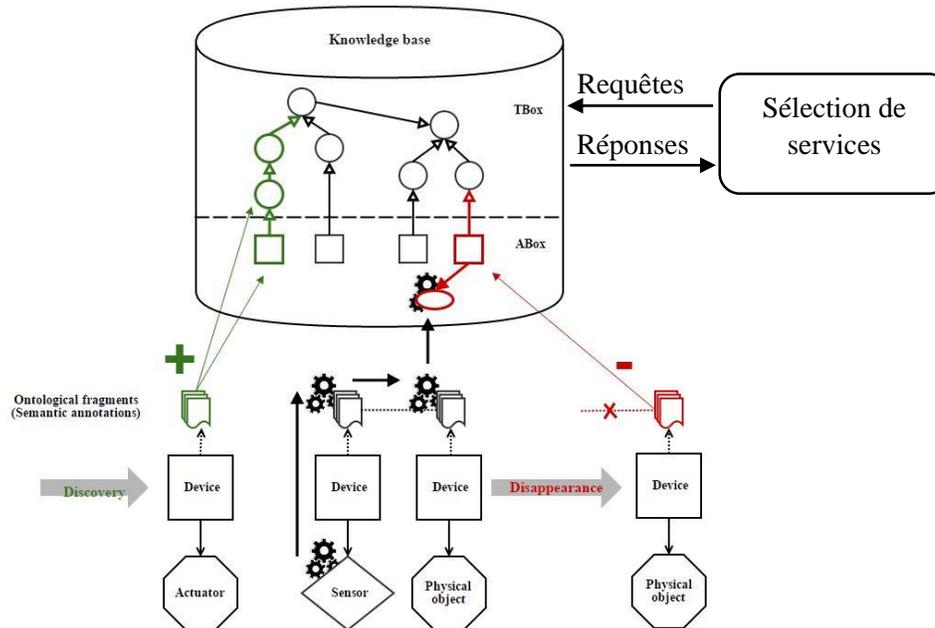


Figure 2 : Niveaux de dynamisme dans SWoT

Cette description de la connaissance permet d'obtenir une représentation dynamique du contexte à partir de laquelle il est possible de déduire, d'une part, la pertinence des dispositifs et des services par rapport à un objectif fonctionnel que l'application doit atteindre, et, d'autre part, leur utilisabilité, c'est-à-dire leur disponibilité dans le temps et dans l'espace.

En effet, la constante augmentation des connaissances terminologiques dans la BdC permet d'accroître la pertinence des dispositifs et des services sélectionnés. Par ailleurs, la propagation des propriétés contextuelles à tous les dispositifs placés dans le même environnement et l'ajout d'une représentation contextualisée de la connaissance à partir de l'approche « context slices » [5] permettent de raisonner sur la validité des dispositifs et des services dans le temps, dans l'espace ou toutes autres dimensions physiques pertinentes dans le cadre d'objets et d'environnements réels (température, qualité de service, etc...). Cela permet d'augmenter la cohérence des services sélectionnés avec le contexte d'exécution de l'application.

A partir de ce travail, un modèle représentationnel et évolutif de la dynamique dans WoT a été établi (Figure 3) qui permet de mettre en lumière certaines problématiques et contraintes liées à l'approche suivie. Dans le cadre de ce projet, nous allons concentrer nos recherches aux problématiques liées à la gestion de la diversité sémantique et l'évolutivité de la connaissance dans la BdC.

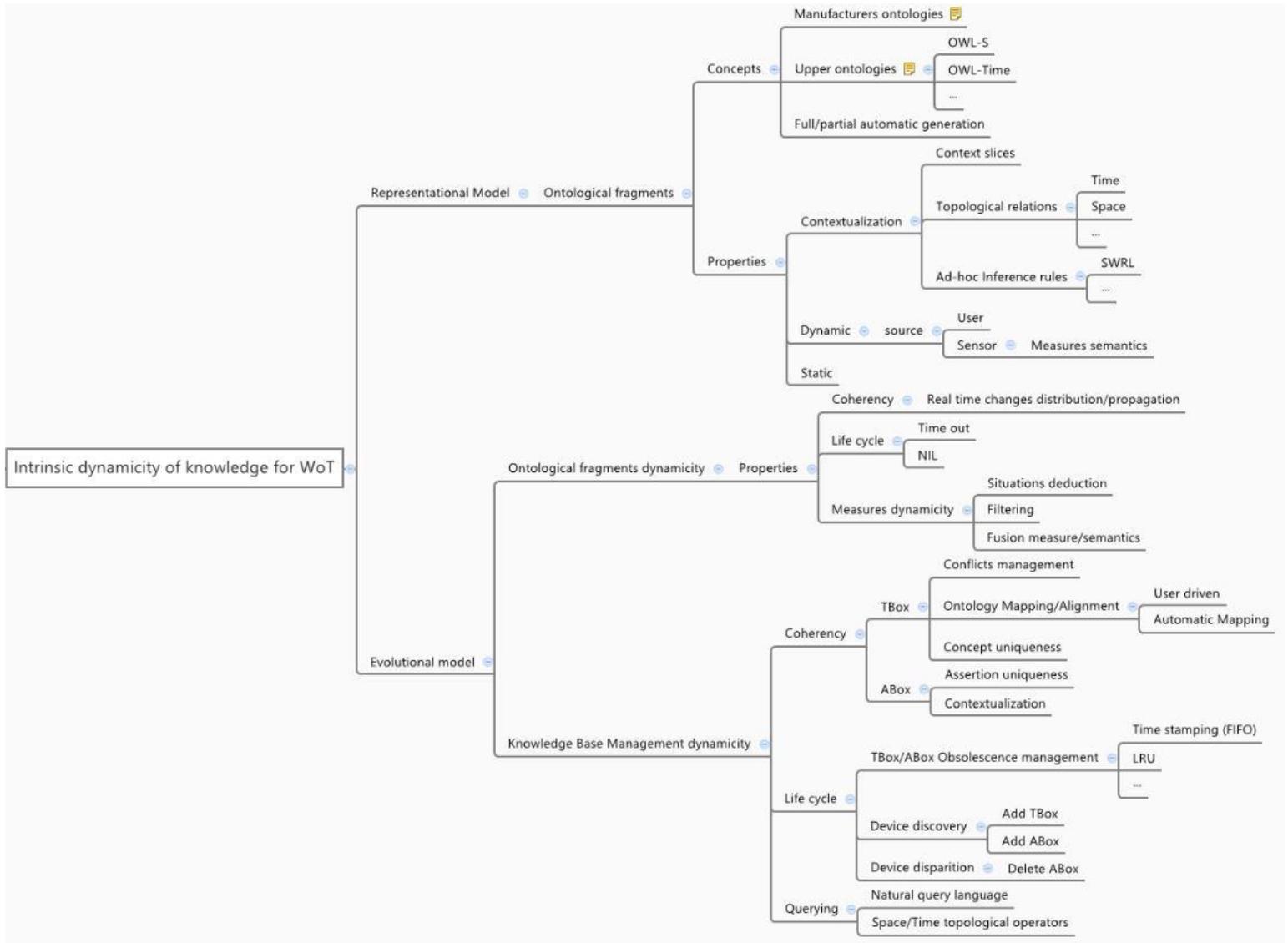


Figure 3 : Modélisation de la dynamique dans WoT

B. Motivations et objectifs

Avec l'approche suivie dans [9], chaque dispositif fournit une description sémantique formelle du domaine qui lui est propre. Une des contributions de cette approche est, d'une part, de permettre de gérer la forte diversité sémantique inhérente à la multiplicité des dispositifs mis en œuvre dans le domaine de l'informatique ambiante, et, d'autre part, de permettre l'évolutivité progressive de la connaissance terminologique dans la BdC dont l'intérêt est l'amélioration de la pertinence des services sélectionnés. Cette approche, cependant, implique des problématiques liées à l'alignement d'ontologies, à la gestion des conflits (un même concept avec des significations sémantiques différentes, etc...). Il conviendra donc d'étudier les différentes approches d'utilisations d'ontologies pour identifier celles qui permettent de répondre aux contraintes d'évolutivité et de diversité sémantique tout en évitant les problématiques citées

précédemment. En suivant une démarche scientifique rigoureuse, l'objectif va être la publication d'un papier exposant notre contribution qui sera alors confrontée avec les chercheurs du domaine.

Par ailleurs, bien qu'évolutive, il est peu probable que la connaissance dans la BdC puisse augmenter indéfiniment. Des limitations dans l'espace (mémoire système) et dans le temps (temps nécessaire pour répondre aux requêtes) sont susceptibles d'apparaître.

Il est donc nécessaire de trouver un compromis entre la gestion de la diversité sémantique (à l'origine de l'augmentation continue de la connaissance terminologique *TBox* et donc de la pertinence des services sélectionnés), les capacités intrinsèques du système informatique lié à l'application logicielle (mémoire) et, enfin, de la satisfaction de l'utilisateur liée aux temps nécessaires pour répondre aux requêtes.

La question de recherche sous-jacente pourrait être formulée de la manière suivante :

«Comment maîtriser le volume de connaissance terminologique dans la BdC tout en assurant l'optimalité de la pertinence des services sélectionnés? »

II. ELEMENTS DE L'ETAT DE L'ART

A. Sur les approches d'utilisation des ontologies

Dans l'approche suivie dans [9], les annotations sémantiques sont des ontologies qui définissent complètement les concepts utilisés pour décrire la connaissance du domaine de chaque dispositif et leurs services. Par ailleurs, des ontologies de plus haut niveau sont utilisées (OWL-S, OWL-Time, etc...). Ces ontologies définissent un vocabulaire et des concepts assez généraux pour pouvoir être utilisés dans n'importe quelle sous-ontologie. Comme expliqué précédemment, cette approche, implique des problématiques liées à l'alignement d'ontologies et à la gestion de conflits. Il nous faut donc tenter d'identifier une approche qui réponde aux contraintes d'évolutivité et de diversité sémantique tout en évitant ces problématiques. Dans [6], les auteurs identifient trois approches distinctes de mise en œuvre des ontologies qu'il conviendra d'étudier:

1. Approche basée sur une ontologie globale

Avec cette approche, une seule ontologie est partagée et décrit la totalité des concepts de tous les domaines. Cette approche apporte une solution aux problématiques de gestion des conflits et d'alignement des ontologies. Du point de vue d'IoT, cela signifie que les concepts apportés par chacun des dispositifs sont issus de cette ontologie globale. Les auteurs indiquent qu'il est envisageable que l'ontologie globale soit l'agrégation de plusieurs sous-ontologies spécialisées.

Des exemples d'ontologies globales peuvent être trouvés dans la littérature. Par exemple :

Ontologie	Référence
OpenCyc	http://sw.opencyc.org/
YAGO	http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/
CEO consumer electronic ontology	http://www.ebusiness-unibw.org/ontologies/consumerelectronics/v1
MMI device Ontology	https://marinemetadata.org/community/teams/ontdevices

Tableau 1: Exemples d'ontologies globales

2. Approche basée sur des ontologies multiples

Cette approche est celle suivie dans [9]. Chaque domaine est décrit par son ontologie propre sans assurance du partage des concepts, donc avec les problématiques de gestion de conflits et d'alignements évoqués précédemment.

3. Approche hybride

Avec cette approche, chaque domaine est décrit avec son ontologie propre mais à partir d'un vocabulaire commun à toutes les ontologies. Cette approche est partiellement suivie dans [9] avec l'utilisation des ontologies OWL-S et OWL-Time définissant respectivement les concepts relatifs aux dispositifs et aux services ainsi qu'au temps.

Des exemples d'ontologies définissant un vocabulaire commun peuvent être trouvés dans la littérature. Par exemple :

Ontologie	Référence
OWL-Time	http://www.w3.org/TR/owl-time/
OWL-S	http://www.w3.org/Submission/OWL-S/
Semantic sensor ontology (W3C SSN)	http://www.w3.org/2005/Incubator/ssn/ssnx/ssn
SENSEI Observation and measurement ontology	http://personal.ee.surrey.ac.uk/Personal/P.B.arnaghi/doc/SENSEI_O&M_Ontology.pdf
BFO	http://ifomis.uni-saarland.de/bfo/

Tableau 2 : Exemples d'ontologies définissant un vocabulaire et des concepts généraux

B. Sur la maîtrise du volume des connaissances terminologiques

La croissance de la connaissance dans la BdC peut être maîtrisée via, par exemple, la mise en place d'un système de gestion de l'obsolescence des éléments terminologiques. Pour ce faire, il peut être envisagé d'appliquer des mécanismes de gestion de la mémoire tels ceux utilisés dans les systèmes d'exploitation, par exemple les algorithmes FIFO (First In First Out), LRU (Least Recently Used), etc... [7]

Cependant, de tels algorithmes, bien adaptés à la gestion de la mémoire dans les systèmes informatiques, ne le sont pas forcément quand ils sont appliqués à la gestion de la connaissance dans une BdC. Il est donc souhaitable de modéliser les problématiques associées au domaine de l'informatique ambiante, à savoir: assurer la pertinence de la sélection des services à partir de la connaissance disponible à chaque instant et ce, avec un temps de traitement des requêtes acceptable du point de vue l'utilisateur final de l'application. Il faut donc être capable de mesurer l'impact de chaque concept terminologique (ou groupe de concepts) sur l'ensemble de la connaissance. De ce point de vue, les algorithmes mis en œuvre dans l'analyse des réseaux sociaux peuvent être étudiés (clusters, etc...) [8].

III. EVALUATION

L'évaluation des différentes approches nécessite la mise en place de mesures sur le contenu de la BdC (le contenu se découpe en deux types d'éléments : les éléments terminologiques *TBox* et les assertions *ABox*), de mesures sur les temps de traitement des requêtes ainsi que l'empreinte mémoire de la BdC. La mesure de la pertinence des services sélectionnés ne peut pas être mesurée numériquement. La mise en place d'un protocole expérimental de validation, à partir de contenus ontologiques prédéfinis, pour lesquels les résultats attendus sont connus est alors nécessaire.

IV. DECOUPAGE EN LOTS ET PLANNING PREVISIONNEL

Le travail de ce projet s'inscrit dans la continuité du PFE exposé dans [9]. Nous allons, dans un premier temps, nous concentrer sur l'aspect incrémental des connaissances terminologiques dans la base de connaissance et sur son intérêt dans le but d'améliorer la pertinence des services sélectionnés. A partir d'une démarche scientifique rigoureuse et de mesures qui viendront étayer nos résultats, l'objectif va être la publication d'un papier exposant notre contribution qui sera alors confrontée avec les chercheurs du domaine. Dans un second temps, nous allons tenter d'améliorer le mécanisme de sélection de services en lui permettant de maîtriser l'évolution de la connaissance terminologique dans le temps sans dégrader la pertinence des services sélectionnés.

C. Article relatif à l'apport de l'enrichissement de la connaissance terminologique

i. Définition des tâches

Tâche#	Dénomination	Début	Fin	Commentaires
T_1.1	Ajout des mesures des paramètres permettant de caractériser : <ol style="list-style-type: none"> 1. le contenu <i>TBox/ABox</i> de la BdC et son évolution dans le temps, 2. Les temps de traitement des requêtes, 3. La mémoire système utilisée. 	16/03	3/04	Prérequis à la publication des résultats.
T_1.2	Compléter et archiver (I3S) le rapport du projet de fin d'étude	23/03	10/04	<ol style="list-style-type: none"> 1. Traduction en anglais 2. Ajout des résultats préliminaires montrant l'évolution de la connaissance dans la BdC au cours du temps. 3. Compléments bibliographiques sur des travaux de recherche du domaine de IoT adressant les mêmes problématiques scientifiques.
T_1.3	Etude bibliographique sur les différentes approches d'utilisation des ontologies	13/04	17/04	
T_1.4	Recherche de benchmarks et mise en place d'un protocole expérimental de validation de l'amélioration de la pertinence des services sélectionnés (L_1.1)	20/04	24/04	A partir de travaux de recherches dans le domaine.
T_1.5	Mise en place de dispositifs avec des annotations sémantiques formelles décrivant les concepts terminologiques	27/04	28/04	<ol style="list-style-type: none"> 1. Enrichissement dynamique de la BdC. 2. Caractérisation de l'évolution du contenu terminologique de la BdC dans le temps en fonction de l'apparition et de la disparition des dispositifs.
T_1.6	Acquisition des résultats	29/04	4/05	
T_1.7	Ecriture de la publication (L_1.2)	5/05	22/05	<u>Contributions</u> : l'apport en connaissance d'un dispositif permet à un ancien dispositif non sélectionné sur une précédente requête de l'être maintenant.
T_1.8	Relecture et correction de la publication	25/05	29/05	

ii. Définition des livrables

Livable#	Dénomination	Deadline	Commentaire
L_1.1	Protocole de validation	1/05	
L_1.2	Papier (poster ou communication workshop WoT)	29/05	Ubicomp 2015

D. Travail relatif à la maîtrise du volume des connaissances terminologiques

i. Définition des tâches

Tâche#	Dénomination	Début	Fin	Commentaires
T_2.1	Etude bibliographique (L_2.1) sur la question de recherche	1/06	26/06	Etude des résultats scientifiques existants.
T_2.1.1	<i>Algorithmes de gestion de mémoire</i>	1/06	12/06	
T_2.1.2	<i>Algorithmes appliqués aux réseaux sociaux</i>	15/06	26/06	Pour l'identification des concepts à conserver ou pas.
T_2.2	Implémentation de l'algorithme identifié	29/06	17/07	
T_2.3	Mesures d'évolution de l'empreinte mémoire dans le temps ainsi que des temps de réponse aux requêtes + pertinence des services sélectionnés à partir du protocole expérimental établi (T_1.4).	20/07	30/07	Validation de l'algorithme.
T_2.4	Intégration de la solution pour la sélection sémantique des services dans la boucle d'auto-adaptation (Figure 1)	4/08	7/08	
T_2.5	Benchmarks de la solution complète	10/08	14/08	Etude de l'impact sur la durée du cycle d'adaptation

ii. Définition des livrables

Livable#	Dénomination	Date	Commentaires
L_2.1	SLR	26/06	Etat de l'art détaillé et ébauche de la solution envisagée pour répondre à la question de recherche.

E. Activités annexes

i. Définition des tâches

Tâche#	Dénomination	Début	Fin	Commentaires
T_3.1	Management	16/03	2/09	Suivi d'avancement du projet (Réunion hebdomadaire)
T_3.2	Ecriture du rapport d'activité intermédiaire	5/05	10/05	15/20 pages (Sur la base du livrable L_1.2)
T_3.3	Ecriture du rapport final (L_3.2)	17/08	1/09	<u>Contributions :</u> Maîtrise du volume de la connaissance terminologique dans la BdC et des temps de traitement des requêtes.
T_3.4	Présentation finale (L_3.3)	17/08	1/09	

ii. Définition des livrables

Livable#	Dénomination	Date	Commentaires
L_3.1	Rapport intermédiaire	10/05	
L_3.2	Rapport final	2/09	40 pages maximum
L_3.3	Slides	2/09	

iii. Planification des congés

Le vendredi 31 juillet et le lundi 3 aout.

V. REFERENCES

- [1] Weiser, M. (1991). The computer for the 21st century. Scientific american, 265(3), 94-104.
- [2] Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. Computer networks, 54(15), 2787-2805.
- [3] Haller, S. (2010). The things in the internet of things. Poster at the (IoT 2010). Tokyo, Japan.
- [4] McGuinness, D. L., & Van Harmelen, F. (2004). OWL web ontology language overview. W3C recommendation, 10(10), 2004.
- [5] Welty, C. (2010). Context Slices: Representing Contexts in OWL. In Proceedings of the 2nd Workshop on Ontology Patterns (WOP 2010).
- [6] Wache, H., Voegelé, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., & Hübner, S. (2001, August). Ontology-based integration of information-a survey of existing approaches. In IJCAI-01 workshop: ontologies and information sharing (Vol. 2001, pp. 108-117).
- [7] Jones, R., & Lins, R. D. (1996). Garbage collection: algorithms for automatic dynamic memory management.
- [8] Carrington, P. J., Scott, J., & Wasserman, S. (Eds.). (2005). Models and methods in social network analysis (Vol. 28). Cambridge university press.
- [9] G.Rocher, R.Daiki (2015) Semantic selection of services for device: dynamic knowledge base management.