



HAL
open science

Le Web Sémantique : une solution fédératrice pour le traitement de l'information

Sébastien Laborie

► **To cite this version:**

Sébastien Laborie. Le Web Sémantique : une solution fédératrice pour le traitement de l'information. Informatique [cs]. Université de Pau et des Pays de l'Adour, 2024. tel-04701289

HAL Id: tel-04701289

<https://univ-pau.hal.science/tel-04701289v1>

Submitted on 18 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HABILITATION À DIRIGER DES RECHERCHES

Le Web Sémantique : une solution fédératrice pour le traitement de l'information

Sébastien LABORIE

JURY

Président de jury	Prof. Christophe Nicolle	Professeur des Universités (CIAD - Dijon)
Rapportrice	Prof. Nathalie Aussenac-Gilles	Directrice de Recherche (IRIT - Toulouse)
Rapportrice	Prof. Marie-Hélène Abel	Professeure des Universités (HEUDIASYC - Compiègne)
Rapporteur	Prof. Jérôme Gensel	Professeur des Universités (LIG - Grenoble)
Examineur	Prof. Philippe Roose	Professeur des Universités (UPPA - Anglet)
Examineur	Dr. Christian Sallaberry	Maître de Conférences HdR (UPPA - Pau)

29 août 2024

Remerciements

En premier lieu, je souhaite remercier mes collègues Philippe Roose et Christian Sallaberry pour les relectures régulières ainsi que pour leurs précieux commentaires et conseils lors de l'élaboration de ce manuscrit. Ils m'ont aussi beaucoup encouragé durant la rédaction et je leur en suis extrêmement reconnaissant.

Je tiens ensuite à remercier tous les membres de mon jury d'habilitation. J'adresse mes plus profonds remerciements à mes trois rapporteurs : Nathalie Aussenac-Gilles, Marie-Hélène Abel et Jérôme Gensel. C'est un grand honneur pour moi que vous ayez accepté d'être rapporteurs de mon HDR et je vous remercie d'avoir pris de votre temps pour évaluer ce manuscrit. Enfin, je remercie sincèrement Christophe Nicolle d'avoir accepté de présider ce jury.

Toutes mes pensées vont aussi vers les collègues de mon laboratoire, et plus particulièrement les membres de l'équipe T2i : Marie-Noëlle Bessagnet, Pantxika Dagoret, Yon Dourisboure, Patrick Etcheverry, Annig Le Parc Lacayrelle, Philippe Lopisteguy, Franck Luthon, Christophe Marquesuzaà, Thierry Nodenot, Philippe Roose, Christian Sallaberry et Nathalie Vallès-Parlangeau. Je suis le petit jeune de la bande depuis toutes ces années et vous m'avez toujours fait confiance avec joie et bonne humeur.

Je pense également à tous les chercheurs et à toutes les personnes avec qui j'ai pu collaborer depuis mes débuts en Recherche. Cela fait 20 ans depuis le début de ma thèse en 2004! Le temps passe si vite. Évidemment, j'ai une pensée pour les doctorants que j'ai pu accompagner durant leur thèse : Nhan, Regina, Irvin, Asma, Nathalie, Nabila, Jorge, Charlotte et Yunji. Vous voir vous épanouir et devenir progressivement Chercheur a toujours été pour moi un réel privilège.

Je remercie l'IUT de Bayonne et du Pays Basque, le LIUPPA et l'Université de Pau et des Pays de l'Adour de nous permettre de nous épanouir au quotidien dans un environnement si agréable sur la Côte Basque. Les équipes pédagogiques, les gestionnaires ainsi que les membres de tous les services contribuent aussi à cet épanouissement.

Enfin, je remercie ma famille qui m'encouragent dans tous mes projets. Je pense aux enfants (Julia, Mickaël et Johann) qui me voient peut-être trop souvent rivé sur l'ordi et évidemment à ma chérie Émilie qui a toujours su me faire rire aux éclats.

Merci à toutes et à tous!

Résumé

Depuis plusieurs années, le Web Sémantique a su évoluer pour proposer des technologies permettant de décrire, de raisonner et d'interroger des descriptions sémantiques. Mes travaux de recherche depuis une dizaine d'années ont permis de démontrer qu'il est une solution fédératrice pour le traitement des informations. En particulier, nos contributions en lien avec le Web Sémantique se situent selon 3 axes, les contributions (1) qui sont centrées sur des technologies du Web Sémantique, (2) celles qui exploitent le Web Sémantique pour renforcer un domaine dédié et (3) celles qui rapprochent différentes communautés grâce au Web Sémantique.

Le premier point met en avant la prise de recul à avoir sur l'usage des technologies du Web Sémantique et notamment au sujet de son formalisme de base qui est RDF. En effet, je démontre qu'une description RDF peut contenir des redondances d'informations ou bien être encodée de façon très verbeuse. Ce phénomène s'accroît lorsque différentes descriptions sont combinées entre elles. Pour des applications qui intègrent de telles descriptions sémantiques, ceci peut s'avérer néfaste que ce soit en terme de stockage, de chargement ou de traitement des informations. Ainsi, nous proposons un processus de normalisation de descriptions RDF qui selon un contexte cible nettoie à la fois la structuration de la description mais aussi sa sérialisation.

Le second point illustre une exploitation des technologies issues du Web Sémantique pour l'adaptation de contenus multimédias. La communauté du Web Sémantique a proposé au travers d'une recommandation du W3C une modélisation possible des profils. La communauté du Multimédia a su prendre en considération ce standard de description pour l'adaptation. Notre prise de recul a permis d'identifier plusieurs limites aux propositions existantes. Ainsi, nous avons développé une nouvelle modélisation sémantique des profils à base de services qui exploite pleinement le formalisme RDF. Nous montrons que cette nouvelle vision du profil peut inspirer à la fois la communauté du Web Sémantique et celle du Multimédia. Nous qualifions ce principe de gagnant-gagnant.

Enfin, le troisième point démontre que le Web Sémantique est une solution pour fédérer des données hétérogènes issues de différentes communautés. Pour réaliser cette démonstration, nous présentons le cas d'une recherche d'explications d'événements au sein d'un réseau de capteurs. Cette situation est intéressante puisqu'elle permet de

s'appuyer sur un environnement que nous qualifions d'hybride car il mêle à la fois les données des capteurs avec un corpus documentaire hétérogène lié à l'environnement. Notre contribution dans ce cadre consiste à intégrer au sein d'un même système les outils du Web Sémantique, les approches en Recherche d'Information ainsi que les jeux de données des Systèmes Cyber-Physiques.

Pour conclure, j'esquisse deux pistes d'ouverture de mes activités de recherche futures en indiquant que le Web Sémantique a encore toute sa place dans les nouveaux enjeux scientifiques actuels, et plus particulièrement dans la gestion de données massives ainsi que dans le contexte de l'informatique éco-responsable.

Sommaire

1	Introduction	1
1.1	Un peu d'histoire en quelques lignes	1
1.2	Un fil conducteur dans mes activités	2
1.3	Organisation du manuscrit	5
2	Les fondamentaux du Web Sémantique	8
2.1	Préambule	8
2.2	Décrire les données	10
2.3	Raisonner sur les données	13
2.4	Interroger les données	16
2.5	Synthèse du chapitre	17
3	Normaliser des descriptions RDF	20
3.1	Le défi de l'harmonisation des descriptions RDF	21
3.2	État de l'art	24
3.3	Notre proposition de normalisation	25
3.4	Expérimentations	27
3.5	Synthèse du chapitre	29
4	Exploiter le Web Sémantique	31
4.1	Le défi de l'adaptation de documents multimédia	32
4.2	État de l'art	34
4.2.1	Modélisation des profils	34
4.2.2	Des approches pour adapter des documents multimédia	37
4.3	Notre proposition de modélisation sémantique	38
4.3.1	Un profil sémantique générique pour décrire un contexte	38
4.3.2	Modélisation sémantique du processus d'adaptation	41
4.4	Expérimentations	44
4.5	Synthèse du chapitre	46
5	Fédérer des données hétérogènes à l'aide du Web Sémantique	48
5.1	Le défi de l'explication d'événements au sein d'un environnement hybride	49
5.2	État de l'art	51
5.3	Notre proposition fédérant des données de communautés différentes	53

5.4	Expérimentations	57
5.5	Synthèse du chapitre	60
6	Conclusion générale	62
6.1	Synthèse de mes activités de recherche	62
6.2	Perspectives sur mon activité de recherche future	63
6.2.1	Gérer des masses de données	64
6.2.2	Prendre en compte l'éco-responsabilité	65
7	Mon parcours : académique & scientifique	68
7.1	Investissement pédagogique	70
7.2	Activité scientifique	71
7.2.1	Publications scientifiques	72
7.2.2	Encadrement doctoral et scientifique	73
7.2.3	Diffusion et rayonnement	74
7.3	Responsabilités collectives	76
7.3.1	Responsabilités administratives	76
7.3.2	Responsabilités et mandats locaux	77
7.3.3	Responsabilités et mandats nationaux	78
7.4	Synthèse du chapitre	78
	Mes publications	80
	Bibliographie	92
	Table des figures	104
	Liste des tableaux	106

Chapitre 1

Introduction

Une Habilitation à diriger des Recherches (HDR) est une occasion unique pour un chercheur de positionner ses travaux par rapport à l’histoire d’une communauté scientifique dans laquelle il évolue (Section 1.1). C’est aussi plusieurs années de collaborations face à divers défis qu’il a fallu relever avec une équipe de recherche (Section 1.2). Dans ce contexte, ce manuscrit met donc en lumière la cohérence de mon activité de recherche depuis mon recrutement en 2010 au sein du Laboratoire d’Informatique de l’Université de Pau et des Pays de l’Adour (Section 1.3).

1.1 Un peu d’histoire en quelques lignes

C’est en 1989 que le Web est né grâce à la proposition¹ de Tim Berners-Lee, à cette époque chercheur au CERN (*Centre Européen de Recherche Nucléaire*). Il souhaitait proposer un système où les données au sein d’une organisation soient liées les unes aux autres permettant ainsi un meilleur partage des connaissances. En 1993, cette proposition se concrétise avec le premier navigateur web exploité dans le monde entier (Mosaic²) qui permet d’afficher des pages contenant du texte ainsi que des images. L’année suivante, lors de la première conférence internationale sur le WWW³, la création du consortium W3C (*World Wide Web Consortium - www.w3.org*) est annoncée. Durant cet événement, Tim Berners-Lee présente les directions futures du W3C et en particulier il expose le besoin de sémantique pour le Web⁴. Pendant plusieurs années, le W3C va ensuite mener des réflexions pour concevoir les technologies qui serviront de base à cette extension du Web. Le Web Sémantique sera promu auprès du grand public en mai 2001 dans un article publié dans la revue *Scientific American* [25]. La communauté du Web Sémantique commence à se structurer en organisant la première conférence internationale du Web Sémantique (ISWC) en 2002.

1. <https://www.w3.org/History/1989/proposal.html>

2. [https://en.wikipedia.org/wiki/Mosaic_\(web_browser\)](https://en.wikipedia.org/wiki/Mosaic_(web_browser))

3. <https://www94.web.cern.ch/WWW94/>

4. <https://www.w3.org/Talks/WWW94Tim/>

Grâce aux efforts de normalisation du W3C, des recommandations importantes vont voir le jour en 2004 en particulier pour d’une part décrire et partager des données avec le formalisme RDF [69] ou encore pour d’autre part raisonner sur les données avec le langage OWL [93]. En 2007, le Web de données (*Linked Data*) prend concrètement forme avec les technologies issues du Web Sémantique grâce au projet DBpedia⁵ qui va engendrer automatiquement un grand volume de descriptions sémantiques au sujet de contenus issus des pages éditées dans Wikipedia. Il s’agit là de l’une des réussites les plus importantes du Web Sémantique qui a permis de prouver qu’un passage à l’échelle de l’exploitation de descriptions sémantiques était tout-à-fait possible. Afin de généraliser et d’étendre encore cette exploitation de descriptions sémantiques, Tim Berners-Lee lancera un appel à de nombreuses communautés en 2009 durant une conférence TED⁶ lors de laquelle il clamera la mise à disposition de données par tous et pour tous. Suite à cet appel, les graphes de connaissance vont se développer à grande échelle. Par exemple en mai 2012, Google intègre dans son moteur de recherche son graphe de connaissance afin de mieux répondre aux utilisateurs⁷. Il sera suivi par d’autres compagnies mondialement connues, telles que Facebook, LinkedIn, Airbnb, Microsoft, Amazon, Uber... En 2019, la première conférence internationale autour des graphes de connaissance (*ICKG - International Conference on Knowledge Graph*) permettra de réunir tous les acteurs qui traitent de ce sujet.

C’est donc dans ce contexte historique que mon activité scientifique va s’inscrire. En effet, dans la partie suivante, je montrerai que le Web Sémantique a alimenté l’ensemble de mes activités de recherche menées depuis une dizaine d’années au sein de mon équipe de recherche.

1.2 Un fil conducteur dans mes activités

Lorsque j’ai rejoint en 2010 l’équipe T2i (*Traitements des informations pour l’adaptation de l’interaction au contexte et à l’utilisateur*) du Laboratoire d’Informatique de l’Université de Pau et des Pays de l’Adour (LIUPPA - liuppa.univ-pau.fr), l’objectif de mon recrutement était double puisqu’il s’agissait d’apporter de nouvelles compétences tout en collaborant au maximum avec les collègues de l’équipe, sachant que ces derniers sont impliqués dans différentes communautés scientifiques, comme par exemple l’Informatique Distribuée, la Recherche d’Information et les Environnements Informatiques pour l’Apprentissage Humain. Étant donné qu’aucun des chercheurs de l’équipe n’exploitait à cette époque les technologies issues du Web Sémantique, les intégrer au sein des multiples activités de cette équipe m’a semblé être une évidence. À mon échelle, j’ai donc essayé de participer à cette dynamique impulsée dans la section 1.1 car il y avait un réel besoin de décrire ainsi que d’exploiter la sémantique des informations traitées par les projets en cours de l’équipe via des graphes de connaissance.

5. <https://www.dbpedia.org>

6. https://www.ted.com/talks/tim_berniers_lee_the_next_web

7. <https://blog.google/products/search/introducing-knowledge-graph-things-not/>

Ces dernières années, j'ai donc respecté cet engagement de collaborations multiples ce qui m'a amené à participer à la formation doctorale de 9 étudiants en thèse (3 en codirection et 6 en co-encadrement). Pour chacun d'entre eux, j'ai pu les former au Web Sémantique afin qu'ils soient en capacité d'intégrer les outils existants à leurs travaux. Dans cette introduction, je souhaite leur rendre hommage ainsi qu'à tous leurs encadrants qui m'ont également fait confiance. Dans ce qui suit, je les présente donc chronologiquement en précisant l'intitulé de leur thèse. Ceci me permet de montrer par la suite que toutes ces activités ont alimenté au fur et à mesure le fil conducteur de mes recherches lié au Web Sémantique.

- **2010-2012 : The Nhan Luong** - *Modélisation centrée utilisateur final appliquée à la conception d'applications interactives en géographie : une démarche basée sur les contenus et les usages*. J'ai participé au co-encadrement à 10% de cette thèse avec Thierry Nodenot (LIUPPA), Christophe Marquesuzaà (LIUPPA) et Patrick Etcheverry (LIUPPA). The Nhan Luong est actuellement ingénieur au sein de l'entreprise Amadeus IT Group (Nice, France).
- **2013-2017 : Regina Paola Ticona Herrera** - *Normalisation de descriptions RDF*. J'ai co-encadré à 33% cette thèse avec Richard Chbeir (LIUPPA) et Joe Tekli (LAU School of Engineering, Liban). Actuellement, Regina est Enseignante-Chercheuse à l'Université Catholique de San Pablo (Pérou) et Directrice de l'École Doctorale dans cette Université.
- **2014-2017 : Irvin Dongo** - *Anonymisation de descriptions RDF*. J'ai co-encadré à 10% cette thèse avec Richard Chbeir (LIUPPA), Firas Al-Khalil (University College Cork, Irlande) et Yudith Cardinale (Universidad Simón Bolívar, Venezuela). Actuellement, Irvin est Enseignant-Chercheur à l'Université Catholique de San Pablo (Pérou).
- **2014-2018 : Asma Saighi** - *Qualité de service pour les adaptations multimédia dans un environnement mobile*. J'ai co-encadré à 33% cette thèse avec Nassira Ghoualmi (Université de Badji Mokhtar, Algérie) et Philippe Roose (LIUPPA). Actuellement, Asma est Enseignante-Chercheuse à l'Université Oum El Bouaghi (Algérie).
- **2015-2018 : Nathalie Charbel** - *Analyse, classification et enrichissement sémantique de corpus de données hétérogènes : application au domaine du bâtiment*. Cette thèse a été effectuée en collaboration avec l'entreprise Nobatek. J'ai co-dirigé à 50% ce travail avec Christian Sallaberry (LIUPPA). Actuellement, Nathalie est ingénieure au sein de l'entreprise Neo4J.
- **2019-2022 : Nabila Guennouni** - *ISEE : Un système d'Information Hétérogène pour l'explication des événements dans les environnements connectés*. Cette thèse a fait partie d'un projet E2S-UPPA dont j'ai été l'un des porteurs. J'ai co-dirigé à 50% ce travail avec Christian Sallaberry (LIUPPA). Actuellement, Nabila est développeuse au sein d'un service R&D de l'entreprise SEI-LKS dans le domaine du Big Data et de la Science de données.

- **2021-2024 : Jorge Andres Larracochea** - *Vers une méthodologie et une DSL pour une informatique verte*. Il s'agit d'une thèse en co-tutelle avec l'Université de Saragosse. J'effectue un co-encadrement à 25% avec Philippe Roose (LIUPPA), Sergio Ilarri (Universidad de Zaragoza, Espagne) et Yudith Cardinale (Universidad Simón Bolívar, Venezuela). La soutenance de cette thèse est prévue en juin 2024.
- **2022-2025 : Charlotte Darricades** - *Contributions info-communicationnelle de la Recherche & Développement à la traduction d'une mutation industrielle énergétique : Le cas TotalEnergies*. Il s'agit d'une thèse CIFRE entre le laboratoire GERiiCO de Lille et l'entreprise Total Energies sous convention de partenariat avec l'UPPA. Mon co-encadrement à hauteur de 25% s'effectue avec Patrice de La Broise (GERiiCO, Université de Lille), Eric Kergosien (GERiiCO, Université de Lille) et Christian Sallaberry (LIUPPA). Cette thèse est toujours en cours.
- **2023-2026 : Yunji Zhang** - *Trajectoires du bien vivre et bien vieillir sur son territoire*. Il s'agit d'une thèse financée par la Région Nouvelle Aquitaine en collaboration avec l'IRIT et la technopôle Domolandes. Je co-dirige à 50% cette thèse avec Philippe Roose (LIUPPA) et Franck Ravat (IRIT). Cette thèse est toujours en cours.

La figure 1.1 résume dans le temps les différentes thématiques abordées durant les encadrements doctoraux que j'ai cités précédemment avec en parallèle toutes les responsabilités administratives que j'ai assurées. Tous les travaux de recherche illustrés sont en lien avec le Web Sémantique et ils sont regroupés selon les trois catégories suivantes :

1. **Les contributions focalisées directement sur les technologies du Web Sémantique** (Fig. 1.1 en bleu). Les travaux de Regina au sujet de la normalisation de descriptions RDF [126] ou bien ceux d'Irvin sur l'anonymisation de descriptions RDF [40] rentrent complètement dans ce cadre.
2. **Les contributions qui exploitent les technologies du Web Sémantique pour renforcer un domaine applicatif dédié** (Fig. 1.1 en orange). Les travaux de Nhan sur la modélisation d'applications interactives en géographie [87], d'Asma sur la modélisation de profils pour l'adaptation de contenus multimédia [107], de Nathalie sur la modélisation sémantique de corpus documentaires hétérogènes dans le domaine du bâtiment [35], comme ceux en cours de Charlotte pour décrire les connaissances des scientifiques ainsi que des communicants d'une organisation et de Jorge sur l'exploitation de services vertueux s'inscrivent dans cette catégorie.
3. **Les contributions qui permettent de fédérer différentes communautés scientifiques au travers des technologies du Web Sémantique** (Fig. 1.1 en vert). C'est le cas des travaux de Nabila qui ont permis de rapprocher le domaine des Systèmes Cyber-Physiques avec celui de la Recherche d'Information documentaire [55] et plus récemment ceux en cours de Yunji sur les trajectoires sémantiques

considérant des masses de données hétérogènes ainsi que des approches de visualisations et de prédictions d'informations.

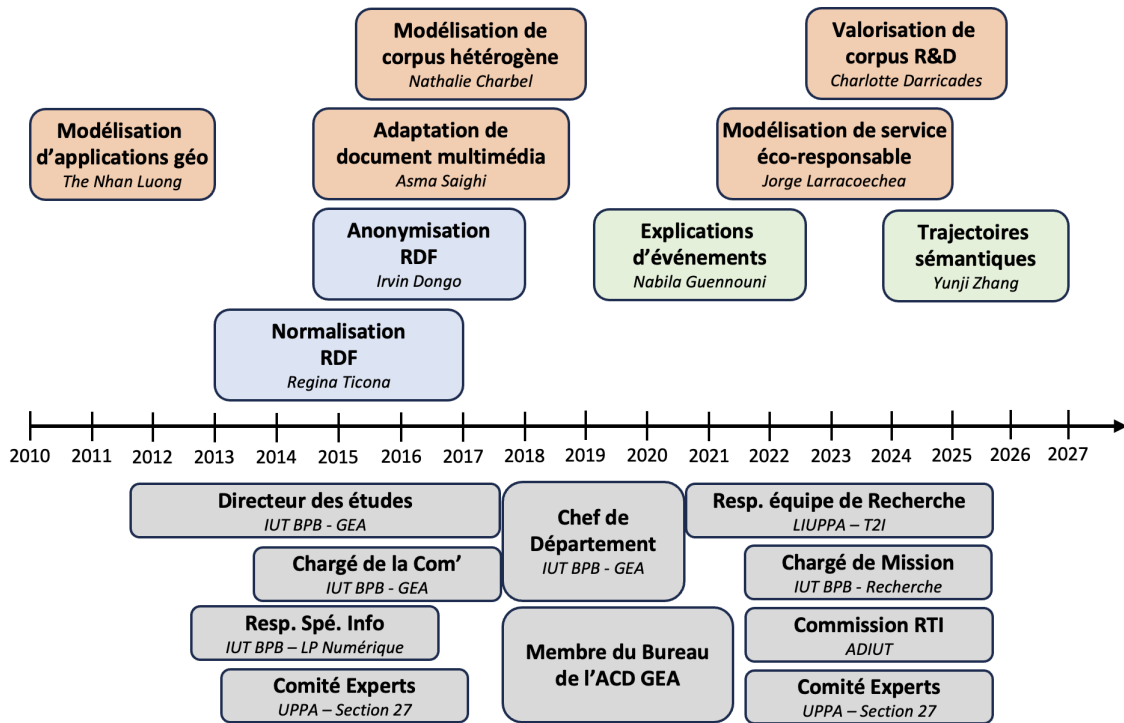


FIGURE 1.1 – Mon implication dans des travaux de thèse depuis 2010

L'organisation de ce manuscrit qui sera décrite dans la partie suivante suivra ces trois grandes catégories d'action. Il ne s'agit pas dans ce manuscrit de détailler toutes les contributions listées précédemment (Figure 1.1) mais plutôt de nous appuyer sur nos expériences acquises dans ces travaux pour mettre en avant des conduites à tenir dans le cadre du Web Sémantique.

1.3 Organisation du manuscrit

Chapitre 2 - Les fondamentaux du Web Sémantique

Ce chapitre est dédié à la présentation succincte de la genèse ainsi que des fondamentaux du Web Sémantique. Les principales technologies pour décrire, raisonner et interroger des données sont présentées au travers d'un exemple pédagogique. L'objectif est de permettre aux lecteurs de comprendre à la fois le vocabulaire ainsi que les technologies de base du Web Sémantique qui seront par la suite évoqués dans les chapitres suivants. Ainsi, le lecteur sera capable d'identifier ce qu'est une ressource, une description ou bien une ontologie. Également, il aura une vue générale des technologies largement exploitées dans ce cadre telles que RDF, RDFS, OWL et SPARQL.

Chapitre 3 - Normaliser des descriptions RDF

Ce chapitre met en avant la prise de recul à avoir sur l'usage des technologies du Web Sémantique et en particulier au sujet de son formalisme de base qui est RDF. En effet, je démontrerai qu'une description RDF peut contenir des redondances d'informations ou bien être encodée de façon très verbeuse. Ce phénomène s'accroît lorsque différentes descriptions sont combinées entre elles. Pour des applications qui intègrent de telles descriptions sémantiques, ceci peut s'avérer néfaste que ce soit en terme de stockage, de chargement ou de traitement des informations. Ainsi, nous proposons un processus de normalisation de descriptions RDF. Son originalité est qu'il nettoie à la fois la structuration de la description mais aussi sa sérialisation. De plus, cette normalisation s'adapte en fonction de l'application cible.

Chapitre 4 - Exploiter le Web Sémantique

Ce chapitre illustre une exploitation des technologies issues du Web Sémantique pour l'adaptation de contenus multimédias. La communauté du Web Sémantique a proposé au travers d'une recommandation du W3C une modélisation possible des profils tout en sachant qu'un profil est un élément essentiel dans ce cadre puisqu'il permet de décrire un contexte d'exécution. La communauté du Multimédia a su prendre en considération ce standard de description pour l'adaptation. Notre prise de recul a permis d'identifier plusieurs limites aux propositions existantes. Ainsi, nous avons développé une nouvelle modélisation sémantique des profils à base de services qui exploite pleinement le formalisme RDF. Nous montrons que cette nouvelle vision du profil peut inspirer à la fois la communauté du Web Sémantique et celle du Multimédia. Nous qualifions dans ce chapitre ce principe de gagnant-gagnant.

Chapitre 5 - Fédérer des données hétérogènes à l'aide du Web Sémantique

Ce chapitre démontre que le Web Sémantique est une solution pour fédérer des données hétérogènes issues de différentes communautés. Pour réaliser cette démonstration, nous présentons le cas d'une recherche d'explications d'événements au sein d'un réseau de capteurs. Cette situation est intéressante puisqu'elle permet de s'appuyer sur un environnement que nous qualifions d'hybride car il mêle à la fois les données des capteurs avec un corpus documentaire hétérogène lié à l'environnement. Notre contribution dans ce cadre consiste à intégrer au sein d'un même système les outils du Web Sémantique, les approches en Recherche d'Information ainsi que les jeux de données des Systèmes Cyber-Physiques.

Chapitre 6 - Conclusion générale

Ce chapitre conclut les chapitres précédents en rappelant que nos contributions en lien avec le Web Sémantique se situent selon 3 axes : (1) les contributions centrées sur des technologies du Web Sémantique, (2) les contributions qui exploitent le Web Sémantique pour renforcer un domaine dédié et (3) les contributions qui rapprochent différentes communautés grâce au Web Sémantique. De plus, cette partie met en avant les grands principes auxquels j'ai souscrit durant toutes ces années : Pédagogie, prise

de recul, principe gagnant-gagnant et fédération de données. Dans cette partie, j'esquisse également deux pistes d'ouverture de mes activités de recherche futures en indiquant que le Web Sémantique a encore toute sa place dans les nouveaux enjeux scientifiques actuels, et plus particulièrement dans la gestion de données massives ainsi que dans le contexte de l'informatique éco-responsable.

Chapitre 7 - Mon parcours académique & scientifique

Ce chapitre offre une synthèse des missions d'enseignant-chercheur que j'ai menées depuis mon recrutement en 2010 au sein de l'Université de Pau et des Pays de l'Adour. J'y aborde mes activités d'enseignement, mes productions et animations scientifiques ainsi que les charges collectives qui m'ont été confiées au niveau de mon établissement et au niveau national.

Chapitre 2

Les fondamentaux du Web Sémantique

Avant de présenter dans les chapitres suivants mes contributions scientifiques en lien avec le domaine du Web Sémantique, il est important dans ce chapitre d'en introduire au préalable les grands principes fondamentaux ainsi que ses principales technologies. Ce chapitre a donc pour objectif d'expliquer de façon pédagogique ces fondamentaux afin de notamment permettre aux lecteurs de mieux appréhender le vocabulaire qui sera utilisé dans la suite du manuscrit.

2.1 Préambule

D'après son inventeur Tim Berners-Lee en 2001 [25], le Web Sémantique est défini comme "*une extension du Web actuel dans lequel on donne un sens bien défini à l'information pour permettre aux machines et aux humains de travailler en coopération*". En effet, les informations sur le Web ont principalement été créées au départ pour la lecture humaine (e.g., des pages internet contenant du texte, des images, des vidéos, etc.). Ces données étaient donc intrinsèquement difficilement exploitables par les machines sans l'appui d'un formalisme de description. Le Web Sémantique vise donc à permettre aux machines de comprendre et d'exploiter automatiquement des ressources disponibles (des documents, des contenus, des données). En conséquence, un défi du Web Sémantique consiste à fournir un ensemble de technologies qui permet de spécifier à la fois toute l'expressivité des données, en leur donnant du sens, mais également de définir les règles qui permettront de raisonner sur ces données et de les interroger. Cette "*sémantisation*" du Web a pour objectif de créer des réseaux d'informations structurées assurant des échanges de données permettant une meilleure interopérabilité entre applications.

Depuis ces dernières décennies, la communauté du Web Sémantique a défini des technologies, aujourd'hui des standards du W3C (*World Wide Web Consortium* - *www.w3.org*), qui se complètent les unes avec les autres comme le montre par exemple la figure 2.1¹. Chaque brique exploite les capacités des briques qui se situent en des-

1. Source W3C : https://commons.wikimedia.org/wiki/File:Semantic_Web_Stack.png

sous d'elle. Les URI (*Uniform Resource Identifier*) sont à la base de toutes les briques puisqu'elles vont identifier de manière unique les ressources à décrire. Le formalisme RDF (*Resource Description Framework*) qui permettra de décrire les ressources exploitera les URI. Les briques supérieures permettront ensuite d'exploiter, d'interroger et de raisonner sur les descriptions RDF. Nous pouvons donc regrouper ces briques selon les trois fonctionnalités suivantes :

- **Décrire les données (Section 2.2) :** Le formalisme RDF avec l'appui des URI sont les fondations de base de la description de données. Ce formalisme permet de produire un graphe de connaissance où chaque ressource à décrire est un nœud du graphe relié par des relations vers d'autres nœuds.
- **Raisonner sur les données (Section 2.3) :** Les ontologies permettent d'ajouter de la sémantique aux données et d'inférer de la connaissance. En effet, chaque élément du graphe de connaissance peut être associé à une classe héritant de propriétés. De plus, des règles d'inférence peuvent permettre de valider la connaissance décrite, voir même de déduire de nouvelles connaissances.
- **Interroger les données (Section 2.4) :** Le langage de requêtes SPARQL permet de rechercher ainsi que de mettre à jour des informations au sein de graphes de connaissance. Ce langage d'interrogation exploite bien entendu la sémantique des descriptions qu'il exploite.

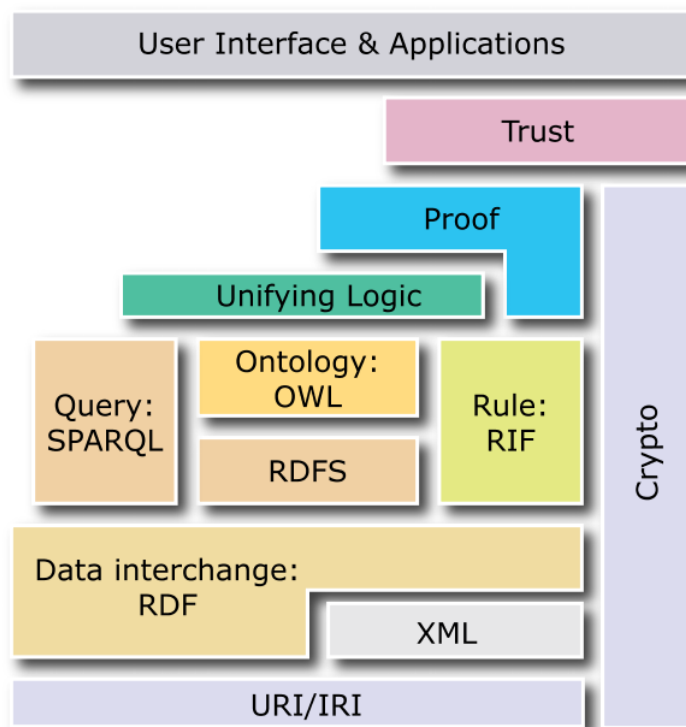


FIGURE 2.1 – La pyramide du Web Sémantique

Dans ce cadre et pour mieux comprendre les fonctionnalités du Web Sémantique, nous présenterons dans les sections suivantes au travers d'un exemple pédagogique comment décrire des données (Section 2.2), comment raisonner sur les données (Section 2.3), et enfin comment interroger des données (Section 2.4). La section 2.5 conclura ce chapitre avec une synthèse qui permettra de faire le lien avec mes activités de recherche.

2.2 Décrire les données

RDF (*Resource Description Framework*) [37, 69] est comme son nom l'indique un cadre de description de ressources. Une ressource peut être quelconque comme un document, une entité physique (e.g., un objet, une personne, une ville) ou bien un concept abstrait (e.g., la liberté, l'amour). RDF permet de faire des déclarations sur des ressources selon la structure suivante : < sujet > < prédicat > < objet >. Dans une telle déclaration, le sujet et l'objet représentent deux ressources avec un prédicat qui précise la nature de leur relation. La relation est formulée de manière directionnelle depuis le sujet vers l'objet. Ainsi, il est possible de représenter visuellement cette déclaration de la façon suivante (Figure 2.2) :

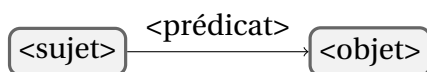


FIGURE 2.2 – Un triplet RDF

Une déclaration RDF est toujours composée d'un triplet (le sujet, le prédicat et l'objet). Bien entendu, une description RDF pourra contenir plusieurs triplets. Une même ressource pouvant être référencée dans plusieurs triplets cela formera un graphe de description. Par exemple, la figure 2.3 illustre un exemple de graphe de description. Dans ce graphe, il y a 7 triplets au total et on peut constater effectivement qu'une ressource comme par exemple <UPPA> est référencée dans 3 triplets que ce soit en tant que sujet (e.g., <UPPA> <locatedIn> <Pau>) ou en tant qu'objet (e.g., <Sebastien Laborie> <employer> <UPPA>).

Plusieurs formats de sérialisation existent pour encoder des descriptions RDF : RDF/XML [52], RDFa [4], N-Triples [34], Turtle [100], JSON-LD [121]... Lors de l'émergence du formalisme RDF au début des années 2000, de nombreux systèmes étaient déjà capables d'exploiter des données décrites dans le langage XML [29]. RDF/XML permet donc de spécifier une description RDF grâce aux balises XML. RDFa permet l'ajout d'attributs au sein de pages Web décrites en HTML. Ainsi, il est possible de combiner descriptions de données avec l'affichage de ces dernières. N-Triples et Turtle sont deux langages plutôt orientés pour les développeurs qui désirent lire facilement les triplets dans une description. Enfin, JSON-LD permet d'encoder une description RDF dans le format JSON qui est aujourd'hui très populaire pour les applications orientées Web.

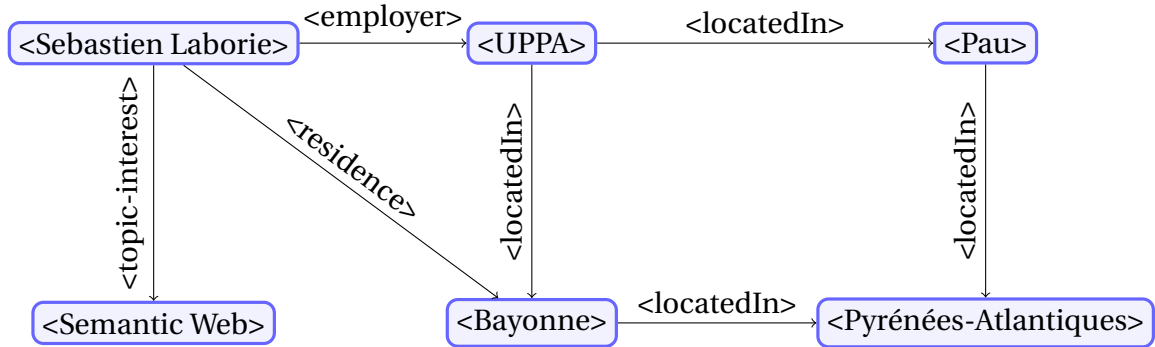


FIGURE 2.3 – Un exemple de description RDF

Afin de sérialiser une description RDF, il existe donc une diversité de formats exploitables selon un contexte d'application cible. Étant donné que le formalisme RDF s'abstrait de ces langages, il est tout à fait possible de traduire une description RDF sérialisée dans un format vers un autre format. En effet, la description de la figure 2.3 respecte le formalisme RDF à base de triplets et elle peut donc être encodée par exemple en RDF/XML, N-Triple, JSON-LD...

Au-delà des sérialisations possibles, un autre atout d'une description RDF est qu'elle doit pouvoir s'agréger avec d'autres descriptions RDF décrites par ailleurs. Ceci permet notamment d'accroître au sein d'une même application la quantité d'informations au sujet d'une même ressource. Pour ce faire, le formalisme s'appuie sur le concept d'URI (*Uniform Resource Identifier*) [113] qui identifie de manière pérenne et unique une ressource. Par exemple dans la figure 2.3, la ressource `<Semantic Web>` peut être associée à l'URI suivante : `<https://www.wikidata.org/wiki/Q54837>`. Tandis que la ressource `<Pyrénées-Atlantiques>` peut faire référence à une autre URI comme : `<http://dbpedia.org/resource/Pyrénées-Atlantiques>`. Les relations peuvent également exploiter des URIs. Par exemple dans la figure 2.3, la relation `<topic-interest>` pourrait faire référence à l'URI : `<http://xmlns.com/foaf/0.1/topic_interest>`.

D'autres descriptions RDF avec de nouvelles informations pourraient utiliser les mêmes URIs que celles citées précédemment pour la figure 2.3. Par exemple, Wikidata fournit une autre description de l'URI `<https://www.wikidata.org/wiki/Q54837>` qui indique entre autres que l'inventeur du Web Sémantique est Tim Berners-Lee employé au W3C. Cette URI est consultable sur Wikidata.org et elle permet de disposer de nombreuses informations complémentaires qui ne sont pas présentes dans la description de la figure 2.3.

Ainsi, lorsque différentes descriptions RDF partageant les mêmes URIs sont intégrées au sein d'une application, toutes les déclarations de triplets sont agrégées entre elles ce qui forme un graphe de données liées. Par exemple, en associant notre description RDF de la figure 2.3 avec Wikidata, étant donné que ces deux sources partagent la même URI associée à la ressource `<Semantic Web>` (i.e., `<https://www.wikidata.org/wiki/Q54837>`), les descriptions de cette ressource vont s'agréger entre elles et par conséquent on obtiendra un graphe contenant à la fois la description de la figure

2.3 (cf., partie bleue de la figure 2.4) et celle de Wikidata (cf., partie rouge de la figure 2.4). Par extension, si l'on intègre d'autres sources d'information comme DBpedia par exemple, puisque la ressource <Pyrénées-Atlantiques> avec l'URI <http://dbpedia.org/resource/Pyrénées-Atlantiques> y est décrite, nous pourrions compléter notre réseau de données avec encore de nouvelles informations (cf., partie verte de la figure 2.4).

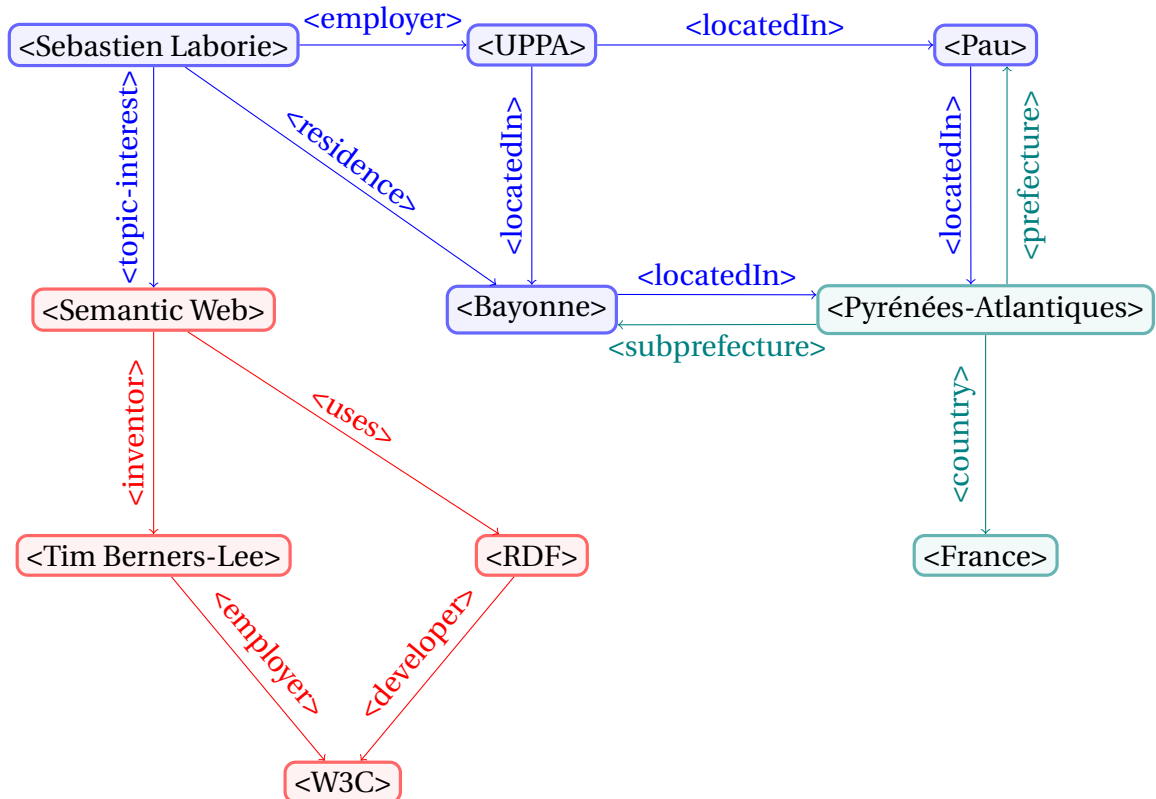


FIGURE 2.4 – Un exemple d'un réseau de données liées

Comme illustré dans la figure 2.4, RDF permet donc de combiner des triplets de n'importe quelle source ce qui permet de représenter et de traiter une grande quantité d'informations. Des jeux de données sont disponibles en ce sens, comme ceux cités précédemment tels que DBpedia² qui publie des données extraites de Wikipedia ainsi que Wikidata³ qui est une source de données collaborative et multilingue. Au travers de notre description (Figure 2.4), les deux URIs présentées précédemment pour <Semantic Web> ainsi que <Pyrénées-Atlantiques> permettent de naviguer au sein de ces jeux de données et elles donnent une vision très concrète de la diversité d'informations disponibles.

Dans la partie suivante, nous allons montrer comment définir un vocabulaire de description et comment raisonner sur celui-ci.

2. <http://fr.dbpedia.org>

3. <http://www.wikidata.org>

2.3 Raisonner sur les données

Le terme "*ontologie*" [127, 54] est incontournable dans la communauté du Web Sémantique puisqu'il permet d'attribuer de la sémantique aux ressources ainsi que de raisonner sur celles-ci. Une définition générale est proposée dans la communauté francophone, comme par exemple dans [36] : "*Une ontologie fournit le vocabulaire spécifique à un domaine de la connaissance et, selon un degré de formalisation variable, fixe le sens des concepts et des relations qui les unissent.*". Afin de spécifier des vocabulaires pour guider la description de ressources, deux langages standards du W3C ont été définis : RDFS (*RDF Schema*) [30] et OWL (*Web Ontology Language*) [93, 131]. Comme il a été mentionné dans la définition précédente et comme illustré dans la pyramide du Web Sémantique (Figure 2.1), ces deux langages offrent différents niveaux de formalisation allant des ontologies dites "légères" en exploitant RDFS à une modélisation sémantique plus complète avec OWL [59]. Dans ce qui suit, nous allons présenter le principe général de fonctionnement des ontologies en prolongeant notre exemple de la section 2.2 et en montrant des différences d'expressivité entre ces deux langages de spécification d'ontologies.

Le langage RDFS dispose de primitives qui permettent d'associer des caractéristiques sémantiques aux ressources d'une description. C'est la raison pour laquelle on parle d'ontologie "légère", c'est-à-dire que ce langage offre le minimum nécessaire pour commencer à définir un vocabulaire de description. La table 2.1 [116] présente des éléments de base du langage RDFS en complément de certains issus du langage RDF.

Primitives RDF & RDFS	Forme syntaxique	Description
Class	C <code>rdf:type</code> <code>rdfs:Class</code>	C est une classe
Property	P <code>rdf:type</code> <code>rdf:Property</code>	P est une propriété
type	I <code>rdf:type</code> C	I est une instance de C
subClassOf	C1 <code>rdfs:subClassOf</code> C2	C1 est une sous-classe de C2
subPropertyOf	P1 <code>rdfs:subPropertyOf</code> P2	P1 est une sous-propriété de P2
domain	P <code>rdfs:domain</code> C	Le domaine de P est C
range	P <code>rdfs:range</code> C	La portée de P est C

TABLE 2.1 – Un exemple de primitives des langages RDF & RDFS

Dans cette table, on constate que la notion de classe (**C**) est exploitée pour spécifier les catégories qui peuvent être utilisées pour classer des ressources. Une propriété (**P**) permettra de définir une relation entre un sujet et un objet de triplet. Les membres

d'une classe sont appelés des instances (**I**) de la classe. Avec RDFS, il est possible de créer des hiérarchies de classes et de sous-classes (**C1, C2...**), mais également des hiérarchies de propriétés et de sous-propriétés (**P1, P2...**). Enfin, des informations supplémentaires sur les sujets et les objets de triplets peuvent être spécifiées via les primitives RDFS *domain* et *range* respectivement. La figure 2.5 présente un exemple de vocabulaire décrit conformément avec les éléments listés dans la table 2.1. Ce schéma RDF définit plusieurs classes : la classe <Person>, la classe <Organization> et la classe <University> qui est une sous-classe de <Organization>. Cet exemple définit également la propriété <affiliation> ainsi que sa sous-propriété <employer>. Enfin, toutes instances reliées par la propriété <affiliation> aura son type associé à la classe <Person> (si cette instance est un sujet de la propriété) ou aura son type associé à la classe <Organization> (si cette instance est un objet de la propriété).

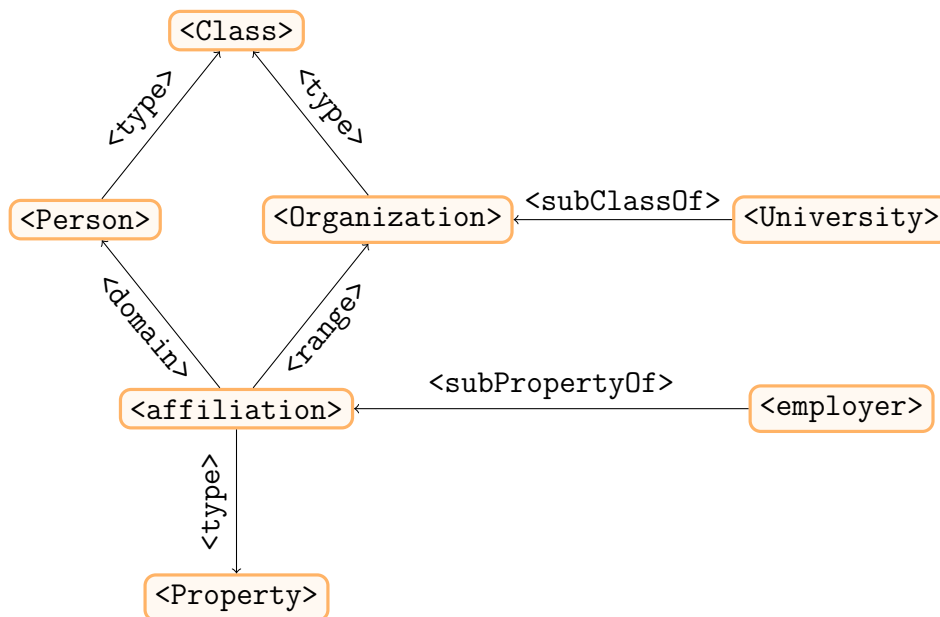


FIGURE 2.5 – Un exemple de schéma RDF

En exploitant ce vocabulaire sur la description de la figure 2.4, il est possible de déduire que les ressources <Sebastien Laborie> et <Tim Berners-Lee> sont des instances de la classe <Person>. Également, on pourra déduire que <UPPA> et <W3C> sont des instances de la classe <Organization>. En effet, la propriété <employer> est une sous-propriété de <affiliation> et son domaine (<Person>) ainsi que sa portée (<Organization>) permettent de caractériser les instances mises en jeu.

RDFS permet donc de donner des informations sur le vocabulaire exploité par les triplets des descriptions RDF. L'organisation hiérarchique des classes ainsi que des propriétés avec les éléments de domaine et de portée permet la réalisation de raisonnements basiques comme nous l'avons vu précédemment. Néanmoins, RDFS ne permet pas l'expression de relations plus complexes entre classes (e.g., l'union, la disjonction), ni l'expression de contraintes plus précises sur des classes et des propriétés (e.g., tous

les membres d'une classe auront une propriété particulière, ou seulement quelqu'un d'entre eux).

Pour une modélisation plus riche des classes et des propriétés associées à un vocabulaire, il est nécessaire de spécifier l'ontologie à l'aide du langage OWL. Afin d'avoir un aperçu du pouvoir plus expressif de ce langage, voici quelques exemples ci-après :

- **Création de classes par combinaison ensembliste d'autres classes.** Par exemple, supposons une classe <Woman> et une classe <Man>, la classe <Person> peut être définie en OWL comme étant l'union disjointe (`owl:disjointUnionOf`) de ces deux classes. Également, la classe <Human> peut être définie comme étant équivalente (`owl:equivalentClass`) à la classe <Person>. Ainsi, une instance de la classe <Person> appartiendra soit à <Woman>, soit à <Man>. De plus, cette instance appartiendra à la classe <Human>.
- **Caractériser des propriétés.** Par exemple, la propriété <locatedIn> (Figure 2.4) est transitive en OWL (`owl:TransitiveProperty`) ce qui permet de déduire que si <UPPA> est localisée à <Bayonne> et que <Bayonne> est localisée dans <Pyénées-Atlantiques> alors <UPPA> est localisée dans <Pyénées-Atlantiques>. D'autres caractérisations OWL existent pour définir des propriétés inverses (e.g., <contains> est la propriété inverse de <locatedIn>), des propriétés symétriques ou asymétriques (e.g., la propriété <locatedIn> est asymétrique car $X \text{ <locatedIn> } Y \neq Y \text{ <locatedIn> } X$), des propriétés réflexives ou irreflexives (e.g., la propriété <employer> est irreflexive car on ne peut s'employer soit même) ou des propriétés fonctionnelles (e.g., la propriété <birthDate> est fonctionnelle car elle cible au moins et au plus une valeur possible).
- **Définir des cardinalités.** Par exemple, on peut spécifier en OWL avec les opérateurs `owl:minCardinality` et `owl:maxCardinality` qu'une personne a toujours 2 parents, et qu'un parent est une personne qui a au moins un enfant.

Comme nous venons de le voir dans les exemples précédents, OWL est un langage qui offre donc des primitives supplémentaires pour spécifier des ontologies. Il décrit de façon plus précise les classes ainsi que les propriétés, et par conséquent il permet d'augmenter les capacités de raisonnement sur les informations décrites.

Dans ce manuscrit, nous n'irons pas plus loin dans la présentation des ontologies, les idées générales sur leur fonctionnement ayant été suffisamment présentées pour la compréhension du lecteur. Pour les lecteurs intéressés, des exemples de vocabulaires ont été définis dans différentes communautés : FOAF⁴ pour les réseaux sociaux, DC Ontology⁵ pour les métadonnées documentaires, Music Ontology⁶ pour l'industrie musicale, SUMO⁷ qui intègre de multiples ontologies de domaine (e.g., Nourriture, Sport, Économie)... Dans la partie suivante, nous allons maintenant montrer comment interroger des descriptions.

4. <http://xmlns.com/foaf/0.1/>

5. <http://dublincore.org/specifications/>

6. <http://musicontology.com>

7. <https://github.com/ontologyportal/sumo>

2.4 Interroger les données

SPARQL [123] est le langage de requêtes standard du W3C pour interroger des descriptions RDF. Une requête contient généralement ce que l'on appelle un patron de graphe (*graph pattern*). Un patron de graphe peut être vu comme un ensemble de triplets où le sujet, le prédicat ou l'objet peut être remplacé par une variable. On peut voir cela finalement comme un graphe contenant des inconnus. L'objectif de SPARQL consiste donc à trouver ces inconnus (les variables) en faisant correspondre le patron de graphe avec des parties d'une description RDF. Des patrons de graphe sont illustrés dans la figure 2.6 avec une ou plusieurs variables que l'on identifie avec le symbole "?".

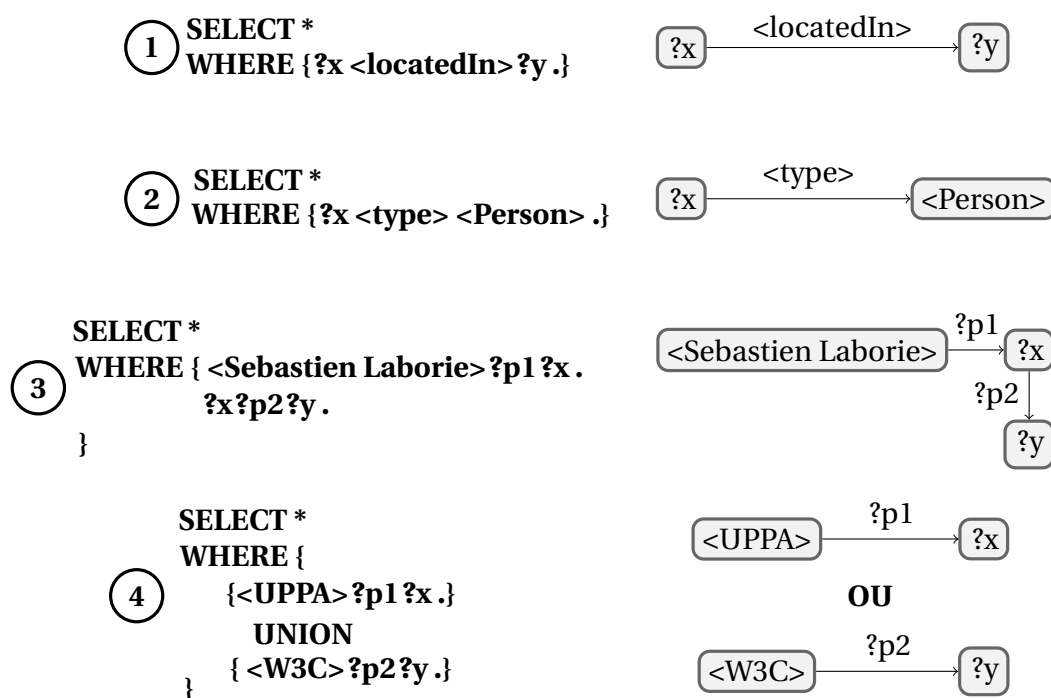


FIGURE 2.6 – Des exemples de requêtes SPARQL et leurs patrons de graphe

La figure 2.6 présente quatre exemples de requêtes SPARQL (sur la gauche de la figure) avec leurs patrons associés (sur la droite de la figure) que nous appliquerons sur la description RDF de la figure 2.4 :

- La requête ① recherche tous les triplets avec le prédicat `<locatedIn>`. On cherche donc les sujets ainsi que les objets impliqués avec cette relation `<locatedIn>` entre eux. Par conséquent, le patron contient 2 variables : l'une pour le sujet (`?x`) et l'autre pour l'objet (`?y`). Une fois exécutée sur la description de la figure 2.4, la requête renvoie respectivement chaque sujet ainsi que chaque objet associé au patron, c'est-à-dire `<UPPA>` et `<Pau>`, `<Bayonne>` et `<UPPA>`, `<Bayonne>` et `<Pyrénées-Atlantiques>`, ainsi que `<Pau>` et `<Pyrénées-Atlantiques>`.

- La requête ② recherche tous les sujets de triplets qui sont de type <Person>. Il n’y a donc dans le patron qu’une seule variable à définir pour le sujet (?x) puisque l’on connaît le prédicat (i.e., <type>) et l’objet (i.e., <Person>). Grâce au vocabulaire de la figure 2.5, cette requête renvoie <Sebastien Laborie> et <Tim Berners-Lee>.
- La requête ③ recherche non seulement les objets dont le sujet est <Sebastien Laborie>, mais également des informations sur ces objets en lien avec la ressource <Sebastien Laborie>. Il y a donc 4 variables à définir dans le patron : le prédicat (?p1) et l’objet (?x) en lien avec la ressource <Sebastien Laborie>, puis le prédicat (?p2) et l’objet (?y) en lien avec la ressource ?x trouvée précédemment. En s’exécutant, la requête fournit toutes les informations au sujet de la ressource <Sebastien Laborie> ainsi que sur les triplets correspondants à ses objets associés.
- La requête ④ recherche tous les triplets dont le sujet est soit <UPPA> ou soit <W3C>. Comme on peut le voir, le langage de requête permet la combinaison de plusieurs patrons de graphe. Par exemple, l’opérateur UNION en SPARQL permet de fusionner les résultats produits par deux patrons différents.

SPARQL est un langage de requêtes assez intuitif qui permet donc de rechercher de l’information au sein de graphes de descriptions sémantiques. Bien entendu, il existe dans ce langage de multiples opérateurs qui permettent par exemple de spécifier différentes contraintes sur la structure du patron, de filtrer et d’ordonner les réponses, d’interroger à la fois différentes sources de données, de produire de nouvelles descriptions RDF, etc. Pour les lecteurs désireux de tester l’interrogation de descriptions RDF, des services de requêtes SPARQL sont disponibles en ligne, notamment pour les jeux de données cités dans la section 2.2. En effet, DBpedia dispose d’un éditeur de requêtes utilisable à cette adresse : <https://dbpedia.org/sparql>. Tandis que Wikidata offre un autre éditeur SPARQL à l’adresse suivante : <https://query.wikidata.org>.

2.5 Synthèse du chapitre

Dans ce chapitre, nous avons présenté les fondamentaux du Web Sémantique pour décrire (Section 2.2), raisonner (Section 2.3) et interroger (Section 2.4) des descriptions sémantiques. Nous avons illustré chaque partie précédente avec un exemple pédagogique qui permet au lecteur d’une part de mieux appréhender le vocabulaire qui sera exploité dans ce manuscrit, et d’autre part de comprendre le fonctionnement général d’une description RDF, d’une ontologie ainsi que d’une requête SPARQL.

Dans ce manuscrit, je vais montrer que le Web sémantique a nourri toutes mes activités de recherche. Au fur et à mesure des années, je me suis fixé des lignes de conduite complémentaires que je vais présenter dans chaque chapitre qui va suivre. La table 2.2 offre une vue globale de mes positionnements. Bien entendu, la pédagogie est à la base de mon engagement avec pour chaque défi scientifique des exemples concrets qui permettent de mieux comprendre les contributions proposées.

Activités de Recherche	Pédagogie	Prise de recul	Principe gagnant-gagnant	Fédérer des données de différentes communautés
Chapitre 2	X			
Chapitre 3	X	X		
Chapitre 4	X	X	X	
Chapitre 5	X	X	X	X
Chapitre 6	X	X	X	X

TABLE 2.2 – Une vue générale de mes contributions dans la cadre du Web Sémantique

Dans le chapitre 3, je me suis essentiellement concentré sur les descriptions RDF afin de les normaliser, c'est-à-dire optimiser la structure du graphe de description ainsi que sa sérialisation. Ce travail a permis de mettre en évidence qu'il est nécessaire d'avoir de bonnes pratiques lorsque l'on souhaite décrire des données que ce soit au niveau de la structuration du graphe de description afin d'éviter notamment des redondances d'informations mais également au niveau de leur sérialisation afin d'éviter une syntaxe trop verbeuse. C'est ce que j'appelle la prise de recul sur les technologies issues du Web Sémantique.

Dans le chapitre 4, nous exploitons les technologies du Web Sémantique comme un outil pour décrire, raisonner et interroger des descriptions. Nous montrons que la communauté du Web Sémantique a développé une recommandation du W3C pour décrire les contextes d'exécution des documents multimédia. Notre prise de recul nous a permis de démontrer que ce standard est limité pour guider la transformation de ce type de document. Dans ce cadre, nous montrons que la sémantique associée aux données peut s'exploiter à de multiples niveaux pour adapter des documents, que ce soit pour les profils des plates-formes cibles, les services d'adaptation ainsi que pour aider à leur combinaison. Ainsi, nos contributions sont une source d'inspiration pour la communauté du Multimédia et celle du Web Sémantique. C'est ce que j'appelle dans ce chapitre le principe gagnant-gagnant.

Dans le chapitre 5, nous montrons que le Web Sémantique permet de fédérer des données provenant de différentes communautés. Nous prenons l'exemple de la recherche d'information au sein d'un réseau de capteurs qui est un domaine de recherche récent. Dans ce cadre, on exploite un corpus de documents hétérogènes avec des données de capteurs issues de la communauté des Systèmes Cyber-Physiques. La mise en relation de ces données avec l'exploitation d'une approche en Recherche d'Information permet en particulier d'expliquer les raisons de déclenchement d'un événement ce qui facilite les prises de décision pour un utilisateur. Bien entendu, nous nous appuyons sur notre prise de recul et le principe gagnant-gagnant.

Dans le chapitre 6, nous concluons ce manuscrit en montrant que nos perspectives de recherche s'inscrivent toujours dans les lignes de conduite que nous avons suivies durant toutes ces années (Table 2.2) : pédagogie (Chapitre 2), prise de recul (Chapitre 3), principe gagnant-gagnant (Chapitre 4) et fédérer des données de différentes communautés (Chapitre 5).

Chapitre 3

Normaliser des descriptions RDF

Grâce aux technologies issues de la communauté du Web Sémantique, il devient aujourd'hui possible de produire, de gérer et d'interroger des descriptions RDF comme nous l'avons vu dans le chapitre précédent (Chapitre 2). Actuellement, des milliards de triplets RDF sont disponibles [66], par exemple au travers de collections ouvertes de données liées¹, telles que DBpedia, Wikidata, LinkedGeoData². . . En complément de ces sources disponibles, les organismes publics comme ceux du secteur privé partagent et s'échangent leurs descriptions selon les différents formats de sérialisation existants (Section 2.2), comme par exemple RDF/XML, Turtle, N-Triples ou JSON-LD.

Cependant, ces descriptions RDF peuvent ne pas être spécifiées de façon optimale que ce soit en terme de structuration du graphe de description ou que ce soit en terme de production syntaxique dans un format de sérialisation cible. En effet, pour une même description le graphe peut contenir par exemple des informations redondantes ou bien être encodé de façon très verbeuse au sein d'une sérialisation. Cela peut provenir des éditeurs eux-mêmes ou bien lorsque des descriptions RDF provenant de différentes sources plus ou moins rigoureuses sont agrégées entre elles afin de compléter de l'information existante. Par conséquent, ceci peut avoir un impact négatif sur de multiples applications qui exploitent des descriptions RDF, comme celles dédiées au stockage, au chargement, aux traitements, aux mesures de similarité ou d'alignement entre descriptions, etc.

Dans ce chapitre, nous allons démontrer que nous avons su apporter un regard critique au sujet des technologies existantes proposées par la communauté du Web Sémantique (Section 3.1). Au-delà des propositions de l'état de l'art (Section 3.2), nous avons proposé une solution originale qui permet de produire des descriptions normalisées en fonction d'un cadre d'application cible (Section 3.3) et des expérimentations ont montré son efficacité (Section 3.4). Nous concluons enfin ce chapitre par une synthèse (Section 3.5).

1. <https://lod-cloud.net>

2. <http://linkedgeo.org>

3.1 Le défi de l'harmonisation des descriptions RDF

Supposons que deux personnes produisent de façon indépendante une description de la ressource <Sebastien Laborie>. La figure 3.1 illustre chacune des descriptions produites avec leur agrégation.

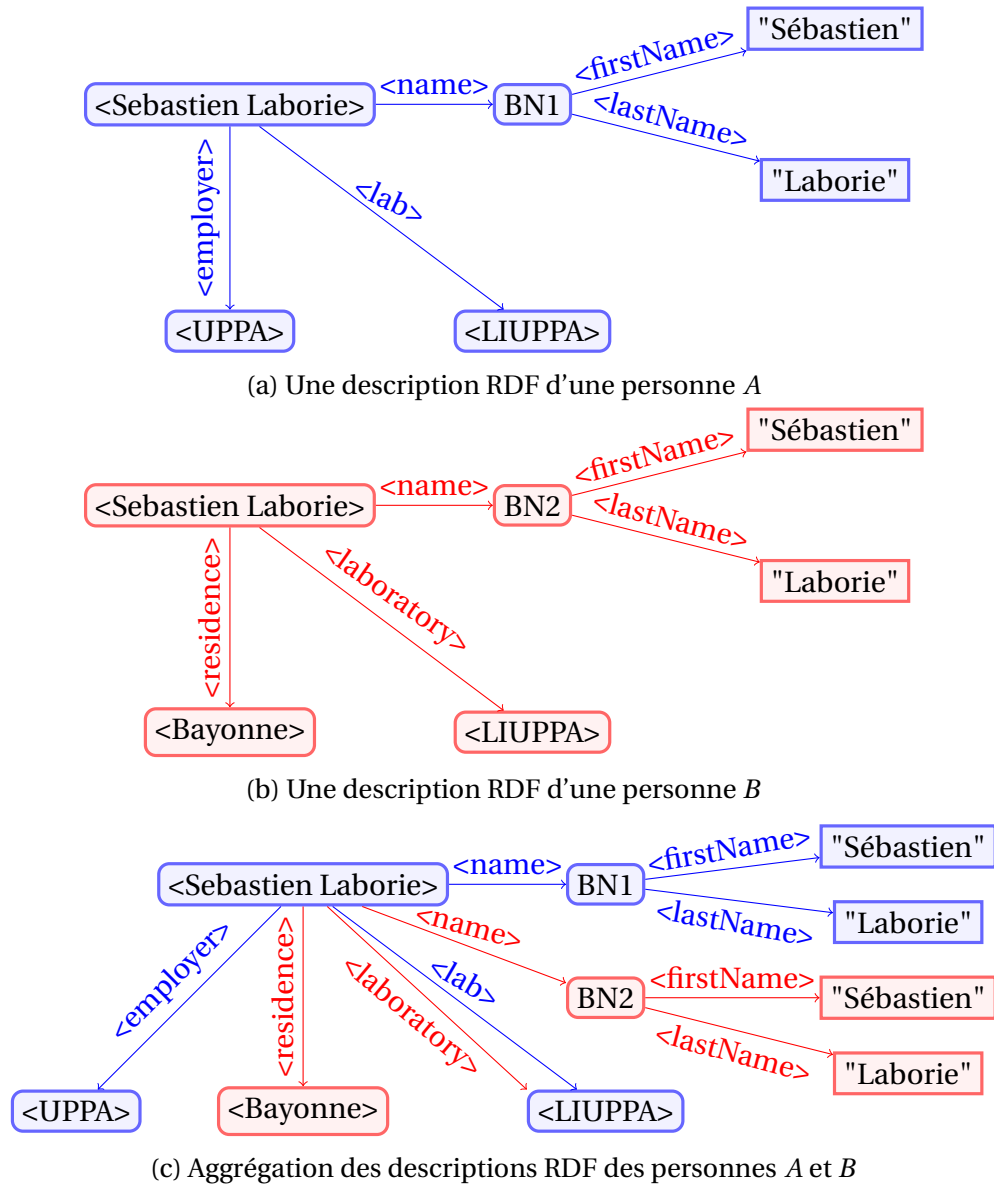


FIGURE 3.1 – Des descriptions RDF au sujet d'une même ressource

D'un côté, la description RDF illustrée dans la figure 3.1a a été réalisée par une personne A. Elle contient 5 triplets qui indiquent que <Sebastien Laborie> est employé par l'UPPA, que son laboratoire est le LIUPPA et qu'il a un nom composé d'un prénom ainsi que d'un nom de famille (avec l'utilisation d'un nœud vide nommé BN1). D'un autre

côté, la description RDF illustrée dans la figure 3.1b a été réalisée par une personne *B*. Elle contient également 5 triplets qui indiquent que <Sebastien Laborie> réside à Bayonne, que son laboratoire est le LIUPPA et qu'il a un nom composé d'un prénom ainsi que d'un nom de famille (avec l'utilisation d'un nœud vide nommé BN2).

Lorsque l'on agrège les descriptions des figures 3.1a et 3.1b pour bénéficier des informations de l'une et de l'autre au sein d'un même graphe cela produit la description illustrée dans la figure 3.1c. Celle-ci contient 10 triplets qui fusionnent toutes les données des deux descriptions initiales produites indépendamment. On remarque aisément que cette fusion d'information peut engendrer de la redondance de données. Par exemple, il y a plusieurs fois le nom ainsi que le prénom de la personne ou encore plusieurs fois le même prédicat qui est mis en jeu (e.g., <name>, <firstName>, <lastName>).

Ce type de problème s'observe également au niveau de la sérialisation des descriptions RDF comme on peut le voir par exemple dans la figure 3.2 avec une sérialisation en RDF/XML. La figure 3.2a représente une sérialisation possible de la description illustrée dans la figure 3.1a, tandis que la figure 3.2b est une sérialisation possible de la figure 3.1b. En agrégeant les balises RDF/XML des figures 3.2a et 3.2b on obtiendra la description de la figure 3.1c. Au-delà du problème de redondance d'informations soulevé précédemment, on peut constater que le langage RDF/XML offre une flexibilité d'écriture des triplets³. En effet, le triplet <Sebastien Laborie> <lab> <LIUPPA> peut s'écrire de différentes façons : de façon condensée comme à la ligne 12 de la figure 3.2a ou bien de façon plus ouverte en vue de la spécification d'autres triplets comme dans les lignes 14-16 de la figure 3.2b.

Notons également que certaines lignes d'une sérialisation peuvent être complètement superflues, ce qui est le cas par exemple pour la ligne 3 et la ligne 5 de la figure 3.2b. Dans le premier cas, l'espace de nom *dc* n'est pas exploité dans la description. Dans le second cas, l'espace de nom *ex2* n'est non seulement pas exploité non plus mais il est en redondance avec l'espace de nom *ex* qui pointe vers la même référence.

Les situations concrètes présentées précédemment nous permettent donc d'apporter un regard critique sur l'usage des technologies de base issues du Web Sémantique, comme RDF. Que ce soit au sujet des structures de description ou bien que ce soit au niveau de leur sérialisation, un usage maladroit pourrait nuire à l'efficacité d'une application, comme par exemple pour le stockage, le chargement, le traitement ainsi que l'interrogation des descriptions. Lors de l'agrégation de descriptions avec de multiples sources, qui est un point très important du Web Sémantique (cf., l'usage des URI décrit dans le Chapitre 2 Section 2.2), ces problèmes de structuration et de syntaxe ne font que s'accumuler et s'accroître sensiblement.

Il existe donc un besoin réel de normalisation de descriptions qui doit s'opérer à la fois sur la structure du graphe de description afin d'éviter la redondance d'informations, mais également sur la sérialisation produite afin d'éliminer par exemple les instructions trop verbeuses ou inutiles.

3. Les autres langages de sérialisation permettent également dans leur syntaxe des variantes de spécification des triplets RDF.

```

1 <?xml version="1.0"?>
2 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
  ns#"
3   xmlns:ex="http://example.org/">
4   <rdf:Description rdf:about="ex:Sebastien-Laborie">
5     <ex:name>
6       <rdf:Description>
7         <ex:firstName>Sebastien</ex:firstName>
8         <ex:lastName>Laborie</ex:lastName>
9       </rdf:Description>
10    </ex:name>
11    <ex:employer rdf:resource="ex:UPPA" />
12    <ex:lab rdf:resource="ex:LIUPPA" />
13  </rdf:Description>
14 </rdf:RDF>
15

```

(a) Un exemple de sérialisation en RDF/XML de la figure 3.1a

```

1 <?xml version="1.0"?>
2 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
  ns#"
3   xmlns:dc="http://purl.org/dc/elements/1.1/"
4   xmlns:ex="http://example.org/"
5   xmlns:ex2="http://example.org/">
6   <rdf:Description rdf:about="ex:Sebastien-Laborie">
7     <ex:name>
8       <rdf:Description rdf:nodeID="BN2">
9         <ex:firstName>Sebastien</ex:firstName>
10        <ex:lastName>Laborie</ex:lastName>
11      </rdf:Description>
12    </ex:name>
13    <ex:residence rdf:resource="ex:Bayonne" />
14    <ex:laboratory>
15      <rdf:Description rdf:about="ex:LIUPPA" />
16    </ex:laboratory>
17  </rdf:Description>
18 </rdf:RDF>
19

```

(b) Un exemple de sérialisation en RDF/XML de la figure 3.1b

FIGURE 3.2 – Des exemples de sérialisation en RDF/XML

Dans la partie suivante, nous présentons un état de l'art au sujet de la problématique de normalisation de description RDF avant de présenter notre solution à ce problème.

3.2 État de l'art

Plusieurs travaux de recherche ont identifié ce besoin de normalisation de descriptions RDF. Certains se sont concentrés sur la structure du graphe de description [50, 64, 86], tandis que d'autres auteurs ont proposé des solutions vis-à-vis de la sérialisation exclusivement [130].

Dans [64], les auteurs démontrent que la notion de graphe RDF n'a pas été clairement définie, au sens mathématique du terme, dans la spécification du formalisme RDF du W3C [37]. De plus, ils montrent également que des redondances de données peuvent apparaître dans leurs structures. Ainsi, ils proposent de représenter les graphes RDF à l'aide d'hypergraphes, tout en sachant que ces derniers peuvent naturellement se représenter en graphes bipartis qui est une modélisation formelle bien connue en mathématiques. Traduire les descriptions RDF dans une telle structure permettra de réduire certaines redondances et de bénéficier d'outils de traitement et de visualisation sur les graphes en général.

Dans [86], les auteurs proposent un même et unique algorithme, intitulé *Universal RDF Graph Normalization Algorithm 2011*, qui normalise des descriptions RDF sérialisées en JSON-LD. En d'autres termes, bien que la solution proposée élimine certaines redondances d'informations, elle traite de façon uniforme les descriptions par exemple avec toujours la même mise à plat de la structure du graphe ainsi que le même ordre imposé des données. Bien entendu, il est mentionné dans leur proposition qu'il n'existe pas un et un seul algorithme possible pour normaliser des descriptions RDF.

Dans [50], une nouvelle structure abstraite intitulée HDT (*Header-Dictionary-Triples*) est proposée pour représenter les descriptions RDF. Celle-ci est composée d'un entête avec des métadonnées générales, d'un dictionnaire spécifiant des identifiants condensés, et d'une structure pour modéliser les triplets sous la forme d'un encodage compact. Un encodeur ainsi qu'un décodeur permettent d'exploiter la structure HDT comme un modèle pivot, c'est-à-dire de RDF vers HDT et inversement. Bien que cette proposition réduise des redondances de données, la normalisation n'est pas leur objectif central puisque leur contribution se concentre plutôt sur l'optimisation du stockage des descriptions afin d'en faciliter les échanges.

Comme nous l'avons illustré dans la figure 3.2, [130] confirme qu'une même description RDF peut s'exprimer de façons différentes en RDF/XML ce qui complexifie le traitement des descriptions. Les auteurs ont défini une méthode de normalisation de sérialisation en utilisant des outils XML, tels que les grammaires DTD (*Document Type Definition*). Avec le soutien d'un expert, ces DTD sont définies semi-automatiquement pour que les sérialisations produites soient conformes avec ces grammaires. Une fois les DTD mises en place, des requêtes SPARQL permettent de construire automatiquement des descriptions RDF/XML normalisées à partir de descriptions initiales.

Toutes ces approches mettent en évidence ce regard critique à apporter sur les technologies du Web Sémantique. Elles proposent différentes méthodes pour combler leurs lacunes, notamment en ce qui concerne la normalisation de descriptions. Néanmoins, celles-ci ne sont pas complètes puisqu'elles n'abordent pas la probléma-

tique de normalisation comme nous le faisons, c'est-à-dire à la fois selon l'angle de la structuration du graphe en plus de celui de la sérialisation. Enfin, celles-ci n'adaptent pas le processus de normalisation en fonction des différents contextes des applications cibles, comme par exemple disposer de descriptions monolingues ou multilingues, autoriser ou non les ressources synonymes, spécifier des ordres d'affichage différents, etc.

3.3 Notre proposition de normalisation

Dans le travail de thèse de Regina Paola Ticona Herrera [126] auquel j'ai participé, nous avons listé de façon précise plusieurs problèmes nécessitant une normalisation de descriptions RDF dont certains ont été illustrés dans les exemples de la partie précédente (Section 3.1). En voici une synthèse dans la table 3.1 qui regroupe chaque problème identifié selon un niveau de traitement possible. Vis-à-vis de la structuration du graphe, des redondances de nœuds et d'arcs peuvent être présentes. Pour la sérialisation utilisée, les éléments disponibles dans le langage de description peuvent ne pas être exploités de façon optimale, par exemple pour les espaces de noms et les variations d'écriture des triplets. Enfin, la sémantique associée aux éléments d'une description peut nous permettre d'identifier des informations équivalentes pouvant être jugées comme de la redondance d'informations (e.g., synonymie de ressources, équivalences de concepts).

Afin de traiter chaque problème listé dans la table 3.1, nous avons associé un ensemble de règles à appliquer sur une description RDF pour les résoudre [125]. L'application de ces règles doit satisfaire *in fine* les 4 propriétés suivantes :

- **Complétude** : Une description RDF D_i est dite complète par rapport à D_j si D_i préserve toutes les informations définies dans D_j .
- **Minimalité** : Une description RDF D_i est dite minimale si toutes les ressources utilisées dans la description D_i sont uniques et que tous les espaces de noms sont exploités.
- **Conformité** : Un fichier RDF F_i est dit conforme s'il vérifie toutes les règles de sérialisation définies dans les standards du W3C.
- **Consistance** : Une description RDF D_i est dite consistante si D_i est à la fois complète, minimale et conforme.

La figure 3.3 présente l'architecture de notre proposition de normalisation de descriptions RDF. Celle-ci prend en entrée une description RDF D et fournit en sortie une description RDF normalisée \tilde{D} . On retrouve dans cette architecture les différents niveaux de traitement illustrés dans la table 3.1.

Selon nous, il n'y a pas de normalisation parfaite puisque cela dépend de l'application cible qui exploitera la description normalisée. Par exemple, si l'application cible nécessite des descriptions RDF monolingues (cf., paramètre 6 *Language Indicator* dans la figure 3.3), notre proposition ne normalisera pas la description de la même façon que pour une description multilingue. Également, une application cible peut nécessiter de

Niveaux de traitement	Problèmes identifiés
Structuration du graphe	1-Redondance de prédicats : Une même relation peut être présente plusieurs fois (e.g., <firstName> en figure 3.1c).
	2-Redondance de sujets ou d'objets : Un nœud peut faire référence à la même information (e.g., "Sébastien" en figure 3.1c).
	3-Redondance de triplets : Un triplet peut être présent plusieurs fois (e.g., <Sebastien Laborie> <lab> <LIUPPA> et <Sebastien Laborie> <laboratory> <LIUPPA> en figure 3.1c).
Sérialisation	4-Duplication d'espaces de noms : Des préfixes différents pointent vers le même espace de nom (e.g., lignes 4 et 5 en figure 3.2b).
	5-Espaces de noms non-exploités : Un espace de nom est déclaré mais il n'est pas utilisé dans la description (e.g., le préfixe dc en figure 3.2b).
	6-Flexibilité d'écriture du langage : Un triplet peut être spécifié à l'aide de différentes constructions (e.g., ligne 12 en figure 3.2a et lignes 14-16 en figure 3.2b).
	7-Ordre de déclaration des triplets : Les triplets peuvent être spécifiés dans un ordre quelconque (e.g., on peut inverser les lignes 11 et 12 en figure 3.2a).
Sémantique	8-Ambiguïté des nœuds vides : Des nœuds vides différents peuvent contenir des informations équivalentes (e.g., BN1 et BN2 en figure 3.1c).
	9-Synonymie des littéraux : Les valeurs d'objets de triplets peuvent être équivalentes (e.g., "Sébastien"@fr et "Sebastian"@es).
	10-Equivalence de sujets ou d'objets : Certains nœuds peuvent être sémantiquement équivalents (e.g., la ressource <UPPA> avec une URI de Wikidata et une URI de DBpedia).

TABLE 3.1 – Les besoins en faveur d'une normalisation de descriptions RDF

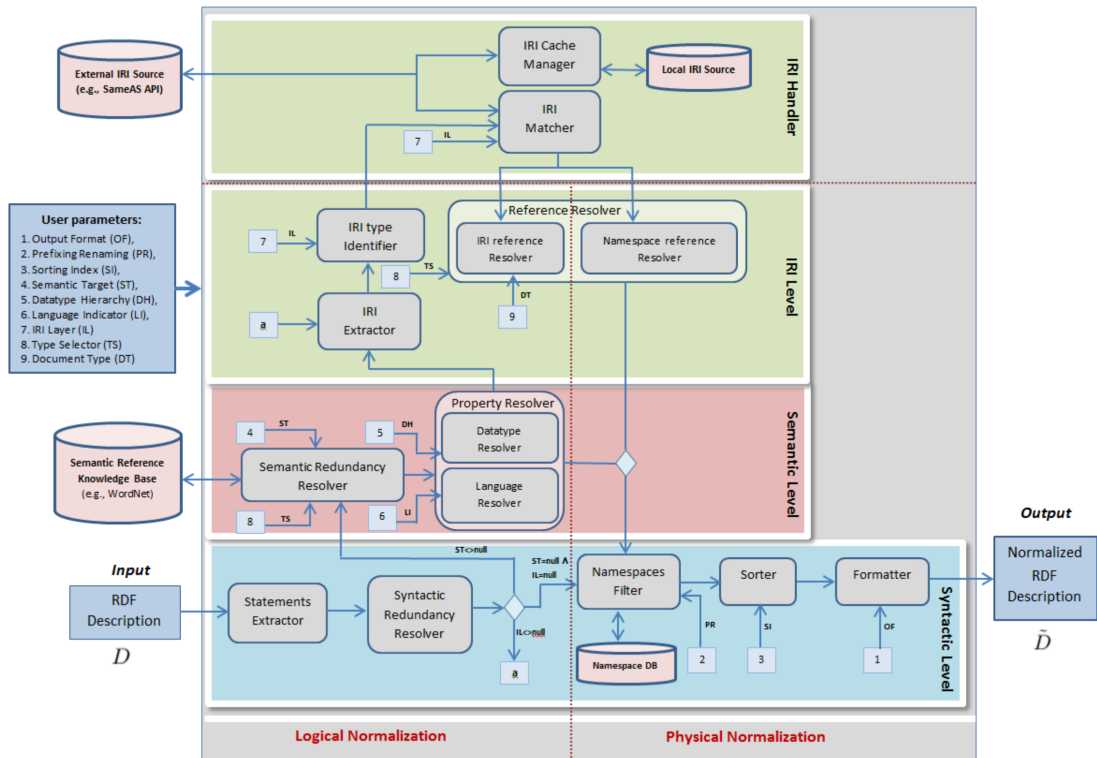


FIGURE 3.3 – Architecture de notre proposition de normalisation de descriptions RDF

conserver une diversité d'informations (e.g., des synonymes, des URI différentes décrivant des ressources équivalentes...) tout en éliminant les répétitions de données, tandis qu'une autre peut désirer d'avoir les strictes informations nécessaires en exécutant un traitement sémantique des informations (cf., paramètre 4 *Semantic Target* dans la figure 3.3), c'est-à-dire sans doublon, ni synonyme, etc. C'est donc à l'utilisateur de paramétrer notre proposition afin que celle-ci lui fournisse des descriptions normalisées selon ses attentes au regard de l'application qui les exploitera.

3.4 Expérimentations

Des expérimentations ont été menées pour valider notre proposition de normalisation de descriptions RDF et en démontrer son efficacité en termes de réduction du volume de données ainsi que de temps de chargement dans une application. Pour ce faire, nous avons au préalable constitué trois groupes de jeux de données :

- **Groupe 1 :** Il s'agit de fichiers RDF/XML que nous avons engendrés nous-même et qui contiennent les problèmes listés dans la table 3.1. Le sous-groupe nommé Syn_DS1 sont des fichiers contenant quelques dizaines de triplets. Tandis que le sous-groupe intitulé Syn_DS2 sont des fichiers avec plusieurs centaines de triplets. Ce groupe permet d'effectuer une preuve de concept de notre approche.

- **Groupe 2** : Il s’agit de descriptions RDF réelles que nous avons récupérées de LinkedGeoData (LGD), DBpedia (DBp) et WordNet (WN). Ce groupe permet de tester notre approche sur des cas réels.
- **Groupe 3** : Il s’agit de fichiers RDF/XML issus du groupe 2 (DBp2 et LGD2) mais pour lesquels nous avons ajouté les problèmes listés dans la table 3.1. Ce groupe permet de tester l’agrégation de données, comme si nous avions combiné les descriptions du groupe 2 avec celles du groupe 1.

Toutes les expérimentations détaillées sont disponibles au sein du manuscrit de thèse de Regina Paola Ticona Herrera [126]. Nous présenterons ci-après uniquement les résultats les plus importants : (1) Réduction du volume des descriptions RDF normalisées et (2) Impact sur leurs chargements dans JENA⁴.

La figure 3.4 illustre les performances de notre processus de normalisation selon les deux points cités précédemment. Comme il est possible de le constater, les taux de réduction des tailles des descriptions sont importants pour le groupe 1 allant de 39.46% en moyenne pour les fichiers du groupe Syn_DS1 jusqu’à 71.11% en moyenne pour les fichiers du groupe Syn_DS2. En effet, plus les fichiers contiennent des duplications d’informations et plus ces derniers sont réduits. Pour le groupe 2, ce taux de réduction est plus faible par rapport au groupe 1. Ceci s’explique par le fait que les fournisseurs de descriptions produisent des données de relativement bonne qualité comme pour DBpedia par exemple qui contient très peu de duplication d’informations. En ce qui concerne le groupe 3, on retrouve des taux de réductions importants ce qui est normal puisque l’on agrège aux descriptions du groupe 2 des problèmes listés dans la table 3.1.

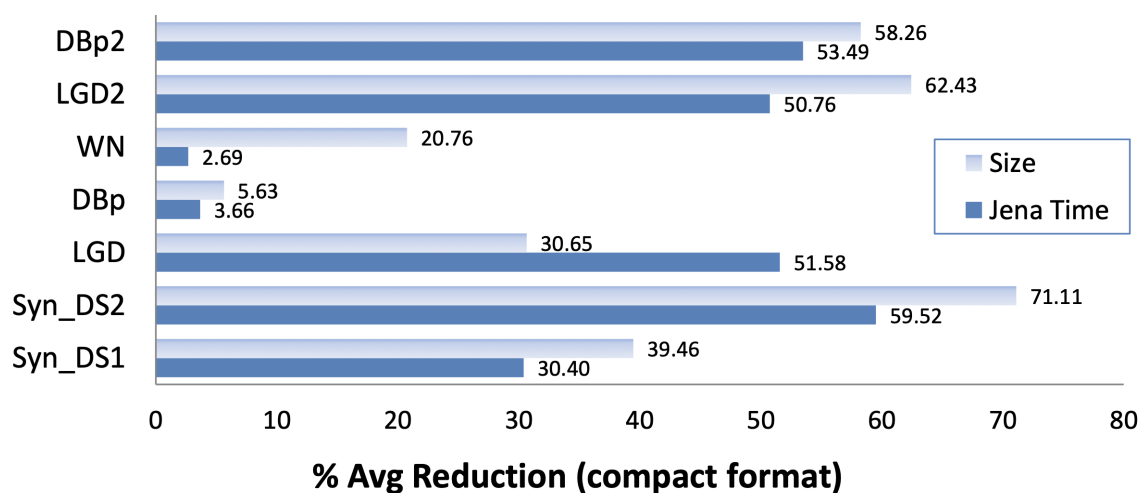


FIGURE 3.4 – Performance de notre processus de normalisation

En ce qui concerne la comparaison du temps de chargement dans JENA des descriptions brutes par rapport aux descriptions RDF normalisées, notre approche offre

4. JENA (<http://jena.apache.org>) permet le développement d’applications supportant des descriptions RDF.

un gain non-négligeable. En effet, pour le groupe 1 ce temps de chargement est réduit de 30.40% en moyenne pour les fichiers légers jusqu'à 59.52% en moyenne pour les fichiers plus lourds. En ce qui concerne le groupe 2, cela varie en fonction des fournisseurs de données. Enfin, pour le groupe 3 on charge les descriptions normalisées deux fois plus rapidement que si nous ne les avons pas normalisées.

Dans cette partie, nous avons démontré qu'une normalisation de descriptions RDF a un impact fort sur le stockage des descriptions ainsi que vis-à-vis de leurs exploitations au sein d'applications.

3.5 Synthèse du chapitre

Dans un premier temps, nous avons montré dans la section 3.1 que nous avons su avoir un regard critique sur l'usage des technologies de base du Web Sémantique, comme les descriptions RDF. En effet, ces descriptions peuvent contenir des redondances d'informations et/ou leurs sérialisations peuvent ne pas être encodées de façon optimale.

Dans un second temps dans la section 3.2, nous avons établi un état de l'art des approches existantes en lien avec la normalisation de descriptions RDF. D'une part, aucune approche dans la littérature ne considère la normalisation à la fois par rapport à la structuration du graphe de description mais aussi vis-à-vis de sa sérialisation. D'autre part, nous proposons un processus de normalisation non-uniforme qui s'adapte selon un contexte d'application cible.

Ensuite, dans la section 3.3, nous avons présenté notre approche de normalisation de descriptions RDF. L'originalité de notre approche est qu'elle ne produit pas une et une seule même forme de normalisation. En effet, selon les besoins d'une application cible, le besoin de normalisation peut être différent. Nous permettons donc à l'utilisateur de paramétrer notre proposition afin de lui fournir une description normalisée adéquate.

Enfin, nous avons présenté des résultats expérimentaux dans la section 3.4 qui démontrent l'impact de notre proposition tant sur la réduction du volume des descriptions que sur leur exploitation au sein d'applications qui doivent les utiliser. Nous avons exploité différents jeux de données allant depuis de petites descriptions (i.e., d'une dizaine à une centaine de triplets RDF) afin d'effectuer des preuves de concept jusqu'à des descriptions réelles pour mesurer le passage à l'échelle de notre proposition via LinkedGeoData ou DBpedia qui fournissent de grandes descriptions (i.e., des milliers de triplets RDF).

Ce travail a été publié en 2015 dans une conférence internationale de rang A au regard du classement CORE :

- Regina Ticona-Herrera, Joe Tekli, Richard Chbeir, **Sébastien Laborie**, Irvin Dongo et Renato Guzman. *Toward RDF Normalization*. In Proceedings of the 34th International Conference on Conceptual Modeling (ER 2015), LNCS 9381, Springer International Publishing, pp. 261–275, 2015.

De plus, il a été exploité dans la thèse d'Irvin Dongo qui traite de l'anonymisation de descriptions RDF [40]. En effet, la normalisation a été utilisée pour nettoyer les descriptions RDF juste avant de déployer un processus d'anonymisation qui a été défini.

Chapitre 4

Exploiter le Web Sémantique

Comme nous avons pu l'introduire dans le chapitre 2, les fondamentaux du Web Sémantique au travers de ses technologies principales (e.g., RDF, OWL, SPARQL) permettent de définir ainsi que d'intégrer des données, de les connecter entre elles et de les échanger entre différents acteurs, pour enfin les enrichir et les réutiliser autant que cela est nécessaire. En une vingtaine d'années, la communauté du Web Sémantique a su diffuser ses acquis auprès de nombreuses communautés scientifiques¹ : Recherche d'Information (e.g., [117]), Multimédia (e.g., [122]), Systèmes Cyber-Physiques (e.g., [106])... Chaque communauté, dont celle du Web Sémantique, a su tirer partie de ces échanges : "*Les avancées dans la poursuite du Web Sémantique nécessitent des contributions de nombreux sous-domaines de l'informatique, l'une des quêtes clés étant d'unifier les contributions*" [67].

Mes activités de recherche ont suivi ce chemin qui consiste d'un côté à exploiter le Web Sémantique en illustrant son apport dans différents cadres originaux, comme l'adaptation de documents multimédia. Puis, en identifier ses limites comme nous avons pu le faire dans le chapitre 3 au sujet par exemple de la normalisation de descriptions RDF. Cette démarche scientifique, je la qualifie de gagnante-gagnante.

Dans ce chapitre, je vais montrer comment j'ai pu mettre en œuvre ce principe d'exploitation gagnant-gagnant. Dans un premier temps, le défi de l'adaptation de documents multimédia qui a été l'une de mes toutes premières problématiques de recherche [71] sera illustré comme exemple (Section 4.1). Dans un second temps, je montrerai que les technologies du Web Sémantique ont été exploitées pour décrire les plates-formes et que différentes approches d'adaptation ont été développées (Section 4.2). Dans une troisième partie, nous présenterons notre modélisation sémantique du contexte et nous verrons que les méthodes d'adaptation elles-mêmes peuvent aussi être modélisées sémantiquement (Section 4.3). La section 4.4 présentera des expérimentations ainsi que des évolutions de nos propositions. La section 4.5 offrira une synthèse ainsi qu'une mise en perspective de ce principe d'exploitation gagnant-gagnant.

1. Je mentionne les communautés scientifiques en lien avec mes activités de recherche menées au sein de l'équipe T2i du LIUPPA. Bien évidemment, d'autres communautés pourraient être citées comme le Traitement Automatique du Langage Naturel, les Bases de Données, l'Intelligence Artificielle...

4.1 Le défi de l'adaptation de documents multimédia

Lorsque j'ai débuté mon doctorat en 2004, la communauté du Multimédia développait plusieurs standards permettant de créer et de consulter des documents composés de différents types de contenus, tels que des vidéos, des objets sonores, des textes et des images. Des exemples typiques de documents multimédias sont les pages web (notamment avec HTML5² qui a introduit entre autres les balises <video> et <audio> [99]), ou encore plus en amont les documents SMIL [32]. Dans ces documents, les contenus multimédias sont généralement organisés dans un espace de présentation et synchronisés dans le temps. L'utilisateur a en outre la possibilité d'interagir avec la présentation et notamment avec tous les éléments multimédias qui la compose.

En parallèle, de nombreux dispositifs mobiles ont commencé à se répandre auprès du grand public. À cette époque au début des années 2000, on parlait de téléphones portables ou de GSM (c'est-à-dire des téléphones mobiles permettant uniquement de téléphoner ainsi que d'envoyer des messages de types SMS ou MMS). Les assistants numériques personnels (PDA), ancêtre du smartphone, étaient moins répandus mais ils permettaient avec leurs écrans tactiles de se connecter au réseau Internet sans fil (comme le Wifi), d'installer des applications et de naviguer sur le Web. Bien entendu, ces dispositifs mobiles étaient déjà capables d'afficher des présentations multimédias afin d'y avoir accès à tout moment et de n'importe où. En effet, les MMS exploitaient le langage SMIL³ et les PDA permettaient l'installation de lecteurs multimédias comme RealPlayer⁴ qui exécute ce type de documents.



FIGURE 4.1 – Un exemple d'hétérogénéité de dispositifs mobiles.

2. <https://html.spec.whatwg.org>

3. http://www.openmobilealliance.org/release/MMS/V1_1-20021104-C/OMA-IOP-MMSCONF-V2_0_0-20020206-C.pdf

4. <https://www.real.com>

Comme illustré dans la figure 4.1, les dispositifs mobiles disposent généralement de capacités hétérogènes : des caractéristiques matérielles variées (e.g., taille d'écran, batterie, capacités d'interaction) mais également des capacités logicielles différentes (notamment liées aux applications et aux codecs). De plus, les préférences de l'utilisateur ou leurs contraintes peuvent restreindre l'exécution de certains contenus. Par exemple, un utilisateur désire lire des textes exclusivement en français (besoin en traduction) et/ou avoir une préférence pour des médias textuels lorsqu'il se trouve en réunion de travail (coupure sonore et besoin de sous-titres). Toutes ces restrictions induisent des contraintes, parfois très riches, qui peuvent être spécifiées dans un profil.

Au sein d'un profil, différentes catégories d'informations doivent être gérées : (1) les caractéristiques du terminal (matérielles et logicielles) avec son état courant (e.g., batterie, luminosité de l'écran, qualité de réception du réseau sans fil...), (2) des informations de contexte liées, par exemple, à des interactions entre l'utilisateur et l'appareil, telles que les préférences de langues, la localisation ou le niveau sonore ambiant, et (3) les informations liées aux constituants des documents, comme les types de contenus qui peuvent être joués ou bien les possibles restrictions de présentation (i.e., mises en page spécifiques, synchronisation temporelle préférée de contenus multimédias...). Par conséquent, si un document multimédia n'est pas conforme à certaines contraintes spécifiées dans un profil, ce document ne sera pas traité correctement sur le terminal cible. Ainsi, pour que les documents multimédias soient accessibles sur de multiples plateformes hétérogènes, ils doivent être adaptés. C'est-à-dire qu'ils doivent être transformés pour satisfaire toutes les contraintes des plateformes cibles.

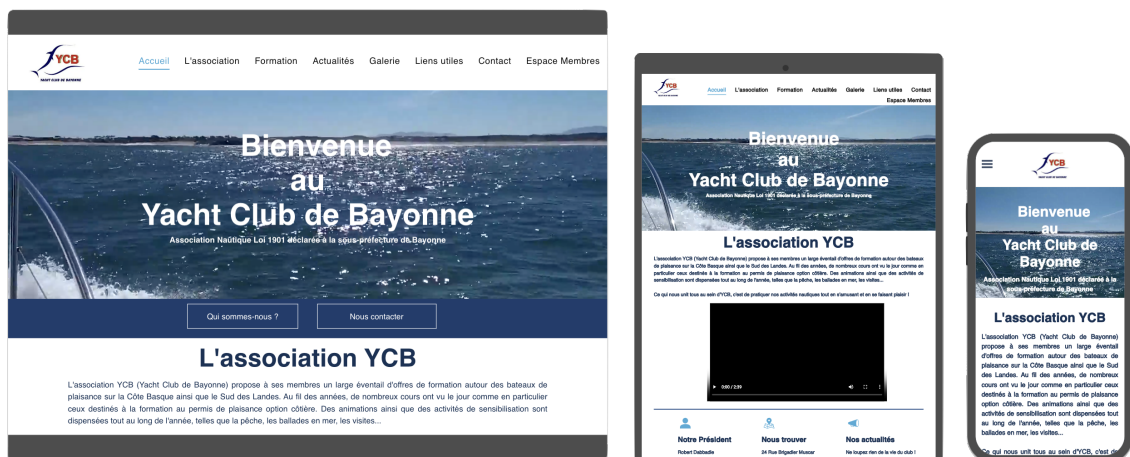


FIGURE 4.2 – Un exemple d'adaptation d'un document multimédia.

La figure 4.2 illustre un document multimédia qui est la page web du Yacht Club de Bayonne (www.yachtclubdebayonne.fr). Pour que cette page web s'affiche correctement sur différents dispositifs, il est nécessaire de l'adapter en fonction des profils cibles. Un ordinateur permettra d'afficher toutes les informations, tandis qu'un dispositif plus limité (e.g., tablette, smartphone) nécessitera de modifier l'affichage.

Dans la partie suivante, nous présentons un état de l'art au sujet de la modélisation des profils ainsi que des méthodes d'adaptation. L'accent sera mis pour illustrer le principe gagnant-gagnant évoqué dans l'introduction de ce chapitre. Par exemple, d'un côté nous montrerons que la communauté du Multimédia (dont je fais partie) a eu un réel besoin de modélisation des profils. D'un autre côté, la communauté du Web Sémantique (dont je fais également partie) a su répondre à ce besoin mais d'une façon limitée. En me retrouvant à la croisée des chemins de l'une et de l'autre des communautés, j'ai pu proposer une modélisation sémantique de profils qui permet d'inspirer mes deux communautés et qui suivra cet état de l'art. C'est justement ce principe que j'appelle gagnant-gagnant.

4.2 État de l'art

Cet état de l'art au sujet de l'adaptation de documents multimédia sera découpé en deux sous-parties. La première se concentrera sur la modélisation des profils (§4.2.1). En particulier, nous présenterons un standard du W3C qui permet la description de profils à l'aide du formalisme RDF. Nous en montrerons également des extensions avec leurs limites. La seconde présentera des méthodes d'adaptation qui exploitent des profils (§4.2.2).

4.2.1 Modélisation des profils

Le groupe de travail Device Independance⁵ du W3C a identifié les pré-requis suivants pour un langage de spécification de profils. En voici les trois principaux :

- **Vocabulaires extensibles** : le langage doit pouvoir être en mesure de supporter divers ajouts de contraintes d'adaptation. En effet, développer des vocabulaires globaux n'est pas concluant puisqu'il est très difficile de déterminer un ensemble fixe de termes pour décrire en avance tous les terminaux possibles, au risque de ne pas penser à ceux qui verront le jour dans le futur.
- **Vocabulaires décentralisés** : un fabricant de terminaux qui désire étendre le vocabulaire de son profil doit souvent passer par une procédure lourde pour le faire évoluer. Le langage doit pouvoir intégrer de nouvelles capacités définies par quiconque sans perturber les définitions de capacités déjà existantes.
- **Simplicité d'intégration des informations à partir de différentes sources** : lors de l'adaptation de documents vis-à-vis d'un profil particulier, l'information dont on a besoin pour l'adaptation peut provenir de différentes sources, comme par exemple le réseau, le terminal, le contexte d'utilisation ou bien les préférences de l'utilisateur. Le service d'adaptation reçoit ces différentes parties d'information séparément et a besoin de les fusionner en un seul modèle avant de faire l'adaptation proprement dite.

5. <https://www.w3.org/2001/di/>

Tenant compte du défi de l'adaptation décrit dans la section 4.1 et des pré-requis précédents pour décrire les profils, la communauté du Web Sémantique s'est impliquée dans la spécification d'un langage standard de description de profils intitulé CC/PP (*Composite Capabilities and Preferences Profiles*) [70]. Cette recommandation du W3C se fonde bien évidemment sur le formalisme RDF. La figure 4.3 illustre un exemple d'un tel profil sérialisé en RDF/XML.

Un profil CC/PP est construit selon une hiérarchie à deux niveaux : un profil contient au moins un composant et chaque composant dispose d'au moins un attribut. Le langage CC/PP offre trois principaux composants par défaut relatifs à la plate-forme matérielle (`TerminalHardware`), aux capacités logicielles (`TerminalSoftware`) et aux applications individuelles comme un navigateur (`TerminalBrowser`). Ce type de profil peut donc contenir un ou plusieurs composants, qui contiennent eux-mêmes un ou plusieurs attributs. Bien entendu, différents composants et attributs ont été définis pour divers usages. Par exemple dans la figure 4.3, le profil précise au sein du composant lié au matériel une largeur (`displayWidth`) ainsi qu'une hauteur (`displayHeight`) d'écran. Pour le composant logiciel, il mentionne le système d'exploitation mobile utilisé (EPOC 2.0 de Symbian). Enfin, le navigateur de la plate-forme cible est la version 5.0 de Mozilla qui supporte les versions HTML 3.2 et 4.0 (`htmlVersionsSupported`).

Comme nous venons de le voir, la communauté du Web Sémantique a donc apporté pour la communauté du Multimédia une pierre à l'édifice de l'adaptation en permettant la spécification de profils en RDE. Dans ce qui suit, nous allons montrer que la communauté du Multimédia a su en tirer partie pour en définir des extensions.

Grâce à l'émergence de CC/PP, de nouvelles extensions ont pu voir le jour via des vocabulaires spécifiques complémentaires à la recommandation initiale du W3C. Par exemple, UAProf (*User Agent Profile*) [134] développé par l'Open Mobile Alliance⁶ est un vocabulaire CC/PP spécifique dédié principalement à la description des caractéristiques des téléphones portables. Des millions d'appareils mobiles de Nokia ou de Samsung ont pu fournir de telles descriptions.

D'autres approches que ce soit pour l'adaptation de contenus multimédias ou bien pour l'adaptation de services web se sont inspirées du modèle CC/PP pour l'étendre. Par exemple, UPS (*Universal Profiling Schema*) [84] qui dispose d'une structure similaire à celle de CC/PP fournit un modèle de spécification de contraintes qui permet l'expression de formules logiques couplé à un processus de traitement des expressions qui facilite la tâche des fournisseurs de contenus multimédia suite à la réception du profil. AWSDL [128] a défini deux nouveaux composants CC/PP (`UserCharacteristic` et `UserContext`) afin de décrire les caractéristiques permanentes et évolutives de l'utilisateur ainsi que son contexte environnant pour adapter des services web. CSCP [31] va encore plus loin dans l'exploitation du formalisme RDF en définissant le concept de sessions qui permet de mettre à jour uniquement certaines valeurs du profil. Pour ce faire, il est structuré et géré de façon modulaire grâce à des sous-profils dédiés référencés avec des URIs ce qui permet un stockage et un accès plus optimal des données.

6. <https://omaspecworks.org>

```

1 <?xml version="1.0"?>
2 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3   xmlns:ccpp="http://www.w3.org/2002/11/08-ccpp-schema#"
4   xmlns:ex="http://www.example.com/schema#">
5   <rdf:Description rdf:about="http://www.example.com/profile#
6     MyProfile">
7
8     <ccpp:component>
9       <rdf:Description
10        rdf:about="http://www.example.com/profile#TerminalHardware">
11         <rdf:type
12          rdf:resource="http://www.example.com/schema#HardwarePlatform"/>
13         <ex:displayWidth>320</ex:displayWidth>
14         <ex:displayHeight>200</ex:displayHeight>
15       </rdf:Description>
16     </ccpp:component>
17
18     <ccpp:component>
19       <rdf:Description
20        rdf:about="http://www.example.com/profile#TerminalSoftware">
21         <rdf:type
22          rdf:resource="http://www.example.com/schema#SoftwarePlatform"/>
23         <ex:name>EPOC</ex:name>
24         <ex:version>2.0</ex:version>
25         <ex:vendor>Symbian</ex:vendor>
26       </rdf:Description>
27     </ccpp:component>
28
29     <ccpp:component>
30       <rdf:Description
31        rdf:about="http://www.example.com/profile#TerminalBrowser">
32         <rdf:type
33          rdf:resource="http://www.example.com/schema#BrowserUA"/>
34         <ex:name>Mozilla</ex:name>
35         <ex:version>5.0</ex:version>
36         <ex:vendor>Symbian</ex:vendor>
37         <ex:htmlVersionsSupported>
38         <rdf:Bag>
39           <rdf:li>3.2</rdf:li>
40           <rdf:li>4.0</rdf:li>
41         </rdf:Bag>
42         </ex:htmlVersionsSupported>
43       </rdf:Description>
44     </ccpp:component>
45   </rdf:Description>
46 </rdf:RDF>

```

FIGURE 4.3 – Un exemple de profil CC/PP sérialisé en RDF/XML

Dans toutes ces extensions et même en ce qui concerne la recommandation CC/PP, un profil est vu comme une structure hiérarchique qui contient des informations quantitatives sur les capacités des plateformes et le contexte de l'utilisateur. Il est étonnant de constater que contrairement à ce que nous avons pu présenter dans le Chapitre 2 pour RDE, la structure sous forme de graphes de description n'est pas envisagée. De plus, la sémantique des données que ce soit au niveau des composants ou bien au niveau des contraintes n'est pas non plus exploitée. Dans la section 4.3, nous montrerons qu'une autre vision du profil est possible tout en exploitant pleinement les capacités offertes par le Web Sémantique.

4.2.2 Des approches pour adapter des documents multimédia

Lorsque j'ai débuté mes travaux de recherche sur l'adaptation de documents multimédia [71], toutes les approches existantes ont pu être regroupées selon trois catégories possibles [81]. La première catégorie consiste à définir des alternatives de présentation (e.g., [32, 51, 105]). C'est un travail fastidieux puisque l'on doit anticiper les différentes présentations du document sur toutes les plates-formes cibles potentielles. La seconde catégorie propose de disposer d'une collection de services d'adaptation qui peuvent être éventuellement appliqués aux documents (e.g., [22, 24, 39, 65, 83, 85]). Il est par conséquent nécessaire d'avoir une certaine diversité de services et donc de pouvoir les décrire afin de les enchaîner pour appliquer les modifications nécessaires. Enfin, la troisième catégorie ne considère pas des documents multimédia directement exécutables mais plutôt des abstractions de documents (e.g., [26, 27, 53]). Ces modélisations que l'on pourrait qualifier de "haut niveau" permettent par la suite la génération de documents exécutables lorsque le profil est défini. L'auteur du document doit donc maîtriser cette modélisation spécifique qui ne correspond pas aux standards.

Nos contributions qui seront décrites dans la partie suivante (§4.3) se situent dans la seconde catégorie d'approches décrite précédemment, c'est-à-dire celle basée sur une collection de services d'adaptation. En partant d'un profil cible, toutes les approches de cette catégorie combinent généralement trois types de services d'adaptation que l'on peut voir comme des opérateurs de modification du document : (a) la transformation qui agit sur les objets de la présentation et/ou son organisation générale (e.g., résumé ou traduction d'un contenu, changement de mise en page), (b) le transcodage qui modifie le type d'encodage d'un objet (e.g., modification d'une vidéo encodée en AVI vers un format en MPEG) et (c) le transmodage qui procède au changement d'apparence d'un objet (e.g., modification d'un contenu textuel en contenu sonore). Il va de soi que la richesse du profil est exploitée pour déterminer un enchaînement pertinent de ces opérateurs qui permettra d'obtenir un document adapté dont l'exécution se fera correctement sur la plate-forme cible.

Dans la partie suivante, nous montrerons que nous avons contribué dans ce cadre sur deux points qui sont (1) la modélisation des profils ainsi que (2) la modélisation des services d'adaptation. L'apport de la sémantique à nos modélisations constitue l'originalité de nos travaux.

4.3 Notre proposition de modélisation sémantique

À l’instar de la partie qui concerne l’état de l’art, nos contributions au sujet de l’adaptation seront découpées en deux sous-parties. La première illustrera une proposition de modélisation sémantique de profils (§4.3.1). Tandis que la seconde présentera notre proposition de modélisation sémantique du processus d’adaptation (§4.3.2).

4.3.1 Un profil sémantique générique pour décrire un contexte

Pour que l’adaptation soit efficace, un profil doit fournir des informations à jour sur les capacités des plateformes, le contexte de l’utilisateur ainsi que les caractéristiques des documents que le terminal est en mesure de prendre en compte. En règle générale (§4.2.1), les descriptions d’un profil se composent de valeurs simples, telles que la taille de l’écran, la langue préférée, la qualité de la bande passante, la puissance de la batterie, la localisation, etc. Cet usage du profil a plusieurs inconvénients : (1) toutes les informations du profil doivent être très régulièrement mises à jour, (2) il n’est pas possible de migrer un profil d’une plateforme vers une autre sans régénérer entièrement le profil avec ses contraintes, (3) les valeurs sont exprimées de façon quantitatives alors que des valeurs qualitatives pourraient être exploitées (e.g., batterie "*faible*" ou température "*élevée*").

Ainsi, nous avons proposé une autre vision possible du profil [41, 42, 43, 44, 61, 72] qui établit plutôt des références vers des services de données exploitables à la demande. Ces services peuvent être regroupés selon différentes facettes selon qu’ils traitent d’informations au sujet du contexte de l’utilisateur (*Context services*), des contenus multimédias (*Document services*) ou de la plateforme utilisée (*Device services*). La figure 4.4 illustre cette vision du profil à base de services organisés en facettes. Dans cet exemple, la facette relative au contexte de l’utilisateur peut contenir différentes catégories de services (e.g., *User services* ou *Weather services*) ce qui permet de mieux les retrouver lors de l’analyse du profil par un processus d’adaptation (e.g., *Existe-t-il un service disponible lié à la météo de l’utilisateur?*). Bien évidemment, les services à proprement parler (e.g., *User location service* ou *User language service*) sont affectés aux différentes catégories de services des facettes.

Afin de mieux guider le processus d’adaptation, des contraintes explicites peuvent être modélisées sur notre proposition de profil. En effet, étant donné que nous exploitons des services, nous avons la possibilité de récupérer des informations sur ces derniers (via des méthodes de type *Get*) mais aussi d’exprimer des actions sur ceux-ci (via des méthodes de type *Set*). Bien entendu, des contraintes riches peuvent être définies puisqu’il est tout à fait possible d’exprimer des conjonctions (\wedge) ou des disjonctions (\vee) entre des évaluations de données sur différents services, de préciser des successions ou des priorités sur les actions à envisager sur des services... Par exemple, comme nous pouvons le constater dans la figure 4.4, la contrainte explicite suivante est décrite (*Constraint 1*) : « *Si l’utilisateur est localisé dans une zone décrite via des coordonnées GPS et que la batterie de son dispositif est inférieure à 10%, alors couper le son des*

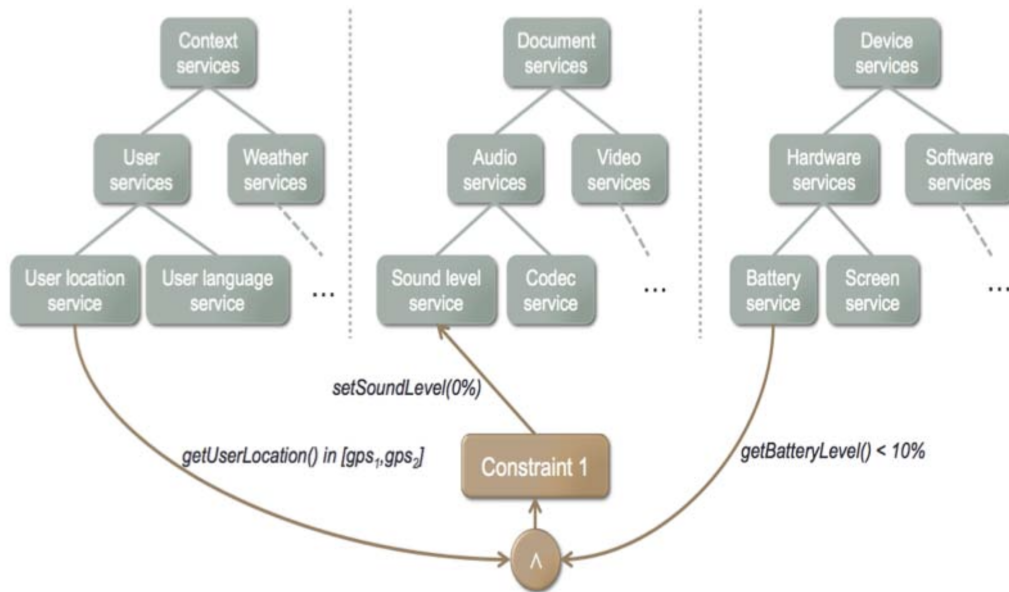


FIGURE 4.4 – Une autre vision du profil à base de services

contenus sonores ». Ici, on remarque l’usage de deux méthodes de type *Get* sur deux services différents (*getUserLocation()* et *getBatteryLevel()*) ainsi qu’une méthode de type *Set* sur un service qui gère les contenus sonores (*setSoundLevel(0%)*). Visuellement, notre vision du profil commence à prendre la forme d’un graphe de description qui met en jeu différentes ressources au travers de relations.

Au-delà de la structure sous forme de graphe, notre proposition exploite également différents niveaux d’expressivité. Par exemple dans la figure 4.4, la contrainte fait référence à des données quantitatives, telles que 10% ou bien un ensemble de coordonnées GPS. En réalité, on pourrait vouloir exprimer que « *si l’utilisateur est proche d’un lieu donné et que la batterie est faible, alors couper le son des contenus sonores* ». Pour permettre l’exploitation de valeurs qualitatives et combler le fossé entre données quantitatives et qualitatives, nous avons proposé la modélisation multi-niveaux suivante (Figure 4.5) [72] :

- **Niveau 1** : Il s’agit du plus haut niveau sémantique. Celui-ci fait référence aux termes qualitatifs à employer dans le profil. Dans la figure 4.5, nous souhaitons par exemple définir le terme « *Low* ».
- **Niveau 2** : Ce niveau correspond aux services potentiels du profil qui peuvent être liés aux termes du niveau précédent. Par exemple, le terme « *Low* » peut être lié aux services qui concernent la batterie et la bande passante.
- **Niveau 3** : Pour un terme qualitatif et un service donné, différents domaines d’applications peuvent mener à différentes interprétations possibles. Dans l’exemple de la figure 4.5, le smartphone ou la tablette peuvent indiquer différentes interprétations du terme « *Low* » pour la batterie, respectivement <12% ou <15%.

- **Niveau i** ($3 < i < n$) : Des sous-domaines d'applications peuvent être identifiés. Par exemple, différentes interprétations possibles du terme « *Low* » pour une batterie associée à certains modèles de téléphones ou de tablettes.
- **Niveau n** : Ce dernier niveau fait référence aux valeurs quantitatives. En effet, pour un terme qualitatif donné, un service associé et des domaines ou sous-domaines identifiés, il faudra renseigner des valeurs quantitatives correspondantes. Par exemple dans la figure 4.5, une batterie faible pour une tablette équivaut à disposer d'un niveau de batterie inférieur à 15%.

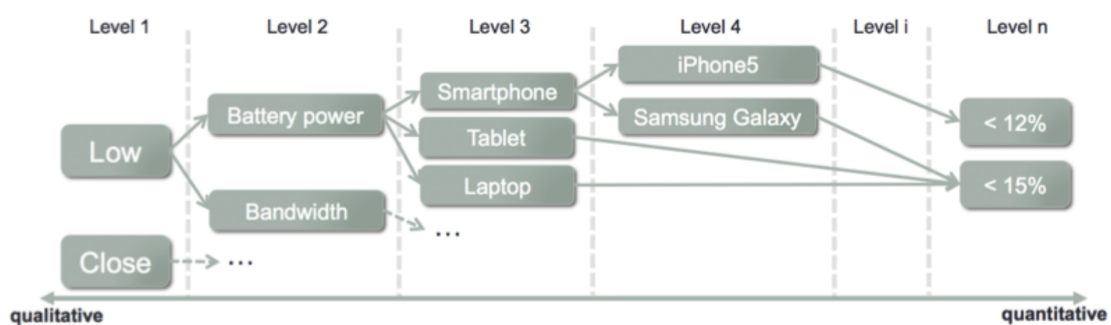


FIGURE 4.5 – Comblent le fossé entre valeurs quantitatives et qualitatives des profils

Au regard de la recommandation CC/PP pour laquelle la communauté du Web Sémantique a contribué et de ses extensions proposées par la communauté du Multimédia (§4.2.1), l'exploitation du formalisme RDF pour décrire un profil est un choix pertinent dès le départ mais notre proposition permet vraiment de l'exploiter pleinement. En effet, notre vision du profil à base de services lève plusieurs limites et met en lumière ce principe gagnant-gagnant que nous souhaitons illustrer dans ce chapitre.

Tout d'abord, notre modélisation sous forme de services et de contraintes associées dessine des graphes de description ce que nous avons pu démontrer visuellement dans la figure 4.4. Bien entendu, les URI couplées au mécanisme d'agrégation de graphes RDF vu dans le Chapitre 2 nous permettent de gagner en modularité. Par exemple, il est possible dans notre proposition de décrire un graphe RDF par facette avec leurs services d'un côté et d'un autre côté les différentes contraintes associées aux ressources. Ensuite, les informations décrites dans notre proposition peuvent être associées à des concepts sémantiques ce que nous avons pu établir par exemple dans la figure 4.4 pour décrire les facettes ainsi que les services du profil mais également dans la figure 4.5 pour exploiter des données qualitatives. Enfin, notre vision originale à base de services évite les mises à jour régulières du profil, permet aux processus d'adaptation de récupérer exclusivement les données qui lui sont nécessaires et il dispose de la capacité à migrer d'une plate-forme à une autre tout en préservant les contraintes exprimées.

Dans la partie suivante, nous montrons que le processus d'adaptation lui-même peut aussi être modélisé sémantiquement ceci lui permettant notamment d'exploiter les capacités d'inférence offertes par le Web Sémantique.

4.3.2 Modélisation sémantique du processus d'adaptation

Au-delà de la modélisation sémantique des profils que nous avons proposé (§4.3.1), nous souhaitons aller plus loin dans l'exploitation du Web Sémantique au sein même du processus global d'adaptation de documents multimédia. Pour ce faire, nous avons proposé de modéliser sémantiquement les services d'adaptation à base d'une ontologie (§4.3.2.1). Afin d'assembler certains services entre eux, nous avons exploité un raisonneur au sein d'un système d'adaptation de documents multimédia que nous avons développé (§4.3.2.2). Notre objectif vise donc à mieux sélectionner les services d'adaptation mais aussi à mieux enchaîner ces services grâce aux mécanismes d'inférence sémantique.

4.3.2.1 Une ontologie pour les services d'adaptation multimédia

Aux alentours de 2010, plusieurs approches ont proposé d'enrichir sémantiquement les descriptions des services que ce soit de façon générale (e.g., [48, 49]) ou bien en tenant compte d'informations sur le contexte d'usage des services (e.g., [28, 75]). En effet, si la description est nécessaire pour l'invocation et la composition de services, l'information sémantique au sujet des paramètres de fonctionnement des services est également importante (i.e., sémantique des paramètres d'entrée-sortie, coût et disponibilité du service, contraintes d'exécution, temps d'exécution, qualité du service, etc.). En 2012, nous proposons une ontologie qui consiste à améliorer la description sémantique des services d'adaptation pour faciliter leur sélection et leur combinaison dans le cadre d'un processus d'adaptation automatique de documents multimédia [18].

La figure 4.6 présente la structure générale de notre ontologie pour décrire des services d'adaptation avec leurs contraintes contextuelles de fonctionnement. Les fournisseurs vont instancier les classes de l'ontologie et publier les instances résultantes sous forme de fichiers OWL ⁷.

Comme on peut le voir dans cette figure, un `Service` dispose d'un nom, d'une version et d'un identifiant. Le concept `Service Role` permet de préciser s'il s'agit d'un mécanisme d'adaptation lié à de la transformation de contenu, à du transcodage ou bien à du transmodage (§4.2.2). Les types de contenus multimédias pris en charge par le service sont précisés dans `Multimedia Content`. Il est également possible de connaître l'état courant du service (`Step`), sa dernière période d'exécution (`Service Time`)... La relation `service_Relation` représente les liens de dépendance, d'équivalence et de substitution entre différents services. Les relations de dépendance entre services (`Dependence_Relation`) précisent les successions ainsi que les précédences possibles entre services (e.g., un service de compression peut être suivi d'un service de décompression). Les relations d'équivalence (`Equivalence_Relation`) signifient que des services jouent le même rôle. Enfin, les substitutions (`Substitution_Relation`) indiquent qu'il est possible de remplacer un service par un autre même si par exemple son coût d'exécution est différent ou bien que d'autres versions sont disponibles.

7. <https://www.iutbayonne.univ-pau.fr/~roose/ontologies/ASQOntology.owl>

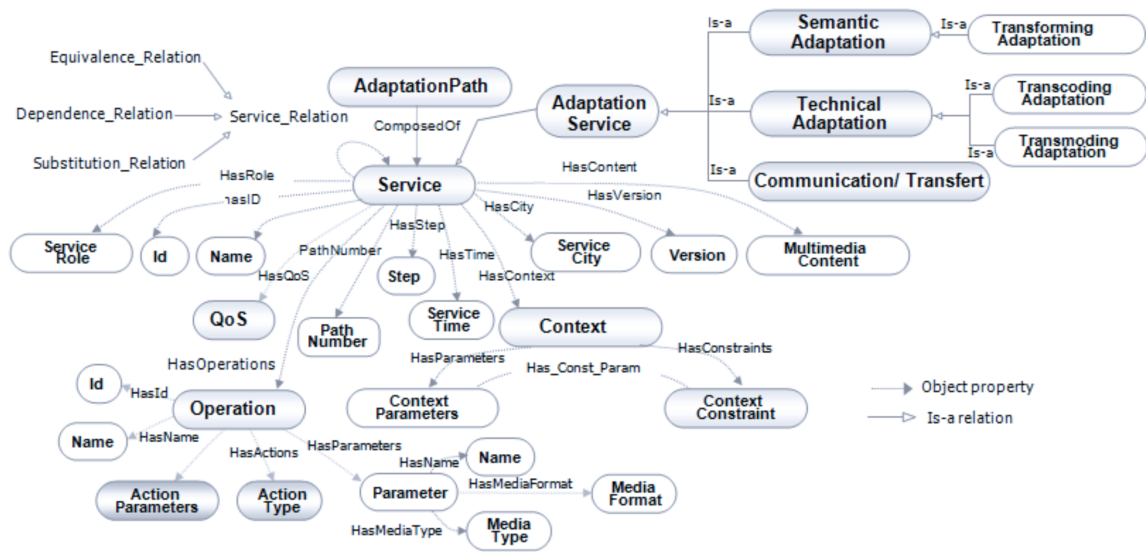


FIGURE 4.6 – Une ontologie pour guider le processus d’adaptation.

Toutes ces informations vont donc permettre de guider l’assemblage automatique des services d’adaptation entre eux. De plus, lorsque les configurations de certains services ne vont pas permettre leurs exécutions effectives dans la chaîne de traitement (AdaptationPath), grâce aux relations entre services des ajustements pourront être effectués en remplaçant les services inopérants par d’autres qui offrent des fonctionnalités similaires quand bien même leur qualité de service (QoS) sera moindre.

Dans la partie suivante, nous allons exploiter ces descriptions sémantiques au sein d’un processus d’adaptation de documents multimédia qui exploitera un raisonneur.

4.3.2.2 Un processus d’adaptation exploitant un moteur d’inférence

Grâce à nos descriptions sémantiques précédentes, c’est-à-dire celles qui concernent les profils (§4.3.1) et celles qui concernent les services d’adaptation (§4.3.2.1), notre processus d’adaptation a la capacité d’exploiter un moteur d’inférence sémantique chargé de raisonner, d’optimiser et de planifier la chaîne de traitement pour mettre en œuvre les transformations nécessaires aux documents sur un serveur multimédia.

La figure 4.7 illustre les principales étapes de ce processus [9]. Notre serveur multimédia qui distribue les contenus aux dispositifs des utilisateurs est constitué d’un gestionnaire de contraintes, d’un planificateur, d’un composant de décision et d’exécution. Le gestionnaire de contraintes analyse le profil et détecte les conflits pouvant avoir lieu avec les ressources multimédia demandées. Selon les conflits, un guide d’adaptation est généré qui liste des directives d’adaptation afin que le document puisse s’exécuter correctement sur la plate-forme cible. Par exemple, si un conflit de format est détecté, le guide proposera de le modifier en un autre format autorisé dans le profil.

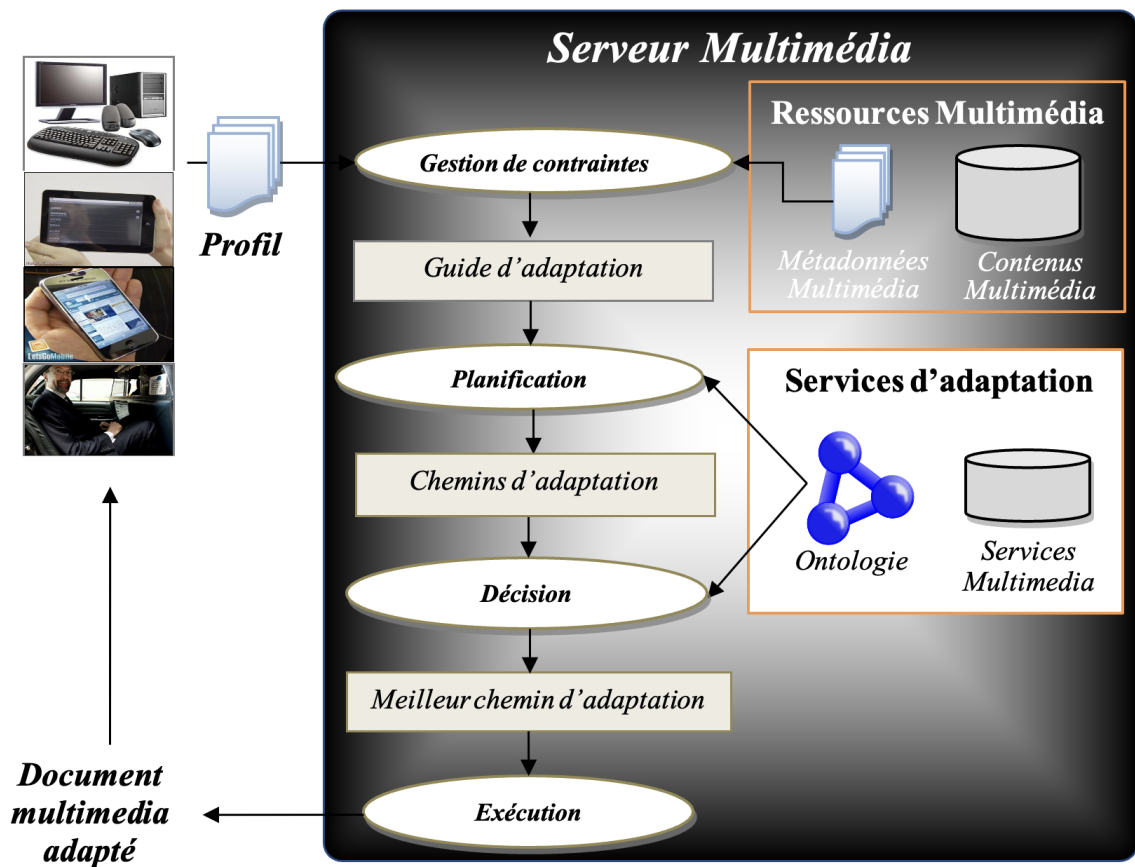


FIGURE 4.7 – Un processus d’adaptation exploitant le mécanisme d’inférence.

Le planificateur identifie les services d’adaptation potentiels et détermine des combinaisons de ces derniers en fonction du guide d’adaptation et des informations sémantiques liées aux services d’adaptation (e.g., le rôle du service, le type ainsi que le format du média en entrée et en sortie, les ressources nécessaires pour exécuter le service, la disponibilité et la qualité du service, etc.). Le composant de décision et d’exécution sélectionne la meilleure combinaison de services d’adaptation et exécute la séquence afin de fournir un document multimedia adapté.

Généralement, les processus d’adaptation définissent leurs propres algorithmes spécifiques pour calculer les combinaisons de services d’adaptation à exécuter. Dans notre cas, nous avons souhaité produire ces combinaisons à l’aide d’un raisonneur utilisé dans la communauté du Web Sémantique. Pour ce faire, le gestionnaire de contraintes ainsi que le planificateur exploitent des règles spécifiées à l’aide du langage SWRL (*Semantic Web Rule Language*) [68]. SWRL est un langage qui permet d’intégrer des règles d’inférences dans les ontologies. Les règles peuvent être écrites dans l’éditeur Protégé⁸ et ensuite être exécutées via un raisonneur, tel que Pellet⁹.

8. <https://protege.stanford.edu>

9. <https://www.w3.org/2001/sw/wiki/Pellet>

Une règle SWRL comprend une ou plusieurs conditions ainsi qu'une ou plusieurs conséquences, séparées par une flèche « \rightarrow ». Si la règle possède plusieurs conditions ou conséquences, elles sont séparées entre elles par une virgule « , » qui a le sens d'une conjonction (\wedge). De plus, les règles SWRL utilisent des variables, dont les noms commencent par « ? », comme pour le langage SPARQL (§2.4). Ces variables peuvent représenter des individus ou des valeurs, mais jamais des classes ou des propriétés.

Description	Traduction avec le langage SWRL
Déterminer des services de transcodage qui s'exécutent en amont et en aval d'autres services de transcodage	<pre> Transcoding_Adaptation(?s1), has_Set_Of_Parameters(?s1, ?p1), has_Parameter_Out(?p1, ?out1), has_Type(?out1, ?type_out1), has_Format(?out1, ?format_out1), Transcoding_Adaptation(?s2), has_Set_Of_Parameters(?s2, ?p2), has_Parameter_In(?p2, ?in2), has_Type(?in2, ?type_in2), has_Format(?in2, ?format_in2), swrlb:equal(?type_out1, ?type_in2), swrlb:equal(?format_out1, ?format_in2) \rightarrow has_Successor(?s1, ?s2), has_Predecessor(?s2, ?s1) </pre>

TABLE 4.1 – Un exemple de règle SWRL.

Par exemple, la table 4.1 illustre une règle SWRL exploitée dans notre processus d'adaptation. Cette règle infère pour des services de transcodage que si les paramètres de sortie d'un service 1 (?s1) sont compatibles avec les paramètres d'entrée d'un service 2 (?s2) alors le service 2 peut succéder au service 1 (`has_Successor(?s1, ?s2)`) et le service 1 peut précéder le service 2 (`has_Predecessor(?s2, ?s1)`). Ce type de règle va permettre au processus d'adaptation de déterminer des chaînes de traitement qui permettront par la suite d'adapter un contenu multimédia.

Dans la partie suivante, nous allons montrer que la sémantisation des profils ainsi que du processus d'adaptation telle que nous l'avons présenté dans les sections 4.3.2.1 et 4.3.2.2 a permis d'ouvrir la voie vers de nombreuses publications que ce soit pour effectuer des expérimentations dans différents contextes d'usage mais également pour faire évoluer notre proposition d'adaptation vers d'autres domaines d'applications que ceux dédiés aux documents multimédia.

4.4 Expérimentations

Depuis 2012, une trentaine de publications scientifiques ont été diffusées en ce qui concerne notre modélisation de profils sémantiques avec le développement de plateformes d'adaptation. Ce travail a tout d'abord été initié en collaboration avec Philippe

Roose du LIUPPA et Adel Alti de l'Université de Ferhat Abbas de Sétif en Algérie. Nos premières propositions d'ontologies et d'implémentations ont été détaillées dans les publications suivantes : [5, 6, 9, 10, 11, 12, 18]. Bien entendu, des expérimentations ont été réalisées comme illustré dans la figure 4.8. Cette figure présente le temps de calcul pour déterminer une chaîne de services d'adaptation sur un smartphone de type LG540 avec un système Android OS 2.1. Naturellement, plus le nombre de services d'adaptation mis en jeu est important, plus la recherche de combinaisons augmente. Dans une précédente approche [60], nous avons proposé un algorithme de composition de service d'adaptation pour des plates-formes de type Android sans exploitation d'ontologies (courbe rose de la figure 4.8). Notre proposition à base d'ontologies et de raisonnement sémantique démontre son efficacité en terme de temps de recherche de combinaisons de services d'adaptation (courbe bleue foncée de la figure 4.8).

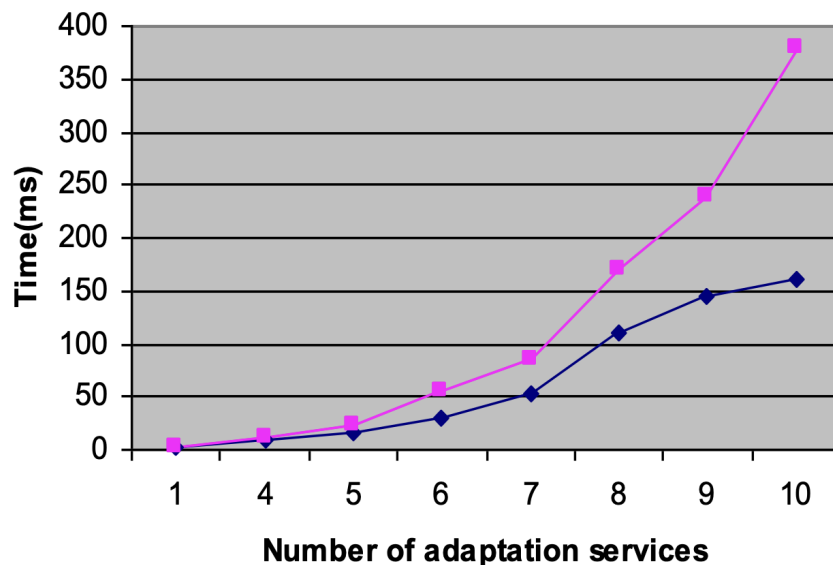


FIGURE 4.8 – Déterminer une chaîne de services d'adaptation sur Android.

Suite à l'encadrement de deux étudiants (Cédric Dromzée puis Hajar Khallouki), nous avons pu étendre et expérimenter nos travaux sur la description de profils sémantiques [41, 42, 43, 44, 61, 72]. Par exemple, en alimentant semi-automatiquement les facettes et les contraintes des profils. Ces propositions ont pu par la suite être appliquées dans le domaine de la santé [2, 8, 7, 16, 82]. La thèse d'Asma Saighi intitulée "*Qualité de service pour les adaptations multimédia dans un environnement mobile*" [107] a permis d'établir une ontologie de contexte de plus haut niveau que nos propositions précédentes car elle se base sur la notion d'handicap afin de toujours guider l'adaptation de documents multimédia. Chaque contrainte contextuelle est liée à un handicap physique (e.g., visuel, sonore). Cette abstraction du profil permet de se détacher des données quantitatives et qualitatives des profils classiques pour optimiser la recherche d'actions d'adaptation grâce à un raisonnement sémantique [108, 110, 111]. Des expérimentations ont été réalisées dans le cadre des maisons intelligentes [109].

Sur la base de notre expérience en matière de modélisation de profils et de développement de plates-formes d'adaptation exploitant des technologies issues du Web Sémantique, des évolutions de nos propositions ont abouti dans différents domaines. Tout d'abord, nous avons proposé d'exploiter les réseaux sociaux [3, 13, 15] ainsi que les technologies du Cloud [1, 14, 19, 91, 95] pour adapter des contenus. Ensuite, comme pour les documents multimédia, nous avons proposé d'étendre notre approche d'adaptation à base d'ontologies pour adapter des applications [17, 73, 74, 119]. En effet, une application peut disposer d'un profil d'exécution qui peut varier d'un contexte d'usage à un autre.

Plus récemment, mon équipe de Recherche au LIUPPA (T2i) s'intéresse au domaine de l'informatique éco-responsable. En effet, la sensibilisation à la consommation d'énergie a motivé les chercheurs à développer des méthodologies qui nous insistent au développement d'applications dites "vertes". Nous avons proposé en ce sens une série de propriétés pour qualifier cette consommation d'énergie sur une application et ses services. Ce travail est engagé dans la thèse de Jorge Andres Larracochea qui s'intitule "*Vers une méthodologie et une DSL pour une informatique verte*". In fine, notre objectif consiste à spécifier des profils pour le Green-IT [76, 77, 78, 79, 80, 104]. Dans ces résultats, la partie sémantique n'a pas encore été abordée car il s'agira d'une contribution future à ces travaux publiés.

Tout ceci démontre bien que notre principe gagnant-gagnant exposé au début de ce chapitre a porté ses fruits. Dans la partie suivante, nous concluerons sur ce leitmotiv.

4.5 Synthèse du chapitre

Dans ce chapitre, j'ai démontré comment nous avons su exploiter le Web Sémantique au sein d'un principe que nous avons qualifié de gagnant-gagnant. J'ai notamment illustré ce principe à l'aide de travaux de recherche qui traitent de la modélisation de profils pour l'adaptation de documents multimédia.

Tout d'abord dans la section 4.1, j'ai présenté le défi de l'adaptation de documents multimédia que j'avais commencé à investir dans mes premiers travaux de recherche durant mon doctorat. Modéliser les profils des plates-formes avec les préférences des utilisateurs est un pré-requis à l'adaptation. Définir de nouvelles approches d'adaptation qui exploitent la sémantique des profils est un défi que nous avons su relever.

Dans la section 4.2, nous avons présenté le langage standard CC/PP proposé par le W3C pour décrire des profils. Cette recommandation s'appuie sur le formalisme RDF et la communauté du Multimédia a su l'exploiter. Néanmoins, nous avons relevé des limites à ce standard avec ses extensions. D'un côté, nous insistons donc sur l'amélioration du standard en exploitant pleinement les capacités offertes par le Web Sémantique. D'un autre côté, nous montrons les nouvelles possibilités offertes à la communauté du Multimédia pour l'adaptation. C'est ce principe que nous appelons gagnant-gagnant. Dans cette partie, nous avons également présenté différentes catégories d'approches d'adaptation de documents multimédia.

Dans la section 4.3, nous avons proposé une nouvelle modélisation sémantique des profils qui permet de guider l'adaptation de contenus. Nous sommes allés également plus loin dans la sémantisation puisque nous avons proposé d'appliquer les technologies issues du Web Sémantique au processus d'adaptation lui-même, c'est-à-dire celui qui permet de composer les différents services d'adaptation. Pour ce faire, nous avons proposé une ontologie décrivant ces services et l'usage du langage SWRL afin d'exploiter les mécanismes d'inférence pour leur composition.

Des expérimentations ainsi que des évolutions de nos propositions ont été exposées dans la section 4.4. Une trentaine de publications ont été diffusées dans des conférences ainsi que dans des journaux à audience nationale et internationale. En ce qui concerne les articles, plusieurs d'entre eux ont été publiés dans des journaux de rang Q2 au regard du classement Scimago (e.g., [2, 9, 17, 91]). Également, on peut noter la publication récente de 2022 dans une conférence de rang A au regard du classement CORE :

- Jorge Andrés Larracochea, Philippe Roose, Sergio Ilarri, Yudith Cardinale, **Sébastien Laborie** et Olga Paulina Vara. *Behavior-Based Consumption Profiles for the Approximation of the Energy Consumption of Services*. In Proc. of the 30th Int. Conf. on Information Systems Development (Cluj-Napoca, Romania), 2022.

Dans ce chapitre, je me suis essentiellement concentré sur l'exploitation du Web Sémantique au service des profils et de l'adaptation. Ce focus m'a permis de mettre en lumière un principe gagnant-gagnant au service des communautés, i.e., ici celle du Web Sémantique avec celle du Multimédia.

Néanmoins, au sein de l'équipe T2i du LIUPPA et en parallèle aux travaux décrits précédemment, j'ai pu collaborer avec d'autres chercheurs qui traitent d'autres thématiques de recherche, comme par exemple celle des EIAH (*Environnements Informatiques pour l'Apprentissage Humain*). En particulier, durant le doctorat de The Nhan Luong intitulé "*Modélisation centrée utilisateur final appliquée à la conception d'applications interactives en géographie*" [87], j'ai pu appliquer ce principe gagnant-gagnant décrit dans ce chapitre. En effet, nous avons proposé une modélisation sémantique d'applications géographiques interactives [45, 88, 89, 92]. Notre modélisation comporte trois facettes en RDF qui permettent de représenter les contenus géographiques manipulés, leur affichage au sein d'une interface graphique et de décrire le comportement de l'application face aux interactions de l'utilisateur. À partir de notre modélisation sémantique, il est possible d'unifier les facettes pour par la suite engendrer automatiquement le code de l'application finale.

Notre principe gagnant-gagnant peut donc s'appliquer à de multiples domaines de recherche. Exploiter le Web Sémantique qui est le titre de ce chapitre n'est donc pas simplement appliquer telles quelles les technologies issues du Web Sémantique dans un domaine spécifique. Il s'agit plutôt d'avoir une conduite à tenir qui consiste d'une part à prendre de la hauteur sur les technologies existantes offertes par le Web Sémantique et à d'autre part en démontrer le potentiel à une communauté cible. C'est ce que je me suis efforcé de faire durant toutes ces années.

Chapitre 5

Fédérer des données hétérogènes à l'aide du Web Sémantique

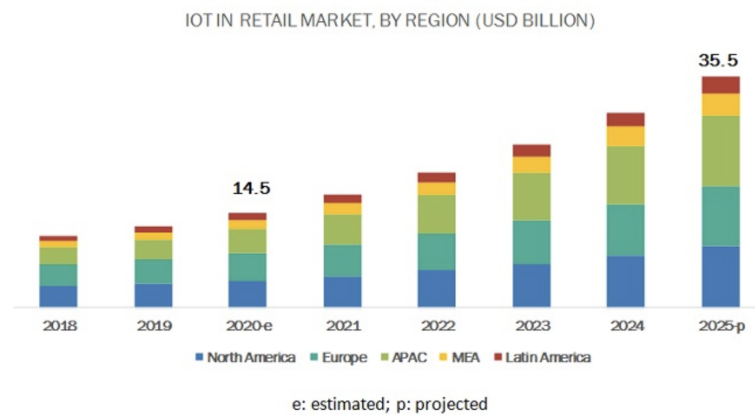
Comme nous avons pu le démontrer dans le chapitre 4, les communautés scientifiques (comme par exemple celle du Multimédia [122]) ont su prendre en main les technologies issues de la communauté du Web Sémantique, comme RDF, OWL et SPARQL. Les communautés qui travaillent dans les domaines de la Recherche d'Information [117] ou encore les Systèmes Cyber-Physiques [106] ont pu également intégrer ces technologies afin d'associer du sens à leurs données tout en perfectionnant leurs systèmes.

Dans ce chapitre, nous allons montrer que nous avons su rapprocher des données exploitées dans différentes communautés grâce aux technologies issues du Web Sémantique. En effet, la recherche d'information au sein d'un réseau de capteurs est un domaine récent. Dans ce contexte, les sources de données peuvent être hétérogènes et complémentaires. Par exemple, la variété des capteurs avec leurs données est considérée dans les Systèmes Cyber-Physiques [98]. La Recherche d'Information étudie également une variété de corpus documentaires [120]. Ces deux catégories de sources de données peuvent cohabiter au sein d'un même système, c'est ce que nous appelons un environnement hybride. Dans ce cadre, le Web Sémantique est une solution possible qui permet de fédérer ces données pour les besoins d'utilisateurs.

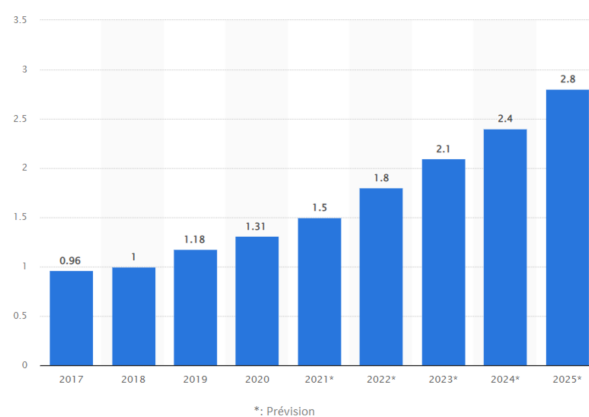
Dans ce qui suit, je vais expliquer comment nous avons pu exploiter les technologies du Web Sémantique pour lier les données de la communauté des systèmes cyber-physiques avec celles exploitées en recherche d'information. Dans un premier temps, je présenterai le défi de l'exploitation de données hétérogènes au sein d'un environnement hybride, c'est-à-dire un environnement qui mêle à la fois des données provenant de capteurs avec un corpus documentaire (Section 5.1). Dans un second temps, je montrerai que chaque communauté a développé des outils complémentaires qui nous servent pour fédérer ces données (Section 5.2). Dans une troisième partie, nous présenterons notre approche qui permet de connecter ces données à l'aide du formalisme RDF pour retrouver, par la suite, des informations utiles pour un utilisateur au sein du graphe de connaissance (Section 5.3). La section 5.4 présentera les expérimentations que nous avons menées et la section 5.5 offrira une synthèse de ce chapitre.

5.1 Le défi de l'explication d'événements au sein d'un environnement hybride

Les environnements connectés et plus généralement les systèmes cyber-physiques sont devenus très présents dans notre vie quotidienne. Ils ont un impact grandissant dans différents secteurs d'activités, tels que le commerce, l'agriculture, la construction, la santé et les transports. La figure 5.1a¹ illustre l'augmentation entre 2017 et 2025 des investissements en particulier dans les bâtiments intelligents. En parallèle, l'accessibilité des technologies de l'information et de la communication n'a fait que renforcer la digitalisation de notre société. La figure 5.1b² met en évidence l'augmentation des dépenses mondiales consacrées à l'effort de la digitalisation au sein des entreprises.



(a) Valeur des investissements dans les bâtiments intelligents



(b) Dépenses mondiales consacrées à la digitalisation

FIGURE 5.1 – Digitalisation et investissement dans les environnements connectés

1. <https://www.marketsandmarkets.com/Market-Reports/retail-iot-market-43188550.html>

2. <https://www.statista.com/statistics/870924/worldwide-digital-transformation-market-size/>

Dans ce contexte, plusieurs travaux ont été menés pour proposer des services basés sur l'utilisation de données collectées par des capteurs [124]. La détection d'événements est l'un de ces services. Généralement, lorsqu'un événement se déclenche dans un environnement doté de capteurs, la réaction naturelle d'un utilisateur en charge de la surveillance consiste à contrôler ce qui s'est passé pour agir (par exemple, donner l'alerte), quand nécessaire, et comprendre et expliquer ce qui se passe, dans tous les cas. Pour ce faire, l'utilisateur peut exploiter différents systèmes d'information comme on peut le voir dans la figure 5.2. D'un côté, il doit se concentrer sur les données issues du réseau de capteurs grâce à un système d'information dédié. D'un autre côté, s'il désire des informations contextuelles complémentaires alors il doit fouiller dans une collection documentaire (parfois avec des données peu structurées) à l'aide d'autres types de systèmes qui exploitent d'autres technologies que celle du réseau de capteurs.

Rechercher une explication au sujet d'un événement dans un tel environnement que nous qualifions d'hybride, c'est-à-dire qui mêle à la fois des données de capteurs avec un corpus documentaire qui comporte des documents peu structurés, comporte plusieurs limites. Tout d'abord, les systèmes associés aux capteurs et ceux liés au corpus documentaire sont actuellement indépendants. Par conséquent, l'utilisateur doit effectuer des allers-retours entre plusieurs systèmes différents pour combiner les informations ce qui est une tâche très fastidieuse. De plus, la quantité d'information disponible dans ces systèmes d'information (réseau de capteurs et corpus documentaire) est à la fois très importante et très hétérogène puisque de multiples structures de données sont utilisées avec différents langages, différents niveaux de détails, différentes versions, etc. Enfin, les éléments de réponses qui fournissent une explication au sujet d'un événement qui est déclenché sont très souvent dispersés entre les différentes sources de données. Il appartient donc à notre utilisateur de filtrer, de combiner et de résumer ces éléments afin de bâtir de lui-même cette explication.

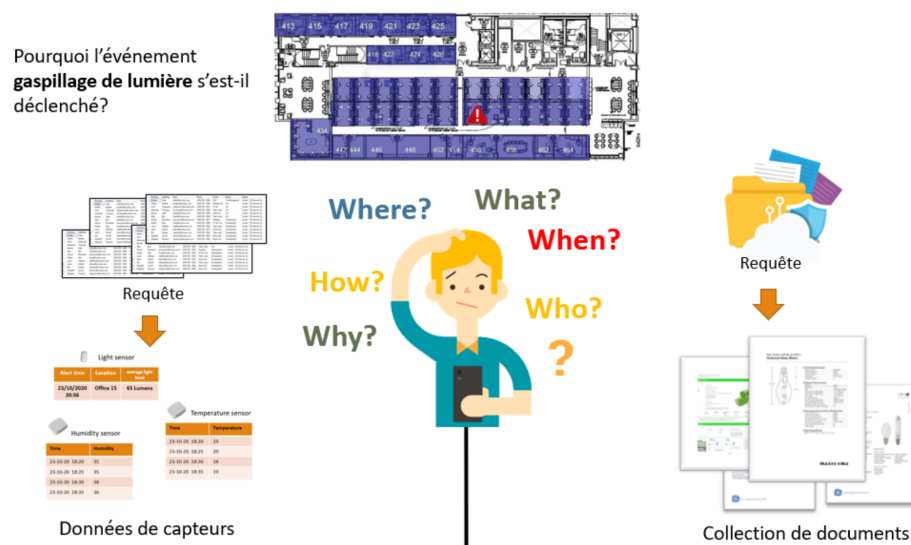


FIGURE 5.2 – Recherche d'explications dans les environnements hybrides

Notre défi ici concerne donc l'intégration ainsi que l'interrogation de ressources de données hétérogènes gérées actuellement par différents systèmes d'information. Pour lever ce verrou, nous avons émis l'hypothèse qu'une représentation sémantique permet de lier de telles données hétérogènes. Nous avons donc exploité le formalisme RDF, comme abordé dans le chapitre 2. En effet, ce choix a été fait car il nous permet de décrire les sources de données associées aux capteurs avec celles issues des documents en une seule et même structure à base de graphes qui permet de lier les différentes informations entre elles par des relations. Ainsi, nous pourrions gérer, analyser et interroger le réseau de données en utilisant toutes les technologies du Web Sémantique.

En complément du défi décrit précédemment, nous devons fournir des pistes d'explication face à un événement déclenché pour faciliter la tâche d'un utilisateur. Pour ce faire, nous exploitons également la sémantique que nous pouvons associer au formalisme RDF, et que nous avons également abordée dans le chapitre 2. En effet, cette sémantique est à la fois utile pour enrichir le graphe de connaissance, par exemple grâce à des concepts ou bien à l'aide de raisonnements effectués par des méthodes d'alignements [46], mais également nécessaire pour parcourir le graphe de description afin d'en extraire des éléments pertinents pour la compréhension d'un utilisateur.

Dans la partie suivante, nous dressons une synthèse de l'état de l'art des travaux concernant la recherche d'information sémantique dans les environnements à base de réseaux de capteurs. Nous montrons également que nous pouvons nous inspirer des systèmes de question-réponse abordés dans la communauté en recherche d'information documentaire pour fournir des explications face à un événement déclenché au sein d'un environnement hybride.

5.2 État de l'art

Contrairement à la recherche d'information documentaire [112], la recherche d'information au sein d'un réseau de capteurs est un domaine beaucoup plus récent. Afin de prendre en compte l'hétérogénéité de ces capteurs avec leurs données produites, de nombreux travaux de recherche ont été menés pour proposer dans ce cadre des approches de recherche d'information sémantique (e.g., [23, 33, 129, 133]). Plusieurs ontologies ont été également mises à disposition pour pousser cette communauté vers un Web Sémantique de capteurs [118], par exemple grâce à la recommandation SSN (*Semantic Sensor Network*) [62] du W3C ou bien à des extensions de cette dernière comme les ontologies MSSN-onto (*Multimedia SSN Ontology*) [20] ou HSSN (*Hybrid Semantic Sensor Networks*) [90].

Plus précisément, la recommandation SSN du W3C est une ontologie générique pour décrire un réseau de capteurs, c'est-à-dire qu'elle n'est pas dépendante d'un domaine spécifique. Elle permet à la fois la modélisation des métadonnées associées aux capteurs mais également la description de leurs observations. Néanmoins, cette recommandation présente deux lacunes principales par rapport à notre défi qui consiste à fournir une explication face à un événement déclenché (§5.1). D'une part, cette on-

tologie ne permet pas de décrire précisément l'environnement où les capteurs sont déployés (e.g., salles, étages, bâtiments) qui sont des informations utiles pour fournir les explications à notre utilisateur. D'autre part, elle ne permet pas de modéliser des événements pouvant être contrôlés et détectés dans l'environnement. L'extension MSSN-onto définit justement la notion d'événement et aborde une plus grande variété de capteurs dits "capteurs multimédia" sans pour autant décrire le contexte environnant associé à ces capteurs. L'ontologie HSSN permet quant à elle de combler toutes ces lacunes en permettant de décrire à la fois des événements, le contexte de déploiement des capteurs, mais aussi la variété de ces derniers et de leurs données produites.

Parallèlement aux travaux présentés précédemment, la recherche d'information s'est intéressée entre autres aux systèmes de question-réponse [94]. Sur la base d'un corpus documentaire, un système de question-réponse consiste à fournir une réponse plus directe et plus précise à une question formulée en langage naturel par un utilisateur. Par exemple, si la question est "*Quelle est la date de naissance de Barack Obama ?*", ce type de système retournera la date du "*4 août 1961*". Contrairement aux moteurs de recherche qui renvoient une liste de documents (éventuellement accompagnés d'extraits) à partir d'une requête saisie sous forme de mots-clés, les systèmes de question-réponse ont pour but d'extraire une réponse précise à une question dans un ensemble de documents.

L'approche 5W1H (*What, Who, When, Where, Why et How*) est exploitée par les systèmes de question-réponse. Dans la littérature, cette approche est utilisée pour extraire des informations structurées sur les événements à partir d'articles de presse en ligne [63, 132]. Dans notre contexte, l'approche 5W1H nous paraît être un excellent choix pour répondre à notre objectif de la représentation des explications face à un événement déclenché. En effet, que ce soit un événement détecté au sein d'un article de presse ou bien un événement produit dans un environnement avec des capteurs, les types de questions que l'on se pose sont souvent du même ordre : pourquoi l'événement a-t-il eu lieu (Why)? Quand est-ce que l'événement a eu lieu (When)? Où est-ce qu'il a eu lieu (Where)? Quels sont les éléments de l'environnement qui ont contribué au déclenchement de cet événement (Who)? et comment (How)? Ainsi, les six questions du 5W1H nous permettent de clarifier les principales raisons liées à un événement déclenché. De plus, nous pensons que cette structuration sous forme de question-réponse en 5W1H facilitera la compréhension d'un utilisateur pour contextualiser les raisons d'un déclenchement.

Les environnements hybrides qui mêlent à la fois des capteurs avec des documents n'ont pas été considérés par la communauté des systèmes cyber-physiques et celle de la recherche d'information. Tenant compte de l'état de l'art et de ce cadre applicatif, la partie suivante démontre que le Web Sémantique est une solution pour rapprocher les contributions de ces deux communautés. En effet, l'ontologie HSSN propre aux capteurs permet d'alimenter un graphe de connaissance avec d'autres ontologies liées aux corpus documentaire alors que l'approche 5W1H est exploitée pour analyser ce graphe afin de fournir à un utilisateur des explications potentielles face à un événement.

5.3 Notre proposition fédérant des données de communautés différentes

Dans le travail de thèse de Nabila Guennouni [55] que j'ai co-dirigé avec Christian Sallaberry, nous avons défini un environnement nommé ISEE (*Information System for Event Explanation*) qui considère deux sources d'informations complémentaires pour fournir une explication sous la forme de réponse 5W1H face à un événement qui se déclenche au sein d'un environnement hybride : une source qui provient d'un réseau de capteurs et une autre qui fait référence à un corpus de documents. À notre connaissance, ISEE est le premier système dédié à l'explication d'événements déclenchés au sein de ce type d'environnement. L'architecture générale de notre proposition est illustrée dans la figure 5.3.

Pour fédérer deux sources de données hétérogènes, nous proposons d'exploiter des ontologies de domaines en nous appuyant sur le formalisme RDF. Cette approche nous permet de lier les données entre elles et de parcourir le graphe de connaissance pour bâtir des explications potentielles face à un événement qui vient d'être détecté. Ainsi, nous supposons que nous disposons dès le départ de plusieurs ontologies de domaine pour modéliser le réseau de capteurs et le corpus de documents. Par exemple, l'ontologie HSSN (cf., §5.2) est utilisée pour décrire les événements ainsi que le réseau de capteurs. Dans le cas d'une entreprise connectée, nous avons exploité l'ontologie EOSK [135] pour identifier à partir de fiches de descriptions des données au sujet d'employés ainsi que l'ontologie SAREF4BLDG³ qui permet à partir d'un inventaire de décrire les éléments d'un bâtiment.

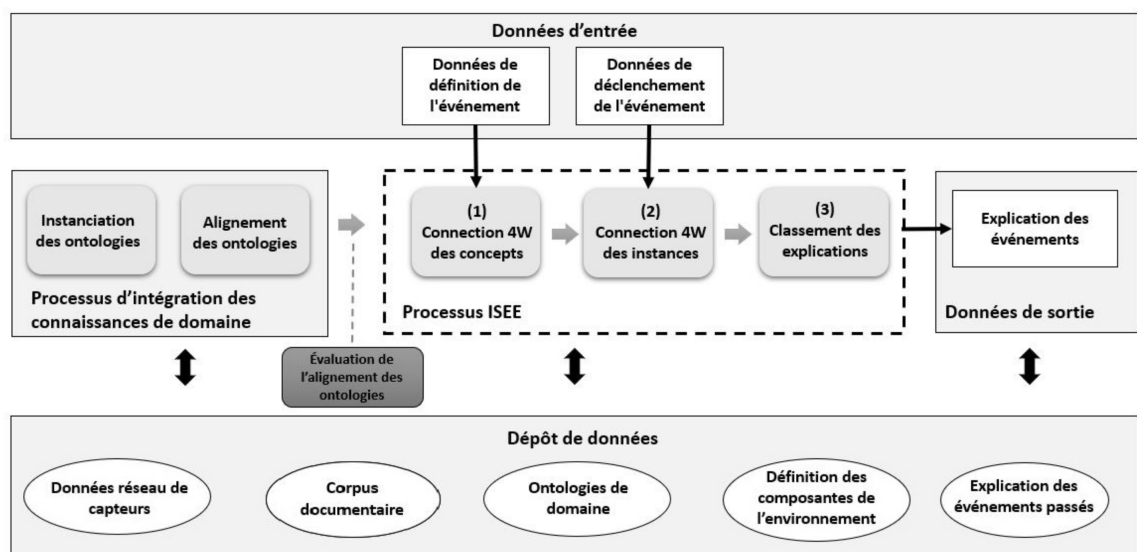


FIGURE 5.3 – Vue d'ensemble de notre approche ISEE

3. <https://saref.etsi.org/saref4bldg/>

Comme il est possible de le constater au centre de la figure 5.3, notre proposition consiste en deux processus principaux : (1) le processus d'intégration des connaissances du domaine qui exploite des outils existants d'instanciation et d'alignement, et (2) le processus ISEE qui selon un événement déclenché analyse le graphe de connaissance selon l'approche 5W1H pour fournir des explications potentielles.

1. **Un processus d'intégration des connaissances de domaine** : ce processus a pour but (i) d'intégrer toutes les informations provenant de l'environnement hybride (i.e., données de capteurs et corpus documentaire) dans les ontologies de domaine sous forme d'instances. Puis, (ii) d'interconnecter ces ontologies en utilisant des techniques existantes d'alignement. Le processus d'intégration des connaissances de domaine est donc composé de deux sous-étapes :

- **Instanciation des ontologies** : des techniques d'instanciation automatique des ontologies sont utilisées pour analyser le corpus de documents et ensuite alimenter les ontologies de domaine avec un ensemble d'instances (e.g., [47]). Par exemple, grâce à une fiche descriptive, on peut identifier une instance d'un employé qui est affecté à un bureau donné avec un rôle spécifique dans son entreprise. Dans la figure 5.4, les instances (*les cercles*) sont ainsi rattachées aux concepts des ontologies (*les rectangles*).
- **Alignement des ontologies** : des techniques d'alignement d'ontologies sont utilisées afin de construire des relations d'alignement entre les concepts des ontologies, telles que des relations d'équivalence, de généralisation et de spécialisation (e.g., [38]). Par exemple dans la figure 5.4, un processus d'alignement peut détecter que le concept *Employee* est lié au concept *Person* par une relation de spécialisation (\subset), tandis que les concepts *Room* et *Office* pourraient être considérés comme similaires (\equiv).

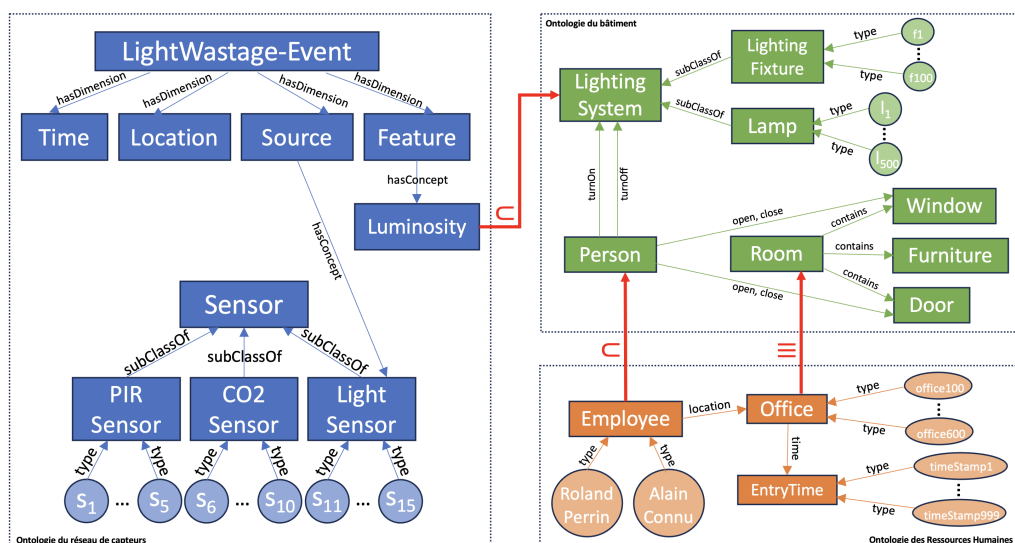


FIGURE 5.4 – Un exemple simplifié illustrant concepts et instances

L'objectif ici consiste donc grâce au formalisme RDF à lier entre elles les différentes informations du système provenant des capteurs et des documents au travers d'ontologies. En effet, initialement les données sont brutes et parfois peu structurées (en particulier, elles sont non-associées à des concepts sémantiques définis dans les ontologies) et les ontologies sont aussi déconnectées entre elles (i.e., il n'y a pas ou peu de liens entre les concepts des ontologies de domaines). Bien entendu, si de nouveaux documents ou de nouveaux capteurs viennent s'ajouter dans le système d'information, l'étape d'instanciation des ontologies doit être réexécutée pour que les nouvelles instances soient intégrées dans les ontologies de domaine. Après l'exécution du processus d'intégration des connaissances de domaine, notre proposition dispose donc d'un graphe de connaissance global pré-connecté (cf., Figure 5.4 grâce aux alignements en rouge) qui va nous permettre de le parcourir afin de construire des explications de déclenchement à travers notre processus ISEE qui va suivre. Toutefois, avant d'entamer cette seconde étape, notre système offre à l'utilisateur la possibilité d'évaluer les liaisons établies entre toutes les ontologies. Cette étape est facultative mais elle permet, par exemple, de savoir si toutes les ontologies mobilisées sont reliées entre elles ou bien s'il existe certaines données isolées.

2. **Notre processus ISEE** : pour illustrer de façon générale les grands principes de ce processus, nous proposons de suivre les étapes de la figure 5.5 ci-dessous en lisant du haut vers le bas.

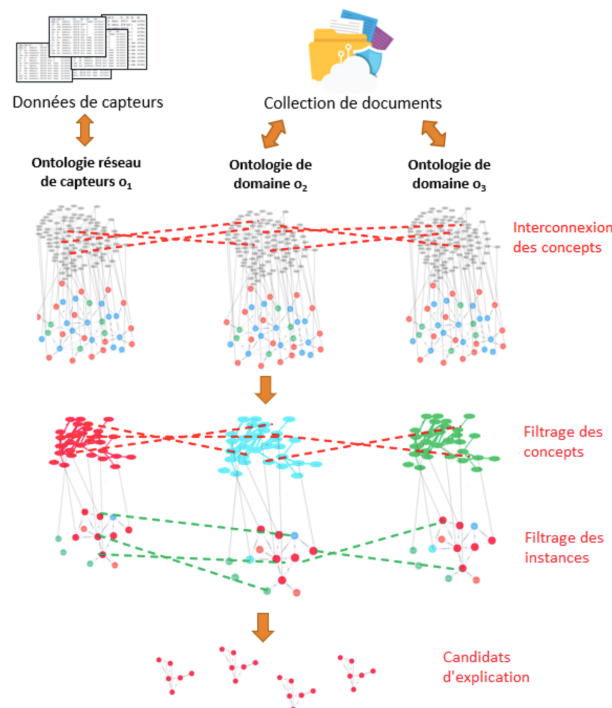


FIGURE 5.5 – Notre proposition ISEE à base de filtrage successifs

Comme nous l'avons vu dans l'étape précédente, ce processus s'appuie donc sur un graphe de connaissance pré-connecté (i.e., un ensemble d'ontologies de domaine instanciées et alignées) à partir de nos jeux de données issus des capteurs et des documents de l'environnement hybride. En haut de la figure 5.5 (Partie *Interconnexion des concepts*), on peut voir justement ce graphe pré-connecté que nous avons également illustré en guise d'exemple dans la figure 5.4 avec les instances des concepts ainsi que leurs alignements symbolisés par des traits rouge. Lorsqu'un événement est défini celui-ci fera intrinsèquement référence à un certain nombre de concepts. Dès lors, avant même son déclenchement, il est possible de lier cet événement à certains concepts du graphe pré-connecté. C'est ce que nous appelons le filtrage des concepts dans la figure 5.5 (ou *Connection 4W des concepts* dans la figure 5.3). En effet, un événement défini comme *Light-Wastage-Event (Gaspillage de lumière)* fait référence au concept de luminosité. Ce concept lié à la définition de l'événement permet d'établir des liens de cause à effet vers d'autres concepts des ontologies de domaine. Par exemple dans la figure 5.4, les lampes et les plafonniers qui sont des systèmes lumineux sont en lien avec ce concept de luminosité. Ceci n'est pas directement le cas pour les concepts tels que *Window, Furniture* et *Door*. Cette définition de l'événement permet donc de se focaliser sur certains concepts du graphe de connaissance tout en écartant d'autres concepts non directement liés à l'événement.

Lorsqu'un événement se déclenche celui-ci fera évidemment référence à des instances particulières. Par exemple, un capteur de lumière détecte vers 20h une luminosité importante dans le bureau 400. Grâce à notre graphe de connaissance filtré au niveau conceptuel, il est possible d'identifier de façon mesurée toutes les instances de l'environnement associées à ces concepts filtrés. Bien entendu, on ne retiendra que les instances en lien avec le déclenchement concerné, c'est-à-dire les instances proches de la fenêtre temporelle ainsi que de la localisation de la détection. C'est ce que nous appelons le filtrage des instances dans la figure 5.5 (ou *Connection 4W des instances* dans la figure 5.3).

À partir de ces instances, la dernière étape de notre processus consistera à fournir des candidats d'explications potentielles sous la forme de réponse en 5W1H (la finalité des figures 5.3 et 5.5). L'utilisateur pourra donc savoir : quel événement a été déclenché (What), quel capteur l'a détecté (Who), les conditions du déclenchement (How), le lieu ainsi que l'horaire du déclenchement (Where et When). Pour les raisons du déclenchement (le Why), nous avons proposé d'y répondre avec une structure en 4W1H (sans un nouveau Why) : quel élément de l'environnement a participé au déclenchement de l'événement (What), qui a pu interagir avec cet élément (Who), où et quand y-a-t il eu une interaction avec l'élément (Where et When) et avons-nous des observations liées à cette interaction (How). Pour ce faire, nous parcourons les instances tout en exploitant la sémantique ainsi que les connections associées à chacune d'entre elles pour répondre à chaque question du 5W1H. Un classement des explications est proposé à l'utilisateur selon que tout ou partie des 5W1H a été complété avec succès.

Les algorithmes liés à chacune des étapes décrites précédemment ont été détaillés dans [55]. De plus, ils ont été publiés dans [56, 57, 58]. Pour conclure et avant de présenter nos expérimentations dans la partie suivante, les avantages principaux de notre approche résident dans le fait qu'elle est :

- **Générique** : du moment où l'on dispose d'ontologies qui modélisent les domaines cibles, le système ISEE peut être utilisé dans tout type d'environnement hybride (hôpitaux connectés, domaines agricoles connectés, commerces connectés...).
- **Flexible** : le système ISEE est basé sur des outils existants (outils d'instanciation et d'alignement). L'utilisateur a donc la possibilité de choisir les outils qui lui conviennent pour indexer ou aligner le graphe de connaissance.
- **Optimisée** : notre approche s'exécute en trois temps : (1) avant la définition d'un événement, puis (2) lorsque l'événement est défini, et enfin (3) lorsque l'événement se déclenche et que l'utilisateur souhaite avoir des pistes d'explications. Par conséquent, les étapes d'interconnexion et de filtrage sont guidées par les données provenant de l'événement. En d'autres termes, nous ne parcourons pas l'ensemble du réseau sémantique, mais plutôt une sous-partie de celui-ci potentiellement en lien avec l'événement. Cela permet de limiter l'analyse du graphe de connaissance uniquement sur certaines données et non toutes les informations acquises.

5.4 Expérimentations

Nous avons développé un prototype intitulé ISEEapp. Une vidéo de présentation de notre outil est disponible en ligne à l'adresse suivante : <https://slaborie.perso.univ-pau.fr/images/site/ISEEapp.mp4>. Pour tester notre prototype, nous avons utilisé entre autres le jeu de données Keti⁴. Ce jeu de données est issu du bâtiment de recherche *Sutardja Dai Hall* de l'Université de Californie. Les données de capteurs proviennent de 51 pièces réparties sur 4 étages. Chaque pièce comprend 5 types de mesures : la concentration de CO₂, le niveau d'humidité, la température ambiante, la luminosité et les données du capteur de mouvement PIR (*Passive InfraRed sensor*). Ces données sont recueillies sur une période d'une semaine, du vendredi 23 août 2013 au samedi 31 août 2013. Le capteur de mouvement PIR est échantillonné une fois toutes les 10 secondes tandis que les autres capteurs sont échantillonnés une fois toutes les 5 secondes. Le tableau 5.1 résume le nombre total d'observations collectées par les 5 types de capteurs ce qui permet de se rendre compte du volume de données acquises sur une semaine pour quelques capteurs.

Le corpus de documents quant à lui est composé de 150 fiches techniques reliées aux différents équipements installés dans le bâtiment de recherche (système d'éclairage, système de ventilation, ordinateurs, projecteurs, capteurs, etc.) et de 130 fiches d'information au sujet des employés pour les différents types de poste.

4. <https://www.kaggle.com/ranakrc/smart-building-system>

Type du capteur	Nombre total d'observations
Concentration de CO2	6 574 059
Niveau d'humidité	6 571 516
Température ambiante	6 571 556
Luminosité	6 571 514
Mouvement PIR	3 594 004
TOTAL	29 882 649

TABLE 5.1 – Observations collectées par les 5 types de capteurs

Comme dans notre exemple de la figure 5.4, un événement intitulé « *Gaspillage de lumière* » est déclaré dans notre système. Il se déclenche lorsque les quatre conditions suivantes sont simultanément vérifiées : (i) un niveau de luminosité dépassant les 50 lux, (ii) absence de mouvement, (iii) les deux conditions (i) et (ii) doivent être vérifiées entre 19h et 7h, enfin (iv) les trois conditions (i), (ii) et (iii) doivent être satisfaites pendant une durée qui dépasse les 10 minutes.

Une fois l'intégration de toutes les données au sein de notre prototype en suivant les étapes décrites dans la section 5.3, ISEEapp vérifie si un événement se déclenche et il fournit en conséquence à l'utilisateur des candidats potentiels d'explication comme illustré dans la figure 5.6. Dans cette réponse en 5W1H, il y a eu un gaspillage de lumière (What) qui a été détecté par un capteur de lumière *LightSensor15* (Who) dans le bureau *Office 400* (Where) à *20:00:14* (When) avec des valeurs observées au-delà de 60 lux pendant plus de 10 minutes (How). Une raison potentielle de ce déclenchement (le Why sur la droite de la figure 5.6) est proposé avec une structure en 4W1H (cf., §5.3) qui indique que deux lampes (*Lamp457* et *Lamp458*) localisées dans le même bureau sont très probablement restées allumées par l'employé *Roland Perrin* qui est affecté à ce bureau. Pour compléter cette explication en 5W1H, nous nous appuyons sur le graphe de connaissance tel que celui illustré en exemple dans la figure 5.4. Bien entendu, d'autres explications potentielles sont présentées sous la même forme à l'utilisateur et il lui appartient d'en vérifier la conformité.

Pour évaluer la clarté de l'affichage des réponses à l'utilisateur, nous avons conduit une expérimentation sur l'usage de notre outil ISEEapp au sein de deux classes d'étudiants en IUT : une classe de 28 élèves en informatique et une autre classe de 34 élèves en génie industriel et maintenance (GIM). L'idée de cette expérimentation consiste à évaluer la facilité de compréhension des explications en 5W1H retournées par notre prototype puisqu'il s'agit d'une originalité de présentation de nos résultats. Pour ce faire, après une rapide présentation de notre application aux étudiants, nous avons créé un questionnaire d'évaluation qui se concentre sur la structure en 5W1H des explications potentielles de déclenchement d'un événement.

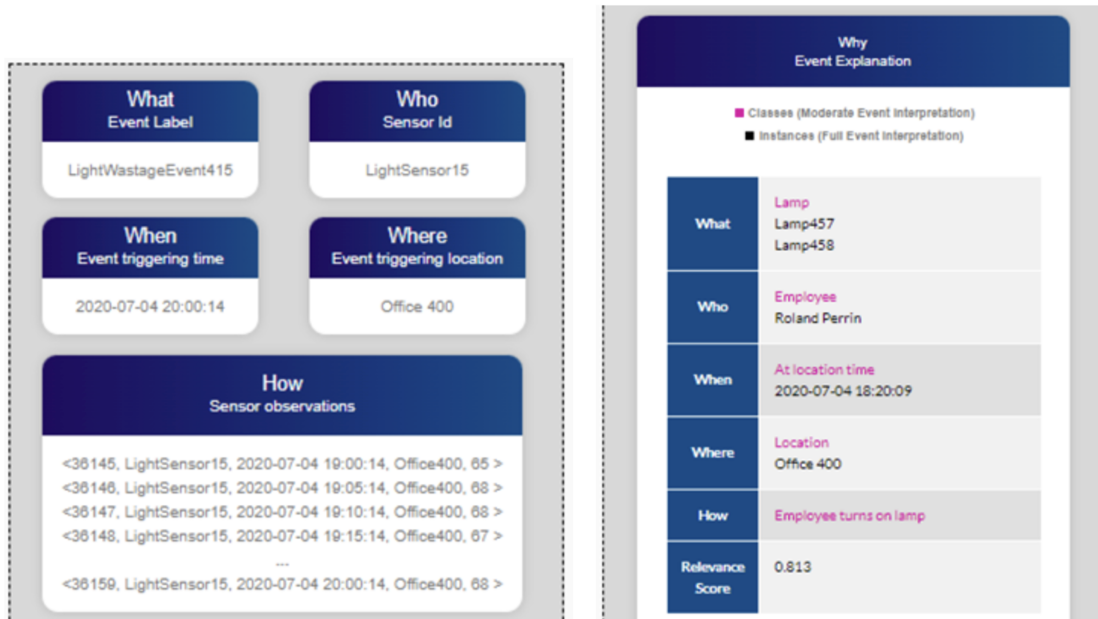


FIGURE 5.6 – ISEEapp : une explication d'un événement déclenché

La figure 5.7 présente les résultats de notre évaluation auprès des étudiants en IUT (GIM en bleu, Informatique en orange et la moyenne totale en gris) avec l'exemple de notre événement « *Gaspillage de lumière* ».

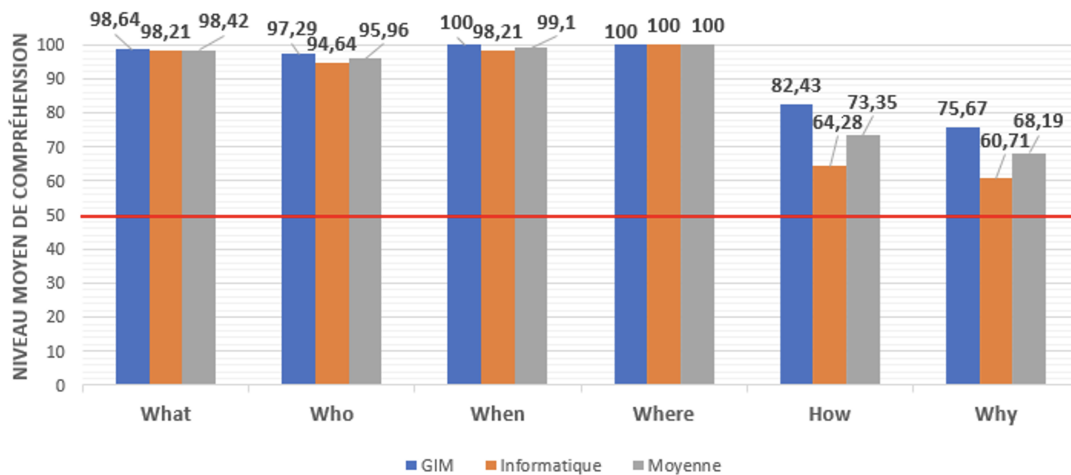


FIGURE 5.7 – Résultats de l'expérimentation d'ISEEapp en IUT

Comme il est possible de le constater, la grande majorité des étudiants comprennent les éléments retournés par ISEEapp qui concernent les réponses aux questions de type quoi (What), qui (Who), où (Where) et quand (When). Le contexte du déclenchement d'un événement est donc parfaitement maîtrisé. En ce qui concerne l'élément comment (how), il y a plus de difficultés de compréhension pour les informaticiens car les

étudiants en GIM sont plus habitués aux valeurs des capteurs observées issues d'un bâtiment. Pour la structuration de l'élément pourquoi (Why), plus de 50% des utilisateurs comprennent les retours de notre outil. Néanmoins, il convient d'améliorer notre interface graphique afin d'augmenter ce résultat, par exemple en expliquant d'avantage ce que signifie le What/Who/When/Where/How au sein du Why ou bien en permettant une navigation au sein de la partie concernée du graphe de connaissance.

5.5 Synthèse du chapitre

Dans ce chapitre, j'ai montré comment nous avons su exploiter le Web Sémantique au sein d'un environnement hybride. Tout d'abord dans la section 5.1, j'ai présenté le défi d'intégration ainsi que d'interrogation de ressources de données hétérogènes gérées actuellement par différents systèmes d'information. Le cas de la recherche d'explications d'un événement déclenché au sein d'un environnement mêlant des données de capteurs avec un corpus documentaire a été illustré.

Dans la section 5.2, nous montrons que chaque communauté a proposé des outils et qu'il est important de les exploiter. En particulier, les outils en lien avec les technologies du Web sémantique vont nous permettre de fédérer toutes nos données au sein d'un graphe de connaissance, comme par exemple des ontologies, des indexeurs sémantiques ou bien des outils d'alignements.

Ensuite, dans la section 5.3, notre approche ISEE a permis d'intégrer et d'analyser toutes ces données au sein d'un graphe RDF en fonction du cycle de vie d'un événement. Tout d'abord, avant la définition de l'événement, les données sources sont instanciées en les affectant aux concepts des ontologies de domaine. Les concepts de ces ontologies sont également alignés pour établir des connections entre toutes les ontologies mobilisées. Lors de la définition d'un événement, certains concepts en lien avec cet événement sont privilégiés. Et enfin, lors du déclenchement de l'événement, les instances associées à ces concepts filtrés sont sélectionnées pour fournir des pistes d'explications potentielles sous la forme de réponse en 5W1H.

Enfin, dans la section 5.4, nous avons effectué des expérimentations entre des données issues des systèmes cyber-physiques et de la recherche d'information. Du côté de la communauté des systèmes cyber-physiques, nous avons exploité des données réelles issues des capteurs ainsi que des extensions de l'ontologie SSN du W3C. De l'autre, nous avons exploité des ontologies de domaine pour indexer un corpus documentaire hétérogène. ISEEapp est le premier prototype permettant de faire le pont entre ces deux communautés scientifiques grâce aux technologies issues du Web Sémantique.

Ce travail récent a été publié en 2021 dans une revue de rang Q1 au regard du classement Scimago :

- Nabila Guennouni, Christian Sallaberry, **Sébastien Laborie**, Richard Chbeir, Elio Mansour. *ISEE : A heterogeneous information system for event explainability in smart connected environments*, Internet of Things, 2021.

Notre prototype ISEEapp a également été publié en 2022 dans une conférence de rang B au regard du classement CORE :

- Nabila Guennouni, **Sébastien Laborie**, Christian Sallaberry, Richard Chbeir, Elio Mansour. *ISEEapp : An Event Explanation Prototype bridging the gap between sensor network and document corpora data*, 18th International Conference on Intelligent Environments (IE 2022), June 2022.

Chapitre 6

Conclusion générale

Fort de cette expérience acquise depuis plus de 10 ans, je souhaite toujours prolonger mes activités de recherche en lien avec le Web Sémantique. Ce dernier a été mon fil conducteur durant toutes ces années comme il est possible de le lire au travers des chapitres de ce manuscrit. Dans cette conclusion, je vais tout d'abord effectuer une synthèse de mes activités de recherche en mettant en avant des grands principes auxquels j'ai souscrit (Section 6.1). En perspective, je montrerai que le Web Sémantique a encore toute sa place dans les nouveaux enjeux liés à la gestion des données que ce soit vis-à-vis des données massives via par exemple les lacs de données (*Data Lakes*) ou bien que ce soit dans un cadre lié à l'éco-responsabilité (Section 6.2).

6.1 Synthèse de mes activités de recherche

J'ai souhaité commencer ce manuscrit par un chapitre dédié à la présentation des grands principes du Web Sémantique (Chapitre 2). Décrire, raisonner et interroger sont des éléments fondamentaux lorsque l'on découvre pour la première fois les technologies issues de cette communauté. Évidemment, ce chapitre permet au lecteur de mieux comprendre tous les autres chapitres de ce manuscrit. Au-delà de cet exercice pédagogique, mon souhait a aussi été de permettre aux personnes désireuses d'appréhender le Web Sémantique de pouvoir en quelques pages avoir une vision globale de cet univers (sans forcément lire la suite des chapitres). J'ai donc pu diffuser ce chapitre auprès de jeunes chercheurs et notamment les doctorants que j'encadre aujourd'hui. Ce chapitre leur a permis de les aider dans la compréhension de leur état de l'art lorsque des travaux exploitent les technologies issues du Web Sémantique. Il m'a aussi permis d'échanger plus rapidement avec eux sur l'opportunité ou non de choisir une technologie en particulier (e.g., a-t-on besoin d'une ontologie légère en RDFS ou plus complète en OWL?).

Le chapitre 3 se concentre essentiellement sur les descriptions à l'aide du formalisme RDF. Que ce soit en terme de structuration ou bien en terme de sérialisation avec un langage tel que RDF/XML, une même description RDF peut se spécifier différemment d'un utilisateur à un autre. Ceci peut donc avoir un impact négatif sur des

applications dédiées notamment au stockage, au chargement et au traitement de descriptions (e.g., calcul de similarité, parcours de graphe). Nous avons ainsi proposé de normaliser les descriptions RDF selon un contexte applicatif. Ces travaux ont permis de démontrer qu'il est nécessaire d'avoir une certaine prise de recul sur les technologies issues du Web Sémantique, quand bien même il s'agit d'une recommandation du W3C.

Dans le chapitre 4, j'exploite les technologies du Web Sémantique au service de l'adaptation de contenus multimédias. J'ai souhaité continuer les travaux de recherche que j'avais initiés dans le cadre de mon doctorat. Nous avons montré que la communauté du Multimédia a eu un réel besoin de modélisation de profils et de services d'adaptation. La communauté du Web Sémantique a su répondre à ce besoin avec notamment la recommandation CC/PP du W3C mais d'une façon limitée. Tenant compte de cette situation avec en outre ma prise de recul énoncée dans le chapitre précédent, cela m'a permis de proposer une nouvelle modélisation sémantique des profils à base de services qui permet d'inspirer ces deux communautés (i.e., Multimédia et Web Sémantique). Ce principe que j'appelle gagnant-gagnant a permis de développer pendant plusieurs années des collaborations que ce soit pour l'exploitation du contexte d'un utilisateur ou bien pour modéliser des applications interactives.

Enfin, le chapitre 5 montre que des données hétérogènes exploitées par différentes communautés peuvent cohabiter au sein d'un même système d'information grâce au Web Sémantique. Nous avons ainsi contribué à un domaine de recherche récent comme la Recherche d'Information au sein d'un réseau de capteurs. Notre proposition a permis de démontrer qu'il est possible de fédérer des données issues de capteurs avec un corpus documentaire grâce à un graphe de connaissance. Dans le même temps, nous avons exploité l'approche 5W1H (utilisée en Recherche d'Information) pour parcourir ce graphe afin de fournir des explications lorsqu'un événement se déclenche.

Pédagogie (Chapitre 2), prise de recul (Chapitre 3), principe gagnant-gagnant (Chapitre 4) et fédération de données issues de différentes communautés (Chapitre 5) sont des éléments qui m'ont tenu à cœur durant toutes ces années de recherche scientifique. Dans la partie suivante, je vais montrer que je souhaite continuer sur cette ligne directrice tout en participant à de nouveaux enjeux, comme par exemple la gestion de masses de données et l'éco-responsabilité.

6.2 Perspectives sur mon activité de recherche future

Le Web sémantique a encore toute sa place dans les enjeux à venir. Je souhaite investir en particulier les deux problématiques suivantes : (1) Comment le Web Sémantique peut contribuer à la gestion ainsi qu'à l'analyse des données de plus en plus massives et hétérogènes? (§6.2.1) et (2) Comment le Web Sémantique s'inscrit au sein de l'informatique éco-responsable? (§6.2.2). Bien entendu, ces deux problématiques sont liées l'une avec l'autre.

6.2.1 Gérer des masses de données

Pour répondre aux nouveaux enjeux liés à la gestion des données massives (*Big Data*), de nombreuses entreprises et collectivités cherchent à mettre en valeur leurs données. Pour les organisations qui sont pilotées ou bien axées sur la donnée, on parle d'organisations dites "*Data-Driven*"¹. Pour valoriser leurs données, ces organisations améliorent continuellement l'ensemble de leurs processus par l'analyse quantitative et qualitative des données en tout temps, en tous lieux et par tous. Ainsi, le management des données devient un enjeu stratégique². L'objectif n'est pas simplement de fournir une solution technologique permettant d'intégrer, de stocker et d'interroger des données, mais bien de les valoriser comme un capital stratégique. Le management de données consiste donc à gérer le patrimoine informationnel, à le compléter par des sources externes hétérogènes, mais aussi à intégrer les transformations de matières brutes en une matière raffinée pour diverses analyses. Il est donc nécessaire de déployer une solution complète de management du cycle de vie de la donnée de sa production jusqu'à sa valorisation. Aujourd'hui, les lacs de données (*Data Lakes*) constituent une solution ambitieuse et prometteuse de management de données massives.

Une approche lac de données permet (1) d'ingérer des données brutes et hétérogènes, (2) de les transformer afin de (3) les rendre disponibles pour différentes analyses [114]. Parallèlement à ces trois étapes, une caractéristique des lacs de données est la gestion des métadonnées permettant de se focaliser sur des éléments d'information particuliers issus de ces données [102, 114]. Les travaux à l'heure actuelle se sont surtout centrés sur la proposition d'architectures multi-sources des lacs de données [102] et sur la définition de méta-modèles spécifiques à différents métiers.

En septembre 2022, une Chaire de recherche industrielle universitaire intitulée « *Digital & Cadre de vie bien vivre et bien vieillir dans son habitat* » a été mise en place avec la technopôle Domolandes par le LIUPPA et l'IRIT. Cette thématique du bien vivre et du bien vieillir sur un territoire s'inscrit complètement dans ce cadre du management et des lacs de données. Pour répondre à ce défi, nous devons prendre en considération une grande quantité d'informations diverses dont les sources peuvent provenir de données ouvertes (*Open data*), de données issues de collectivités (e.g., cartographie, démographie, enquêtes), de données sur les services offerts à la population (alimentation, transport, santé, éducation, loisirs...), de données provenant des réseaux sociaux, etc. Au travers d'un lac de données, l'exploitation des informations doit permettre de réaliser des analyses pertinentes selon différents niveaux de granularité (i.e., individuel, collectif, local, territorial) tenant compte des évolutions dans le temps, dans l'espace et des besoins. Ce patrimoine informationnel mis en valeur, par exemple, grâce à des tableaux de bord doit permettre aux utilisateurs de prendre des décisions. Il s'agit, par exemple, d'aider à la prise de décision dans des domaines tels que l'aménagement urbain, le développement d'activités et d'actions de sensibilisation...

1. <https://www.ipag.edu/blog/domaines-professionnels/marketing-digital/en-quoi-consiste-la-demarche-data-driven>

2. <https://www.lebigdata.fr/data-management>

Récemment en 2023, j'ai rejoint cette collaboration au travers d'un projet Région intitulé « *Bien Vivre & Bien Vieillir sur son territoire* » afin d'apporter en particulier mon expertise sur l'usage du Web Sémantique. En effet, au travers d'une thèse que je co-dirige avec Philippe Roose (LIUPPA) et Franck Ravat (IRIT), nous envisageons d'exploiter le concept des trajectoires de vie [115] pour modéliser et analyser la dynamique d'un territoire dans le temps et selon différents points de vue. Les outils du Web sémantique ont déjà été exploités pour modéliser des trajectoires dites "sémantiques" (e.g., [96, 97, 103]). Néanmoins, les travaux existants n'intègrent pas, en amont, des données massives et hétérogènes, telles que celles disponibles au sein de lacs de données. Enfin, en aval, ces travaux n'offrent pas non plus d'outil de visualisation pour la prise de décision.

Le Web Sémantique peut-il donc établir ce pont entre les données massives et hétérogènes au sujet du bien vivre et du bien vieillir d'une part, et leurs analyses à des fins de prise de décision pour les acteurs d'un territoire d'autre part (Figure 6.1)? Dans cette question, je retrouve totalement la ligne directrice exposée dans les chapitres de ce manuscrit et dans la section 6.1, c'est-à-dire la prise de recul, le principe gagnant-gagnant et le fait de pouvoir fédérer des données de différentes communautés.

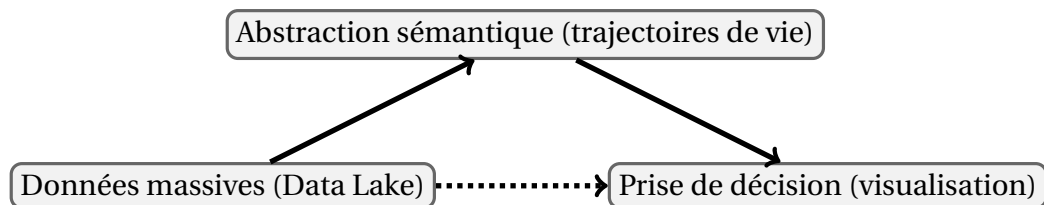


FIGURE 6.1 – Le Web Sémantique : un pont entre données massives et décisions

6.2.2 Prendre en compte l'éco-responsabilité

Si les solutions de gestion des données existantes répondent à des enjeux économiques, ou encore de développement stratégiques et tactiques des organisations et entreprises, elles n'ont toujours pas intégré le facteur aujourd'hui crucial de l'éco-responsabilité. Cette éco-responsabilité peut être vue sous l'angle de l'impact écologique. Sur ce point, beaucoup d'efforts ont déjà été faits sur les problématiques de stockage des données massives en termes d'infrastructures matérielles (e.g., [101]). Cependant, la maîtrise du coût énergétique engendré durant le cycle de vie des données depuis son ingestion jusqu'à son analyse reste encore entière. La figure 6.2 illustre une projection de la demande mondiale d'électricité des centres de données³. Comme il est possible de le constater, quelles que soient les estimations plus ou moins optimistes, la consommation mondiale d'énergie pour les centres de données augmente dans le temps.

3. <https://www.connaissancedesenergies.org/sites/default/files/pdf-actualites/FINALSTUDYEnglishKK-03-20-210-EN-N13072020pdf.pdf>

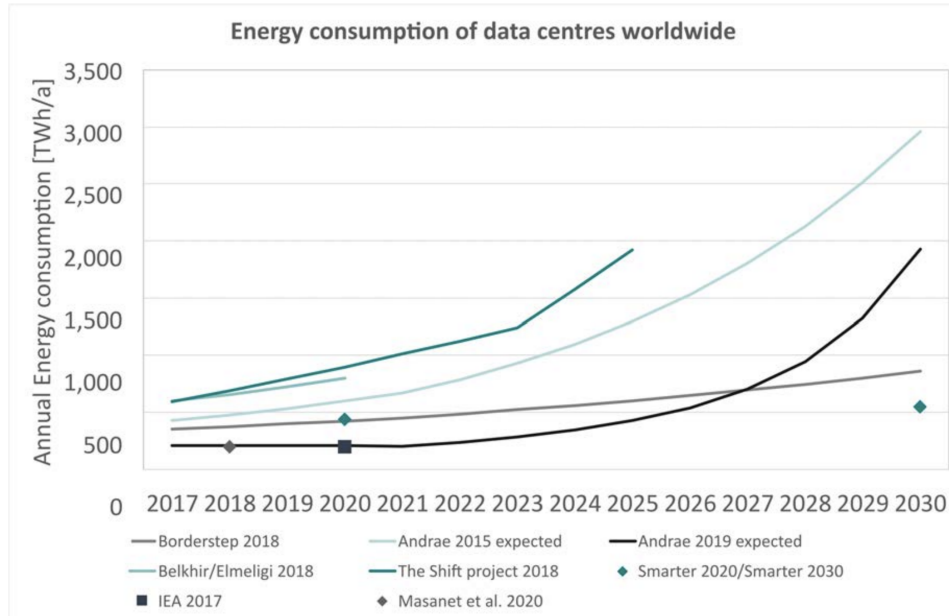


FIGURE 6.2 – Estimation de la demande mondiale en électricité pour les centres de données

À l'heure actuelle, l'efficacité de la gestion de données se manifeste par une prise de décision rapide sur les données pertinentes et de qualité. Ainsi, aux critères de rapidité et de qualité associés à l'efficacité, il est crucial d'y adjoindre le critère de l'écoresponsabilité.

Face à ce constat, l'UPPA au travers de son consortium E2S (*Solutions pour l'Énergie et l'Environnement*⁴) est engagée depuis 2017 pour accompagner la transition énergétique et environnementale des entreprises et collectivités du territoire. Dès le lancement de ce programme au sein de mon Université, je me suis inscrit dans cette initiative en optant pour un financement qui a notamment permis de conduire certains travaux dont celui que j'ai décrit dans le chapitre 5 de ce manuscrit. Plus récemment, j'ai co-encadré un projet avec des étudiantes en Master Technologies de l'Internet pour développer un outil web de visualisation de données énergétiques qui offre de nouveaux indicateurs inspirés du secteur de la Finance pour présenter une consommation d'énergie. Comme j'ai pu l'indiquer dans le chapitre 4, je participe également à des travaux de thèse qui définissent une série de propriétés pour qualifier une consommation d'énergie au sujet d'une application et ses services [76, 77, 78, 79, 80, 104]. Dans ces publications, la partie sémantique n'a pas encore été abordée puisqu'il s'agira d'une contribution future.

Quels que soient les travaux que nous avons menés, nous nous rendons compte que l'une des raisons du gaspillage énergétique lié au numérique est en partie dû à la non-connaissance de la consommation d'énergie par les usagers, qu'ils soient utilis-

4. <https://e2s-uppa.eu>

teurs novices, avancés ou experts. Le caractère virtuel n'encourage bien évidemment pas à cette prise de conscience. En nous inspirant du secteur de l'électroménager avec l'étiquette-énergie ou bien celui de l'agro-alimentaire via l'eco-score⁵, nous souhaitons de la même manière proposer une labellisation écologique au niveau de la gestion et de l'analyse de données. Bien entendu, nous nous intéressons à la chaîne complète du traitement des données : ingestion, stockage, accès, traitement, analyse, archivage et partage. Cet impact écologique de l'intégralité du cycle de vie de la donnée n'a jamais été abordé. En effet, la frugalité des analyses produites pour informer ou bien orienter les concepteurs comme les utilisateurs est encore un domaine à défricher.

Le Web Sémantique a donc toute sa place dans cette prise de conscience d'écoresponsabilité. Non seulement il pourra contribuer à décrire, raisonner et interroger des descriptions sémantiques au sujet de la consommation énergétique par exemple. Mais il pourra également appliquer ces principes sur ses propres descriptions pour un Web Sémantique plus "vert". Cela concerne à la fois l'ingestion des descriptions, leur stockage et leur requête. En effet, faut-il intégrer et interroger des descriptions de données à tout prix?

5. <https://docs.score-environnemental.com>

Chapitre 7

Mon parcours : académique & scientifique

Après un Baccalauréat Scientifique obtenu en 1999 à Aubenas (07 – Ardèche), j’ai souhaité poursuivre mes études supérieures dans le domaine de l’Informatique. Ainsi, à l’Université Joseph Fourier de Grenoble (38 – Isère), j’ai successivement obtenu un DEUG Sciences et Technologies mention MIAS (*Mathématiques et Informatique Appliquées aux Sciences*), une Licence ainsi qu’une Maîtrise d’Informatique, puis un Master Recherche Mathématiques et Informatique spécialité Intelligence, Interaction et Information. En mai 2008, j’ai ensuite obtenu un doctorat en Informatique à l’Université Joseph Fourier de Grenoble. Après deux années en qualité de post-doctorant à l’Université Paul Sabatier de Toulouse (31 – Midi-Pyrénées), j’ai été recruté en 2010 en tant que Maître de Conférences à l’Université de Pau et des Pays de l’Adour (64 – Pyrénées-Atlantiques) affecté à l’IUT de Bayonne Pays Basque ainsi qu’au Laboratoire d’Informatique de l’Université de Pau et des Pays de l’Adour (LIUPPA).

Depuis des années, je m’efforce de mener de front des activités d’enseignement, de recherche et d’administration propres à tout enseignant-chercheur. La figure 7.1 qui suit donne une vue d’ensemble chronologique de mes activités dans ces 3 domaines et la manière dont j’ai essayé de les mener en parallèle. Les sections qui suivent décrivent en détail les éléments qui y sont présentés : investissement pédagogique (Section 7.1), activité scientifique (Section 7.2) et responsabilités collectives (Section 7.3). Des ressources complémentaires sont également disponibles sur mon site web professionnel à l’adresse suivante : <http://slaborie.perso.univ-pau.fr>.

Pour information, en 2022 j’ai obtenu par la voie nationale un avancement de grade de la classe normale vers la hors classe des maîtres de conférences. Également en 2022, j’ai obtenu la Prime individuelle (C3) du Régime Indemnitaire des Personnels Enseignants et Chercheurs (RIPEC) sur l’ensemble de mes activités d’enseignant-chercheur, c’est-à-dire enseignement, recherche et administration.

CHAPITRE 7. MON PARCOURS : ACADÉMIQUE & SCIENTIFIQUE

Statut	Sept. 2010 Août 2011	Sept. 2011 Août 2012	Sept. 2012 Août 2013	Sept. 2013 Août 2014	Sept. 2014 Août 2015	Sept. 2015 Août 2016	Sept. 2016 Août 2017	Sept. 2017 Août 2018	Sept. 2018 Août 2019	Sept. 2019 Août 2020	Sept. 2020 Août 2021	Sept. 2021 Août 2022	Sept. 2022 Août 2023	Sept. 2023 Août 2024	
Etablissement et laboratoire de rattachement	MCF	MCF	MCF	MCF	MCF	MCF	MCF	MCF	MCF	MCF	MCF	MCF	MCF	MCF	
	IUT de Bayonne Pays Basque LUPPA (Bayonne)	IUT de Bayonne Pays Basque LUPPA (Bayonne)	IUT de Bayonne Pays Basque LUPPA (Bayonne)	IUT de Bayonne Pays Basque LUPPA (Bayonne)	IUT de Bayonne Pays Basque LUPPA (Bayonne)	IUT de Bayonne Pays Basque LUPPA (Bayonne)	IUT de Bayonne Pays Basque LUPPA (Bayonne)	IUT de Bayonne Pays Basque LUPPA (Bayonne)	IUT de Bayonne Pays Basque LUPPA (Bayonne)	IUT de Bayonne Pays Basque LUPPA (Bayonne)	IUT de Bayonne Pays Basque LUPPA (Bayonne)	IUT de Bayonne Pays Basque LUPPA (Bayonne)	IUT de Bayonne Pays Basque LUPPA (Bayonne)	IUT de Bayonne Pays Basque LUPPA (Bayonne)	
ADMINISTRATION															
ENSEIGNEMENT	274h	369h	473h	414h	468h	458h	351h	301h	242h	242h	242h	211h	245h	300h	
	- DUT GEA - LP Num - DU TIC - ESTI (Ing.) - M2 IT (Toulouse)	- DUT GEA - DUT Info - LP Num - DU TIC - M1 SIGUS - M2 IT (Toulouse)	- DUT GEA - DUT TIC - DUT Info - LP Banque - LP Num - DU TIC - M1 SIGUS - M2 IT (Toulouse)	- DUT GEA - DUT TIC - DUT Info - LP Banque - LP Num - DU TIC - M1 SIGUS	- DUT GEA - DUT Info - LP Banque - LP Num - DU TIC - M1 SIGUS	- DUT GEA - DUT Info - LP Banque - LP Num - DU TIC - M1 SIGUS	- DUT GEA - DUT Info - LP Banque - LP Num - DU TIC - M1 SIGUS	- DUT GEA - DUT Info - LP Banque - LP Num - DU TIC - M1 SIGUS	- DUT GEA	- DUT GEA	- DUT GEA	- DUT GEA	- BUT GEA - M1 TI	- BUT GEA - BUT TIC	
RECHERCHE															

FIGURE 7.1 – Mon parcours d’enseignant-chercheur à l’UPPA.

7.1 Investissement pédagogique

Mes activités d'enseignement occupent une place centrale dans mon quotidien. Comme vous pouvez le constater dans la figure 7.1, j'ai pu intervenir pour différents niveaux de Licence (L) à Master (M) et ces enseignements m'ont permis d'intervenir auprès d'étudiants en formation initiale, en formation continue et en alternance.

Dès mon recrutement en 2010 au sein de l'IUT de Bayonne Pays Basque, j'ai été affecté au Département Gestion des Entreprises et des Administrations (GEA). Mes enseignements (Figure 7.1) sont par conséquent majoritairement dispensés au sein de ce département que ce soit en DUT/BUT et en Licence Professionnelle. Néanmoins, j'ai souhaité intervenir dans d'autres départements de l'IUT tels que le Département Informatique et le Département Techniques de Commercialisation (TC) afin de diversifier les publics visés mais aussi de collaborer avec différentes équipes pédagogiques. Je réalise chaque année des heures d'enseignement au-delà de mon service statutaire de 192H eq. TD. Mes thématiques d'enseignement sont centrées dans le domaine de l'Informatique. J'interviens principalement auprès de futurs gestionnaires et commerciaux désireux d'exploiter l'outil informatique pour la gestion d'entreprise. J'y côtoie divers types d'étudiants de jeunes bacheliers aux adultes en reconversion professionnelle qui ne sont pas toujours familiers des nouvelles technologies et qui nécessitent un accompagnement individualisé.

Pour chaque intervention, je me suis toujours pleinement investi en créant de nouveaux supports ou de nouveaux modules intégralement. J'ai pu dispenser des cours magistraux, des travaux dirigés, des travaux pratiques, du suivi de projets, de stages et d'étudiants en alternance. Je suis également responsable actuellement de plusieurs modules de formation en BUT GEA et TC. De plus, j'exploite systématiquement les plates-formes pédagogiques universitaires tout en offrant aux étudiants des pratiques ainsi que des usages d'outils pédagogiques innovants, comme par exemple ma chaîne Youtube à l'adresse suivante : <https://youtube.com/@SebastienLaborie-GEA>. Cette chaîne a été alimentée en 2020 suite à la période du COVID-19 et elle compte aujourd'hui plus de 13000 vues avec certaines vidéos comptabilisant plus de 2000 vues.

Durant plusieurs années, j'ai aussi eu l'opportunité d'enseigner au niveau Master sur plusieurs formations et sur des matières en relation avec mes activités de recherche que nous décrivons dans la partie suivante (§7.2). Par exemple, de 2008 à 2013, j'ai créé intégralement un module pour le Master IT (*Informatique et Télécommunications*) à l'Université Paul Sabatier de Toulouse qui s'intitulait "*Créer et interroger des descriptions RDF*". Je présentais aux étudiants comment créer des descriptions RDF avec ses diverses sérialisations, comment interroger ces descriptions avec SPARQL et comment intégrer ces descriptions au sein de pages web à l'aide du langage RDFa. Également, de 2012 à 2016, je suis intervenu au sein du Master SIGLIS (*Systèmes Informatiques pour le Génie de la Logistique Industrielle et des Services*) à l'UFR Sciences Côte Basque d'Anglet dans un module intitulé "*Intégration des données dans les Systèmes d'Information*". J'ai créé des supports qui présentent les technologies autour d'XML (essentiellement DTD, XMLS, XPath et XSLT) avec des cas pratiques.

7.2 Activité scientifique

En septembre 2010, j'ai été recruté en tant que Maître de Conférences en Informatique à l'Université de Pau et des Pays de l'Adour (UPPA). Mes travaux de recherche (Figure 7.1) sont menés actuellement au sein du Laboratoire d'Informatique de l'Université de Pau et des Pays de l'Adour (LIUPPA) dans l'équipe T2I (*Traitement des Informations pour l'adaptation de l'Interaction au contexte et à l'utilisateur*)¹. Cette équipe est composée de 13 enseignants-chercheurs permanents (3 PR, 1 MCF HDR et 9 MCF). Les thématiques de recherche auxquelles je contribue dans l'équipe concernent principalement (1) la recherche d'information (*des modèles issus de l'indexation jusqu'aux requêtes*) ainsi que (2) l'adaptation au contexte (*de la modélisation des profils jusqu'à la mise en œuvre opérationnelle*).

Pour la première partie (1- la recherche d'information), l'activité porte sur des informations contenues dans des documents ainsi que des flux de données hétérogènes à des fins de représentation, d'indexation, de recherche et d'exploitation : modélisation des informations ciblées et constitution de ressources nécessaires à des traitements sémantiques, analyse (*segmentation, traitement sémantique, annotation, indexation*) et exploitation (*filtrage, fouille, recherche d'information, aide à la décision, présentation des résultats*).

En ce qui concerne la seconde partie (2 - l'adaptation au contexte), l'activité consiste en la modélisation du contexte (*de manière formelle ou non*), à sa prise en compte (*exploitation et évaluation*) jusqu'à l'adaptation (*approche type de la boucle MAPE-K [21] : Monitor, Analyse, Plan and Execute + Knowledge*). Cela pose des problèmes de modélisation de la connaissance, de la captation des données (*avec les problèmes classiques d'hétérogénéité, de normalisation...*), mais également des problèmes de décidabilité lorsque le contexte est utilisé à des fins d'autonomique.

Le fil rouge de toutes mes activités de recherche consiste à démontrer que les technologies issues du Web Sémantique (*modélisation RDF, ontologies, alignements, requêtes SPARQL, Web de données liées...*) peuvent être mises au service de thématiques scientifiques pour répondre à différents défis. Nous avons donné des exemples dans le cadre de l'adaptation de documents multimédia dans le chapitre 4, mais également en recherche d'information comme dans le chapitre 5 où l'on couple au sein d'un environnement hybride par exemple un corpus documentaire avec des données issues de capteurs.

Les défis auxquels je me consacre s'inscrivent dans les champs thématiques suivants issus de la classification ACM² : Information systems, Human-centered computing, Computing methodologies et Applied computing. Mes activités s'inscrivent également dans les axes de recherche prioritaires de l'Université de Pau et des Pays de l'Adour : Patrimoine & Tourisme³ ainsi qu'Énergie & Environnement⁴.

1. <https://liuppa.univ-pau.fr>

2. <https://dl.acm.org/ccs>

3. <https://recherche.univ-pau.fr>

4. <https://e2s-uppa.eu>

Dans les parties suivantes, je vais synthétiser mes publications (§7.2.1), mon encadrement doctoral et scientifique (§7.2.2) ainsi que mes activités de diffusion et de rayonnement d'activités de Recherche (§7.2.3).

7.2.1 Publications scientifiques

Depuis le début de mon doctorat en Informatique en 2004, j'ai rédigé 102 publications comme co-auteur dans des revues d'audience internationale (17), dans une revue d'audience nationale (1), dans des conférences d'audience internationale (35) ou nationale (10), dans des ateliers internationaux (10) ou nationaux (5), dans des chapitres d'ouvrages (6), pour la direction d'ouvrages (5), ainsi que pour l'élaboration d'actes scientifiques (13).

La liste complète de mes publications qui est classée de façon chronologique et groupée par type de production est disponible sur HAL (idHAL : sebastien-laborie)⁵. La table 7.1 ci-dessous présente une synthèse de mes publications regroupées par type, par audience et par classement. Le classement SJR (Scimago)⁶ est exploité dans ce tableau pour les articles de revues et le classement CORE⁷ est utilisé pour les articles de conférences.

Type	Audience	Rang	Nombre	Total
Journal	Internationale	Q1	3	18
		Q2	5	
		Q3	3	
		Aucun	6	
	Nationale	-	1	
Conférence	Internationale	A	2	45
		B	6	
		C	8	
		Aucun	19	
	Nationale	-	10	
Atelier	Internationale	-	10	15
	Nationale	-	5	
Chapitre	Internationale	-	4	6
	Nationale	-	2	
Direction ouvrage	Internationale	-	4	5
	Nationale	-	1	
Actes	-	-	13	13
Total :			102	

TABLE 7.1 – Type et classement de mes publications.

5. <https://cv.hal.science/sebastien-laborie>

6. <https://www.scimagojr.com>

7. <http://portal.core.edu.au/conf-ranks/>

Dans ce qui suit, je présente une liste de quatre publications récentes classées :

— **Dans une revue internationale de Rang Q1 :**

[A] Nabila Guennouni, Christian Sallaberry, **Sébastien Laborie**, Richard Chbeir, Elio Mansour. *ISEE : A heterogeneous information system for event explainability in smart connected environments*. Internet of Things, Elsevier, Décembre 2021, volume 16, pp.379-398.

— **Dans une revue internationale de Rang Q2 :**

[B] Nathalie Charbel, Christian Sallaberry, **Sébastien Laborie**, Richard Chbeir. *FEED2SEARCH : a framework for hybrid-molecule based semantic search*. International Journal of General Systems, Elsevier, pp.1-41, 2023.

— **Dans une revue internationale de Rang Q3 :**

[C] Asma Saighi, Zakaria Laboudi, Philippe Roose, **Sébastien Laborie**, Nassira Ghoualmi-Zine. *On Using Multiple Disabilities Profiles to Adapt Multimedia Documents*. International Journal of Information Technology and Web Engineering, 15(3), pp.34-60, 2020.

— **Dans une conférence internationale de Rang A :**

[D] Jorge Andrés Larracochea, Philippe Roose, Sergio Ilarri, Yudith Cardinale, **Sébastien Laborie**, Olga Paulina Vara. *Behavior-Based Consumption Profiles for the Approximation of the Energy Consumption of Services*. The 30th International Conference on Information Systems Development (ISD 2022), Août 2022, Cluj-Napoca, Romania.

Comme je l’ai indiqué précédemment, pour toutes ces publications j’apporte mes compétences dans le domaine du Web Sémantique. Les publications [A] et [B] traitent essentiellement de la thématique qui concerne la recherche d’information. Il s’agit d’optimiser l’indexation ainsi que l’interrogation de données que ce soit dans un environnement connecté [A] ou avec des documents hétérogènes [B]. La publication [C] porte sur la modélisation du contexte afin de déployer efficacement des composants logiciels ou des services d’adaptation pour des utilisateurs. Enfin, la publication [D] se concentre sur l’Informatique éco-responsable qui est un nouveau champ de recherche émergent au sein de l’équipe T2I et plus généralement dans mon laboratoire. J’investis donc actuellement cette thématique porteuse avec d’autres collègues notamment pour modéliser des profils énergétiques sémantiques de services.

Les publications citées précédemment ont été réalisées dans le cadre d’encadrements de thèses en informatique. Dans la partie suivante, je vais lister tous les encadrements que j’ai menés depuis mon recrutement au sein de l’UPPA.

7.2.2 Encadrement doctoral et scientifique

Depuis 2010, j’ai co-encadré sans discontinuer des doctorants avec plusieurs collègues de mon laboratoire. Vous trouverez ci-dessous une synthèse de mes encadrements (Table 7.2). Les sujets de thèse avec les encadrements ont été listés dans l’Introduction de ce manuscrit (§1.2).

Doctorant(e)	Rôle	Taux	Cotutelle	Début	Soutenance
Yunji Zhang	Codirecteur	50%	Non	Nov. 2023	-
Charlotte Darricades	Co-encadrant	25%	Non	Sept. 2022	-
Jorge Larracochea	Co-encadrant	25%	Oui	Mars 2021	juin 2024
Nabila Guennouni	Codirecteur	50%	Non	Avril 2019	28/06/2022
Nathalie Charbel	Codirecteur	50%	Non	Avril 2015	21/12/2018
Asma Saighi	Co-encadrant	33%	Non	Sept. 2014	18/10/2018
Irvin Dongo	Co-encadrant	10%	Non	Sept. 2014	20/12/2017
Regina Ticona	Co-encadrant	33%	Non	Sept. 2013	06/07/2016
The Nhan Luong	Co-encadrant	10%	Non	Sept. 2010	12/12/2012

TABLE 7.2 – Thèses codirigées et co-encadrées.

Bien entendu, j’ai aussi co-encadré plusieurs étudiants de niveau Master (7 au total) dans le cadre de projets ou de stages sur des durées de 4 à 6 mois. Comme pour les étudiants en thèse de doctorat, j’essaie de transmettre ma passion du métier d’enseignant-chercheur en les conseillant sur les aspects méthodologiques et la rigueur à acquérir lorsqu’on déroule une approche scientifique.

7.2.3 Diffusion et rayonnement

J’ai participé à plusieurs projets de recherche afin d’accroître mes collaborations au niveau national et international ainsi que de valoriser mes productions scientifiques. Vous trouverez dans la table 7.3 une synthèse des projets dans lesquels je me suis impliqué. Comme il est possible de le constater dans ce tableau, les types de projets dans lesquels j’ai évolué ont été variés allant de projets locaux (i.e., Université, Région et Industriel) à des projets de niveau national (ANR) et européen (ITEA2). Pour information, j’ai pu aussi participer et soumettre des propositions de projets de recherche en tant que membre ou en tant que porteur mais malheureusement non-acceptés (13 au total depuis 2010).

Nom du projet	Type de financement	Rôle	Début	Fin
B&V Landes	Région Aquitaine	Membre	2023	2026
BISE2	UPPA/E2S	Co-porteur	2018	2022
NobaSearch	Industriel	Membre	2015	2018
ROSCO	Transfrontalier	Membre	2013	2014
MOJITO	Région Aquitaine	Membre	2011	2014
MOANO	ANR	Membre	2010	2013
IWAAT	Région Aquitaine	Membre	2010	2011
LINDO	Européen	Membre	2008	2010

TABLE 7.3 – Projets de recherche menés.

Au-delà des collaborations au sein de projets de recherche, je suis membre d'associations et des réseaux de recherche suivants : GdR MaDICS (*Masses de Données, Informations et Connaissances en Sciences*)⁸, GdR MAGIS (*Méthodes et Applications pour la Géomatique et l'Information Spatiale*)⁹, ATIEF (*Association des Technologies de l'Information pour l'Éducation et la Formation*)¹⁰, SIF (*Société Informatique de France*)¹¹, et des réseaux régionaux de recherche tels que sur le Numérique en Éducation (R3 NumEd) ou Transition Énergétique Sur la Nouvelle Aquitaine (R3 TESNA).

Également, j'ai été membre de comités de rédaction pour les journaux et actes suivants : Numéro spécial en 2024 intitulé « *e-Health Systems and Technologies* » pour la revue Sensors de MDPI, Numéro spécial en 2020 intitulé *e-Health Pervasive Wireless Applications and Services* de l'Information Journal, Numéro spécial en 2018 intitulé « *Open et Linked Data dans les Systèmes d'Information* » pour la revue Ingénierie des Systèmes d'Information (ISI), Numéro spécial en 2015 intitulé *Social Media and networks for multimedia content management* de l'International Journal of Virtual Communities and Social Networking, Actes de l'atelier *e-Health Pervasive Wireless Applications and Services (eHPWAS)* chaque année de 2013 à 2023 en collaboration avec la conférence internationale WiMob, et enfin responsable scientifique des actes du 8th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP 2013).

De plus, j'ai participé à plusieurs comités de lecture de revues scientifiques ainsi que de conférences et ateliers nationaux et internationaux que ce soit en tant que membre ou de relecteur externe :

- **Revues scientifiques (8)** : Numéro spécial de la revue Ingénierie des Systèmes d'Information sur le thème INFORSID 2020, Information Systems Journal (2017), Numéro spécial de la Revue Ingénierie des Systèmes d'Information sur le thème "*Web de données : publication, liage et capitalisation*" (2016), Numéro spécial de la revue Ingénierie des Systèmes d'Information sur le thème INFORSID 2014-2015 (2015), Engineering Applications of Artificial Intelligence Journal (2015), World Wide Web Journal (2014), Computer Standards and Interfaces Journal (2014) et IEEE Multimedia Special Issue on Multimedia Data Management in Mobile Computing (2014).
- **Conférences & Ateliers (42)** : CNRIUT 2025, CNRIUT 2024, CSAE 2023, SMAP 2023, CNRIUT 2023, SMAP 2022, INFORSID 2022, eHPWAS de 2013 à 2023, SMAP 2021, SMAP 2020, INFORSID 2020 (JCJC), SMAP 2019, SMAP 2018, MEDES 2017, INFORSID 2017, SMAP 2017, Ubimob 2016, SMAP 2015, IWCIM 2014, SMAP 2014, INFORSID 2014, RTE 2013, Ubimob 2013, MEDES 2013, CORIA 2013, INFORSID 2013, Ubimob 2012, RTE 2012, RTE 2011, INFORSID 2011 pour la thématique Systèmes d'information pervasifs et adaptation, SMAP 2010, SMAP 2009 et SMAP 2008.

8. <https://www.madics.fr>

9. <https://gdr-magis.cnrs.fr>

10. <http://www.atief.fr>

11. <https://www.societe-informatique-de-france.fr>

Enfin, j'ai également participé aux comités d'organisation de plusieurs conférences et ateliers que ce soit à un niveau national ou à l'international :

- **Au niveau international (24)** : International Conference on Intelligent Environments 2022, e-Health Pervasive Wireless Applications and Services (eHPWAS) chaque année de 2013 à 2023 en collaboration avec la conférence internationale WiMob, Smart Ubiquitous Context Service Discovery & Adaptation (SuCDA) chaque année de 2016 à 2024 en collaboration avec la conférence internationale ANT, SMAP 2017 (responsable de la publicité), co-responsable de l'atelier « *Social Media and Networks for Tourism* » pour MEDES 2016, MEDES 2015 (webmaster), AIAI 2015, SMAP 2013 (co-responsable avec Philippe Roose).
- **Au niveau national (10)** : responsable du CNRIUT 2025, co-porteur de l'atelier « *Open et/ou Linked Data dans les systèmes d'information* » pour INFORSID 2017, INFORSID 2015, co-porteur de la session « *Impact des *-Data sur les Systèmes d'Information* » pour INFORSID 2014, co-porteur de la session « *Big, Open et Linked Data* » pour INFORSID 2013, EGC 2013, NOTERE/CFIP 2012, Ubimob 2012, Atelier RTE 2011 (co-responsable) et INFORSID 2009.

En résumé, je participe activement à l'animation scientifique au niveau international et national que ce soit pour des événements récurrents (e.g., eHPWAS, SMAP, SuCDA, CNRIUT) ou bien pour répondre à des sollicitations ponctuelles telles que l'édition de numéro spéciaux, la relecture d'articles, l'organisation de congrès... Dans la partie suivante, je vais montrer que je me suis également engagé dans des responsabilités collectives au sein de mon Université mais également au-delà.

7.3 Responsabilités collectives

Au sein d'un établissement tel qu'un IUT, il est primordial de s'investir dans les tâches collectives contribuant au bon fonctionnement des formations (§7.3.1). Au-delà de ma composante d'affectation, il me semblait également important de contribuer à des responsabilités plus larges au sein de l'Université (§7.3.2) ainsi qu'à un niveau national (§7.3.3). Aussi, depuis que j'ai été recruté en 2010 et suite à ma titularisation en 2011, j'ai toujours assuré des responsabilités afin de participer pleinement à la vie universitaire (Figure 7.1).

7.3.1 Responsabilités administratives

Quatre responsabilités administratives importantes ont fait partie de mon parcours au sein de l'UPPA que ce soit à l'IUT de Bayonne Pays Basque ou au sein du LIUPPA : (1) Directeur des études pour le DUT GEA de l'IUT de Bayonne Pays Basque, (2) Chef du Département GEA de l'IUT de Bayonne Pays Basque, (3) Responsable de l'équipe de Recherche T2I du LIUPPA, et (4) Chargé de mission " *Soutien à la Recherche*" pour l'IUT de Bayonne Pays Basque.

De septembre 2011 à août 2017, j'ai assuré la charge de Directeur des études pour le DUT GEA de l'IUT de Bayonne Pays Basque. Une de mes premières missions a consisté à automatiser et gérer les emplois du temps du Département. Une autre mission a été d'assister le Chef de Département pour produire les services de tous les enseignants du Département et de lui apporter mon aide lors des délibérations au sein des jurys de DUT.

De septembre 2017 à août 2020, j'ai assuré les fonctions de Chef du Département GEA de l'IUT de Bayonne Pays Basque. J'ai organisé la vie du Département GEA que ce soit au niveau pédagogique et au niveau administratif. J'ai piloté le recrutement post-bac (*avec APB puis Parcoursup*), dirigé l'action des enseignants ayant accepté des charges de responsabilité pédagogique dans le département, participé au recrutement d'enseignants (*vacataires, PAST, CDI*) ainsi qu'à l'évaluation des personnels BIATSS du Département (4 secrétaires), représenté le Département à l'Assemblée des Chefs de départements et à toutes réunions concernant le Département (*Conseil de Direction, Conseil d'IUT, réunions internes et externes*), géré le budget du Département, organisé les sous-commissions de jury pour chaque semestre, délibéré en jury d'IUT et animé dans sa globalité le Département.

Depuis septembre 2020, je suis responsable de l'équipe T2I du LIUPPA. Je représente l'équipe au sein du Conseil de Laboratoire, de présenter les résultats de l'équipe lors par exemple d'évaluation HCERES (*ce que j'ai effectué en février 2021 pour l'évaluation HCERES du LIUPPA*), de gérer le budget de l'équipe, d'inciter et de recenser les projets de recherche, de promouvoir les activités de l'équipe (*via le site web du laboratoire*) ainsi que d'animer la vie de l'équipe (*via des séminaires ainsi que des réunions mensuelles*).

En parallèle à la responsabilité de l'équipe T2I, je suis chargé de mission « Soutien à la Recherche » pour l'IUT de Bayonne Pays Basque depuis septembre 2021. Cette responsabilité consiste à accueillir les personnels permanents et non-permanents ayant des activités de recherche à l'IUT, promouvoir les activités de recherche au sein de l'Institut (*Communication interne et externe*), soutenir l'organisation d'événements scientifiques (ex., colloques, soutenances de thèse, fêtes de la Science, séminaires) et aider financièrement les actions relevant de la Recherche.

7.3.2 Responsabilités et mandats locaux

Au sein de l'Université de Pau et des Pays de l'Adour, je participe à la vie universitaire au-delà de mon IUT d'affectation et de mon équipe de recherche.

Depuis septembre 2020, je suis membre du Conseil du Laboratoire d'Informatique de l'UPPA. Notre Conseil de laboratoire se réunit tous les mois et il définit la politique du LIUPPA. De façon globale, j'ai pu débattre des campagnes d'emplois, des fiches de postes, des budgets ainsi que des actions stratégiques pour notre laboratoire. Pour chaque conseil, j'ai établi des compte-rendu que je mets à disposition des collègues.

Depuis octobre 2022 et auparavant de septembre 2012 à août 2015, j'ai été élu membre du Comité d'experts de l'UPPA pour la section CNU 27ème Informatique. Durant ce mandat, j'ai notamment participé aux recrutements des candidats aux postes d'ATER et de PAST pour l'Université.

7.3.3 Responsabilités et mandats nationaux

Au-delà de l'Université de Pau et des Pays de l'Adour, je collabore activement avec des instances à rayonnement national comme l'ACD GEA, l'ADIUT ou le CNU.

De septembre 2017 à février 2021, j'ai été élu au sein du Bureau de l'Assemblée des Chefs de Départements GEA (ACD GEA). Ce Bureau contient 10 membres élus par environ 80 Chefs de Départements GEA de France. Durant mes mandats annuels, on m'a confié la responsabilité de la Communication interne et externe du réseau GEA de France. J'ai également créé le site officiel des Départements GEA de France pour en faire la promotion auprès du grand public : www.but-gea.fr. Durant cette période, j'ai également été fortement impliqué dans la mise en place du nouveau diplôme du Bachelor Universitaire de Technologie (BUT), et en particulier pour tout ce qui concerne l'élaboration du référentiel national de compétences du nouveau Programme National en GEA.

Depuis décembre 2021, je suis membre de la commission RTI (Recherche, Transfert de technologie et Innovation) de l'ADIUT. Je participe à la relecture ainsi qu'à l'organisation des CNR IUT (*Congrès National de la Recherche en IUT*). De plus, nous engageons des réflexions ainsi que des initiatives pour favoriser la Recherche en IUT (*annuaire des acteurs de la Recherche en IUT, recensement des plateaux techniques, renforcement du lien BUT et Recherche...*).

En septembre 2023, je me suis porté candidat sur une liste électorale du collègue B de la Section 27 pour les élections qui concernent le renouvellement des membres de la CNU. La CNU est l'instance nationale qui gère, entre autres, les dossiers de promotions, la qualification et le suivi en général de la carrière ainsi que des missions des enseignants-chercheurs.

7.4 Synthèse du chapitre

Mes activités d'enseignant-chercheur ont été diverses et dynamiques durant toutes ces années passées que ce soit au niveau de l'enseignement, de la Recherche et des responsabilités collectives. C'est la vision que je me fais de ce poste : s'engager pleinement dans toutes les facettes de nos missions.

Au niveau de l'enseignement, j'ai dû construire de nouveaux supports afin qu'il soit conforme à la réforme du BUT et je propose aux étudiants l'usage d'outils pédagogiques innovants que ce soit via des plates-formes universitaires d'enseignement en ligne ou d'outils interactifs (e.g., chaîne Youtube). Je m'implique fortement dans mon équipe pédagogique tant au niveau des responsabilités (Chef de Département et res-

ponsable de plusieurs modules) que dans la vie au quotidien de l'IUT (accompagnement des étudiants, réunions pédagogiques. . .).

Au niveau de la Recherche, je forme plusieurs doctorants au métier d'enseignant-chercheur. Mes publications régulières permettent de valoriser l'ensemble de mes activités au travers des collaborations ainsi que des projets de recherche. Enfin, je m'implique dans l'animation scientifique de mon laboratoire via l'organisation de congrès ainsi que d'ateliers au niveau national et international, et en tant que membre de comités de rédaction et de lecture pour des revues ou des conférences. En qualité de responsable d'une équipe de recherche, j'ai pu valoriser nos résultats lors de l'évaluation HCERES du LIUPPA en 2021 et je m'efforce à inciter mes collègues à des dépôts de projets ainsi qu'à développer des collaborations de part notamment l'organisation de séminaires avec des chercheurs invités.

Au niveau des responsabilités collectives, j'ai souhaité m'impliquer dans différentes structures que ce soit au niveau local à l'UPPA comme au sein de l'IUT de Bayonne Pays Basque (Chef de Département, Chargé de mission Soutien à la Recherche) et de mon laboratoire de Recherche (responsable de l'équipe de recherche T2I et membre du CdL du LIUPPA), mais aussi que ce soit au niveau national (membre du Bureau de l'ACD GEA, membre de la commission RTI de l'ADIUT, membre d'une liste électorale de la CNU 27).

Pour conclure, je souhaitais indiquer dans ce manuscrit que je m'épanouis pleinement dans l'ensemble de mes missions d'enseignant-chercheur. Je souhaite continuer à me fixer de nouveaux objectifs de réalisation dans mes tâches et responsabilités. Notre métier nous offre l'opportunité de diversifier nos actions au niveau local, national et international. Je crois saisir humblement ces opportunités car elles impulsent une dynamique sur notre investissement et enrichit notre expérience.

Mes publications

La liste complète de mes publications est également disponible sur HAL à l'adresse suivante : <https://cv.archives-ouvertes.fr/sebastien-laborie>.

Articles dans revues internationales à comité de lecture (17)

- Nathalie Charbel, Christian Sallaberry, **Sébastien Laborie**, Richard Chbeir. FEED2SEARCH : a framework for hybrid-molecule based semantic search. *International Journal of General Systems*, 2023, pp.1-41. [⟨10.1080/03081079.2023.2195173⟩](https://doi.org/10.1080/03081079.2023.2195173) (rang Q2)
- Nabila Guennouni, Christian Sallaberry, **Sébastien Laborie**, Richard Chbeir, Elio Mansour. ISEE : A heterogeneous information system for event explainability in smart connected environments. *Internet of Things*, Elsevier, 2021. [⟨10.1016/j.iot.2021.100457⟩](https://doi.org/10.1016/j.iot.2021.100457) (rang Q1)
- Asma Saighi, Zakaria Laboudi, Philippe Roose, **Sébastien Laborie**, Nassira Ghoualmi-Zine. On Using Multiple Disabilities Profiles to Adapt Multimedia Documents. *International Journal of Information Technology and Web Engineering*, Idea group publishing, 2020, 15 (3), pp.34-60. [⟨10.4018/IJITWE.2020070103⟩](https://doi.org/10.4018/IJITWE.2020070103) (rang Q3)
- Abderrahim Lakehal, Adel Alti, **Sébastien Laborie**, Philippe Roose. A Semantic Agile Approach for Reconfigurable Distributed Applications in Pervasive Environments. *International Journal of Ambient Computing and Intelligence*, IGI Pub, 2020, 11 (2), pp.48-67. [⟨10.4018/IJACI.2020040103⟩](https://doi.org/10.4018/IJACI.2020040103) (rang Q3)
- Achouri Mounir, Alti Adel, Derdour Makhoulf, **Sébastien Laborie**, Philippe Roose. A New Two-Level Clustering Approach for Situations Management in Distributed Smart Environments. *International Journal of Ambient Computing and Intelligence*, IGI Pub, 2019, 10 (2), pp.91-111. [⟨10.4018/IJACI.2019040107⟩](https://doi.org/10.4018/IJACI.2019040107) (rang Q3)
- Mounir Achouri, Adel Alti, Derdour Makhoulf, **Sébastien Laborie**, Philippe Roose. Smart fog computing for efficient situations management for smart health environments. *Journal of Information and Communication Technology (JICT)*, 2018, 17 (4), pp.537-567. (rang Q2)
- Kamel Mansouri, Adel Alti, Philippe Roose, **Sébastien Laborie**. Dynamic semantic-based green bio-inspired approach for optimizing energy and cloud services qua-

- lities. Transactions on emerging telecommunications technologies, Wiley-Blackwell, 2018, 29 (5). [⟨10.1002/ett.3305⟩](#) (rang Q2)
- Asma Saighi, Philippe Roose, Nassira Ghoulmi, **Sébastien Laborie**, Zakaria Laboudi. HaMA : A Handicap-based Architecture for Multimedia Document Adaptation. International Journal of Multimedia Data Engineering and Management (IJMDEM), 2017, 8 (3), pp.55-96. [⟨10.4018/IJMDEM.2017070104⟩](#)
 - Alti Adel, Philippe Roose, **Sébastien Laborie**. Multimedia Documents Adaptation Based on Semantic Multi-Partite Social Context-Aware Networks. IJVCSN, International Journal of Virtual Communities and Social Networking, George Washington University, State University of New York at Plattsburgh, 2017, 9 (3), pp.44-59. [⟨10.4018/IJVCSN.2017070104⟩](#)
 - Christophe Marquesuzaà, Patrick Etcheverry, **Sébastien Laborie**, Thierry Nodet, The Nhan Luong. User-Centered Design of Geographic Interactive Applications : From High-Level Specification to Code Generation, from Prototypes to Better Specifications. Transactions on Large-Scale Data and Knowledge-Centered Systems, Springer Berlin / Heidelberg, 2016, LNCS 10140, pp.1-29. [⟨10.1007/978-3-662-54173-9_1⟩](#)
 - Adel Alti, Abderrahim Lakehal, **Sébastien Laborie**, Philippe Roose. Autonomic Semantic-Based Context-Aware Platform for Mobile Applications in Pervasive Environments. Future internet, MDPI, 2016, 8(4). [⟨10.3390/fi8040048⟩](#) (rang Q2)
 - Adel Alti, **Sébastien Laborie**, Philippe Roose. A Community-Based Semantic Social Context-Aware Driven Adaptation for Multimedia Documents. International Journal of Virtual Communities and Social Networking, 2015, 7 (2), pp.31-49. [⟨10.4018/IJVCSN.2015040102⟩](#)
 - Adel Alti, **Sébastien Laborie**, Philippe Roose. Dynamic semantic-based adaptation of multimedia documents. Transactions on emerging telecommunications technologies, Wiley-Blackwell, 2014, 25 (2), pp.239-258. [⟨10.1002/ett.2677⟩](#) (rang Q2)
 - **Sébastien Laborie**, Bouchra Soukkaïeh, Florence Sèdes. On using generic profiles and views for dynamic web services adaptation. Journal of Modern Internet of Things, 2013, vol. 2 (n° 3), pp. 18-23.
 - Cédric Dromzée, **Sébastien Laborie**, Philippe Roose. A Semantic Generic Profile for Multimedia Documents Adaptation. Dimitris Kanellopoulos. Intelligent Multimedia Technologies for Networking Applications : Techniques and Tools, IGI Global, pp.225-246, 2013. [⟨10.4018/978-1-4666-2833-5.ch009⟩](#)
 - **Sébastien Laborie**, Jérôme Euzenat, Nabil Layaida. Semantic adaptation of multimedia documents. Multimedia Tools and Applications, Springer Verlag, 2011, 55 (3), pp.379-398. [⟨10.1007/s11042-010-0552-9⟩](#) (rang Q1)
 - **Sébastien Laborie**, Ana-Maria Manzat, Florence Sèdes. Managing and querying efficiently distributed semantic multimedia metadata collections. IEEE Multime-

dia Special Issue on Multimedia-Metadata and Semantic Management, 2009, 16 (4), pp.12-21. [10.1109/MMUL.2009.68](https://doi.org/10.1109/MMUL.2009.68) (rang Q1)

Article dans une revue nationale à comité de lecture (1)

- Patrick Etcheverry, **Sébastien Laborie**, Christophe Marquesuzaà, Thierry Node-not, The Nhân Luong. Conception d'applications web géographiques guidée par les contenus et les usages : cadre méthodologique et opérationnalisation avec l'environnement WINDMash. Journal d'Interaction Personne-Système, Association Francophone d'Interaction Homme-Machine (AFIHM), 2014, 3 (1), pp.1-42.

Chapitres d'ouvrage à audience internationale (4)

- Tayeb Lemlouma, **Sébastien Laborie**, Philippe Roose, Abderrezak Rachedi, Kenza Abdelaziz. mHealth Contents and Services Delivery and Adaptation Challenges for Smart Environments. Sasan Adibi. mHealth Multidisciplinary Verticals, CRC Press/Taylor & Francis, pp.295-314, 2014, 978-1-4822-1480-2.
- Adel Alti, **Sébastien Laborie**, Philippe Roose. Enrich the Expressiveness of Multimedia Document Adaptation Processes. Evaggelos Spyrou, Dimitris Lakovidis, Phivos Mylonas. Semantic Multimedia Analysis and Processing, CRC Press, pp.185-217, 2014, Digital Imaging and Computer Vision, 978-1-4665-7549-3.
- **Sébastien Laborie**, Antoine Zimmermann. On using information retrieval techniques for semantic media adaptation. Manolis Wallace, Marios Angelides, Phivos Mylonas. Advances in Semantic Media Adaptation and Personalization, CRC Press, pp.137-157, 2009.
- **Sébastien Laborie**, Jérôme Euzenat. An incremental framework for adapting the hypermedia structure of multimedia documents. Manolis Wallace, Marios Angelides, Phivos Mylonas. Advances in Semantic Media Adaptation and Personalization, 93 (8), Springer, pp.157-176, 2008, Studies in computational intelligence, 978-3-540-76359-8. [10.1007/978-3-540-76361_8](https://doi.org/10.1007/978-3-540-76361_8)

Chapitres d'ouvrage à audience nationale (2)

- Adel Alti, **Sébastien Laborie**, Philippe Roose. Enrichir l'expressivité des profils et des processus d'adaptation de contenus multimédias. Chantal Soulé-Dupuy. Variabilité et adaptation de l'accès à l'information, 17(2), Lavoisier, pp.101-124, 2014, Document Numérique, 9782746246683. [10.3166/DN.17.2.101-124](https://doi.org/10.3166/DN.17.2.101-124)
- Max Chevalier, Christine Julien, **Sébastien Laborie**, Chantal Soulé-Dupuy. Adaptation des systèmes d'accès à l'information et des documents. Philippe Lopis-

téguy, Dominique Rieu, Philippe Roose. L'adaptation dans tous ses états, Cépaduès, pp.39-71, 2012.

Publications dans des conférences internationales (35)

- Jorge Andrés Larracochea, Philippe Roose, Sergio Ilarri, Yudith Cardinale, **Sébastien Laborie**, Olga Paulina Vara. Behavior-Based Consumption Profiles for the Approximation of the Energy Consumption of Services. International Conference on Information Systems development (ISD), Aug. 2022, Cluj-Napoca, Romania. (rang A)
- Nabila Guennouni, **Sébastien Laborie**, Christian Sallaberry, Richard Chbeir, Elio Mansour. ISEEapp : An Event Explanation Prototype bridging the gap between sensor network and document corpora data. The 18th International Conference on Intelligent Environments (IE 2022), Jun 2022, Biarritz, France. [⟨10.1109/IE54923.2022.9826763⟩](https://doi.org/10.1109/IE54923.2022.9826763) (rang B)
- Jorge Andrés Larracochea, Philippe Roose, Sergio Ilarri, Yudith Cardinale, **Sébastien Laborie**, Mauricio Jacobo González. Towards Services Profiling for Energy Management in Service-Oriented Architectures. The 17th International Conference on Web Information Systems and Technologies (WEBIST 2021), Oct. 2021. **Best Student Paper Award** (rang C)
- Philippe Roose, Sergio Ilarri, Jorge Andrés Larracochea, Yudith Cardinal, **Sébastien Laborie**. Towards an Integrated Full-Stack Green Software Development Methodology. The 29th International Conference on Information Systems Development (ISD 2021), Sept. 2021, Valencia, Spain. (rang A)
- Nabila Guennouni, Christian Sallaberry, **Sébastien Laborie**, Richard Chbeir. A Novel Framework for Event Interpretation in a Heterogeneous Information System. MEDES '20 : 12th International Conference on Management of Digital EcoSystems, Nov 2020, Virtual Event United Arab Emirates, France. pp.140-148. [⟨10.1145/3415958.3433073⟩](https://doi.org/10.1145/3415958.3433073) (rang C)
- Nathalie Charbel, Christian Sallaberry, **Sébastien Laborie**, Richard Chbeir. Hybrid Molecule-based Information Retrieval. The 34th ACM/SIGAPP Symposium On Applied Computing (ACM SAC 2019), Apr 2019, Limassol, Cyprus. pp.808-815. [⟨10.1145/3297280.3297358⟩](https://doi.org/10.1145/3297280.3297358) (rang B)
- Mounir Achouri, Alti Adel, Makhoulf Derdour, **Sébastien Laborie**, Philippe Roose. A New Two-Layered Architecture for Efficient Situations Management in Smart Environments. The 9th International ACM Conference on Management of Digital EcoSystems (MEDES 2017), Nov 2017, Bangkok, France. pp.44-59. [⟨10.1145/3167020.3167022⟩](https://doi.org/10.1145/3167020.3167022) (rang C)
- Nathalie Charbel, Christian Sallaberry, **Sébastien Laborie**, Gilbert Tekli, Richard Chbeir. LinkedMDR : A Collective Knowledge Representation of a Heterogeneous Document Corpus. The 28th International Conference on Database and Expert

- Systems Applications (DEXA 2017), Aug 2017, Lyon, France. pp.362-377. [⟨10.1007/978-3-319-64468-4_28⟩](https://doi.org/10.1007/978-3-319-64468-4_28) (rang C)
- Asma Saighi, Philippe Roose, Nassira Ghoulami, **Sébastien Laborie**, Zakaria Laboudi. Multimedia Presentation Adaptation Architecture Assisted by Ontological Reasoning. The 2nd International Conference on Pattern Analysis and Intelligent Systems (PAIS 2016), Nov 2016, Khenchela, Algeria.
 - Asma Saighi, Philippe Roose, Nassira Ghoulami, **Sébastien Laborie**, Zakaria Laboudi. Ontological Reasoning for the Personalization of Multimedia Adaptation Services in Smart Home Domain. The Second International Symposium on Informatics and its Applications (ISIA 2016), Nov 2016, M'sila, Algeria.
 - Sarah Slimani, Adel Alti, **Sébastien Laborie**, Philippe Roose. An Advanced Adaptation Approach for Pervasive Applications. The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016), May 2016, Madrid, Spain.
 - The Nhan Luong, Christophe Marquesuzaà, Patrick Etcheverry, Thierry Node-not, **Sébastien Laborie**. Facilitating the Design/Evaluation Process of Web-based Geographic Applications : a Case Study with WINDMash. The 2nd International Conference on Future Data and Security Engineering 2015 (FDSE), Nov 2015, Ho Chi Minh City, Vietnam. pp.259-271. [⟨10.1007/978-3-319-26135-5_19⟩](https://doi.org/10.1007/978-3-319-26135-5_19)
 - Adel Alti, Bellal Wassim, **Sébastien Laborie**, Philippe Roose. CSDMAP : Cloud Semantic-based Dynamic Multimodal Adaptation Platform. The 7th International Conference on Management of computational and collective Intelligence in Digital EcoSystems (MEDES'15), Oct 2015, Caraguatatuba/Sao Paulo, Brazil. pp.161-168. [⟨10.1145/2857218.2857255⟩](https://doi.org/10.1145/2857218.2857255) (rang C)
 - Adel Alti, **Sébastien Laborie**, Philippe Roose. Cloud Semantic-based Dynamic Multimodal Platform for Building mHealth Context-aware Services. The 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2015), Oct 2015, Abu Dhabi, United Arab Emirates. pp.357-364, [⟨10.1109/WiMOB.2015.7347984⟩](https://doi.org/10.1109/WiMOB.2015.7347984) (rang B)
 - Regina Ticona Herrera, Joe Tekli, Richard Chbeir, **Sébastien Laborie**, Irvin Dongo, Renato Guzman. Toward RDF Normalization. The 34th International Conference on Conceptual Modeling (ER 2015), Oct 2015, Stockholm, Sweden. pp.261-275. [⟨10.1007/978-3-319-25264-3_19⟩](https://doi.org/10.1007/978-3-319-25264-3_19)
 - Adel Alti, **Sébastien Laborie**, Philippe Roose. USSAP : Universal Smart Social Adaptation Platform. The 6th International Conference on Ambient Systems, Networks and Technologies (ANT-2015), Jun 2015, London, United Kingdom. pp.670-674, [⟨10.1016/j.procs.2015.05.070⟩](https://doi.org/10.1016/j.procs.2015.05.070)
 - Adel Alti, **Sébastien Laborie**, Philippe Roose. Semantic Context-aware Adaptation Platform Architecture. The 5th International Conference on Ambient Systems, Networks and Technologies (ANT-2014), Jun 2014, Hasselt, Belgium. pp.959-964. [⟨10.1016/j.procs.2014.05.518⟩](https://doi.org/10.1016/j.procs.2014.05.518)

- Adel Alti, **Sébastien Laborie**, Philippe Roose. A framework for managing and optimizing the adaptation process qualities. The 7th IEEE GCC Conference and Exhibition (GCC'13), Nov 2013, Doha, Qatar. pp.125-130.
- Adel Alti, **Sébastien Laborie**, Philippe Roose. Towards a Context-Aware Service and Quality Multimedia Adaptation for Healthcare Applications. The International Conference on Digital Information Processing, E-Business and Cloud Computing (DIPECC2013), Oct 2013, Dubai, United Arab Emirates. pp.113-122.
- Adel Alti, **Sébastien Laborie**, Philippe Roose. Automatic Adaptation of Multimedia Documents. International Symposium on Frontiers in Ambient and Mobile Systems (FAMS-2013) in conjunction with the 4th International Conference on Ambient Systems, Networks and Technologies (ANT-2013), Jun 2013, Halifax, Canada.
- Tayeb Lemlouma, **Sébastien Laborie**, Philippe Roose. Toward a Context-Aware and Automatic Evaluation of Elderly Dependency in Smart Homes and Cities. The Fourteenth International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM/SCUCA), Jun 2013, Madrid, Spain. (rang B)
- Adel Alti, **Sébastien Laborie**, Philippe Roose, Reffad Hamza. Dynamic Quality Management of Emergency Systems via Rich Explicit Constraints. The 5th International Conference on Web and Information Technologies (ICWIT'13), May 2013, Hammamet, Tunisia. pp.531-540.
- Adel Alti, Philippe Roose, R. Saffidine, **Sébastien Laborie**. Towards an Automatic Adaptation of Heterogeneous Multimedia Mobile Applications. NOTERE / CFIP 2012, Oct 2012, Anglet, France. pp.79-87.
- Quang Pham Hai, **Sébastien Laborie**, Philippe Roose. On-the-fly Multimedia Document Adaptation Architecture. Pervasive Environments (SUPE'2012), in conjunction with the 9th International Conference on Mobile Web Information Systems (MobiWIS'12), Aug 2012, Ontario, Canada. pp.1-6.
- Mihaela Brut, **Sébastien Laborie**, Ana-Maria Manzat, Florence Sèdes. Generic Information System Architecture for Distributed Multimedia Indexation and Management. The 15th East-European Conference on Advances in Databases and Information Systems (ADBIS'11), Sep 2011, Vienne, Austria. pp.347-360, [⟨10.1007/978-3-642-23737-9_25⟩](#) (rang C)
- The Nhan Luong, **Sébastien Laborie**, Thierry Nodenot. A Framework with Tools for Designing Web-based Geographic Applications. The 11th ACM Symposium on Document Engineering, Sep 2011, Mountain View, California, United States. pp.33-42. [⟨10.1145/2034691.2034699⟩](#) (rang B)
- Dana Codreanu, **Sébastien Laborie**, Florence Sèdes. Modelling Multimedia Indexing Algorithms Description for Implicit and Explicit Indexation. 2011 IEEE/FTRA International Conference on Multimedia and Ubiquitous Engineering (MUE), Jun 2011, Crete, France. [⟨10.1109/MUE.2011.49⟩](#)

- **Sébastien Laborie**, Ana-Maria Manzat, Florence Sèdes. A Generic Framework for the Integration of Heterogeneous Metadata Standards into a Multimedia Information Retrieval System. International Conference on Adaptivity, Personalization and Fusion of Heterogeneous Information, Apr 2010, Paris, France. pp.1-4.
- Mihaela Brut, **Sébastien Laborie**, Ana-Maria Manzat, Florence Sèdes. A Generic Metadata Framework for the Indexation and the Management of Distributed Multimedia Contents. 3rd International Conference on New Technologies, Mobility and Security (NTMS 2009), Dec 2009, Cairo, Egypt. [⟨10.1109/NTMS.2009.5384702⟩](https://doi.org/10.1109/NTMS.2009.5384702)
- Mihaela Brut, **Sébastien Laborie**, Ana-Maria Manzat, Florence Sèdes. A Framework for Automatizing and Optimizing the Selection of Indexing Algorithms. the Third International Conference on Metadata and Semantics Research (MTSR'09), Sep 2009, Milan, Italy. pp.48-59, [⟨10.1007/978-3-642-04590-5_5⟩](https://doi.org/10.1007/978-3-642-04590-5_5)
- Mihaela Brut, **Sébastien Laborie**, Ana-Maria Manzat, Florence Sèdes. Integrating Heterogeneous Metadata into a Distributed Multimedia Information System. COGIS 2009, Nov 2009, Paris, France.
- Faisal Alkhateeb, **Sébastien Laborie**. Towards extending and using SPARQL for modular document generation. Proc. 8th ACM symposium on document engineering (DocEng), Sep 2008, São Paulo, Brazil. pp.164-172, [⟨10.1145/1410140.1410174⟩](https://doi.org/10.1145/1410140.1410174) (rang B)
- Jean-François Baget, **Sébastien Laborie**. Bi-Intervals for Backtracking on Temporal Constraint Networks. TIME : Temporal Representation and Reasoning, Jun 2007, Alicante, Spain. pp.163-168. [⟨10.1109/TIME.2007.44⟩](https://doi.org/10.1109/TIME.2007.44) (rang C)
- **Sébastien Laborie**, Jérôme Euzenat, Nabil Layaïda. A Spatial Algebra for Multimedia Document Adaptation. The 1st International Conference on Semantic and Digital Media Technologies (SAMT), Dec 2006, Athènes, Greece. pp.7-8. (rang C)
- **Sébastien Laborie**. Spatio-temporal proximities for multimedia document adaptation. Proc. 12th Artificial intelligence : methodology, systems and applications (AIMSA), Sep 2006, Varna, Bulgaria. pp.128-137. [⟨10.1007/11861461_15⟩](https://doi.org/10.1007/11861461_15)

Publications dans des conférences nationales (10)

- **Sébastien Laborie**, Nabila Guennouni, Christian Sallaberry. ISEE : Un système pour l'explication d'événements dans les environnements hybrides. Congrès National de la Recherche des IUT (CNRIUT), Jun 2023, Saint-Pierre (La Réunion), France.
- Jorge Andrés Larracochea, Philippe Roose, Sergio Ilarri, **Sébastien Laborie**, Judith Cardinale. RADIANCE : A tool for software behavior design and energy consumption categorization. INFORSID 2023, Mai 2023, La Rochelle, France.
- Charlotte Darricades, **Sébastien Laborie**, Christian Sallaberry, Eric Kergosien, Patrice de La Broise. Identifier les traces pertinentes dans la documentation scien-

- tifique d'une entreprise à des fins de communication institutionnelle. Le cas de la Compagnie TotalEnergies. INFORSID 2023, Mai 2023, La Rochelle, France.
- Jorge Andrés Larracochea, Philippe Roose, Sergio Ilarri, Yudith Cardinale, **Sébastien Laborie**. Profilage de services pour la gestion de l'énergie dans les architectures orientées services. INFORSID 2022, Mai 2022, Dijon, France.
 - Nathalie Charbel, Christian Sallaberry, **Sébastien Laborie**, Gilbert Tekli, Richard Chbeir. LinkedMDR : un modèle sémantique de représentation de corpus de documents multimédia. 35ème congrès INFormatique des ORganisations et Systèmes d'Information et de Décision (INFORSID 2017), Mai 2017, Toulouse, France. pp.11-26.
 - **Sébastien Laborie**, Franck Ravat, Jiefu Song, Olivier Teste. Combining Business Intelligence with Semantic Web : Overview and Challenges. 33ème Congrès sur l'INFormatique des ORganisations et Systèmes d'Information et de Décision (INFORSID 2015), Mai 2015, Biarritz, France.
 - Cédric Dromzee, **Sébastien Laborie**, Philippe Roose. Profil générique sémantique pour l'adaptation de documents multimédias. 30ème congrès INFORSID (INFormatique des ORganisations et Systèmes d'Information et de Décision) - INFORSID'12, Mai 2012, Montpellier, France. pp.16.
 - **Sébastien Laborie**, Dana Codreanu, Florence Sèdes. Spécification et utilisation d'un modèle de description d'algorithmes d'indexation. Colloque Veille Stratégique Scientifique et Technologique (VSST 2010), Oct 2010, Toulouse.
 - **Sébastien Laborie**, Ana-Maria Manzat, Florence Sèdes. Création et utilisation d'un résumé de métadonnées pour interroger efficacement des collections multimédias distribuées. 27ème Congrès Informatique des Organisations et Systèmes d'Information et de Décision (INFORSID 2009), May 2009, Toulouse, France. pp.227-242.
 - **Sébastien Laborie**, Jérôme Euzenat, Nabil Layaida. Adaptation spatiale efficace de documents SMIL. 15e congrès francophone Reconnaissance des Formes et Intelligence Artificielle (RFIA), 2006, pp.127.

Publications dans des ateliers internationaux (10)

- Mohamed Amine Boudia, Philippe Roose, **Sébastien Laborie**. Smart city : Road flow management. 3rd International Workshop on Information Systems Engineering for Smarter Life (ISESL) – CAISE 2023, Jun 2023, Zaragoza, Spain.
- Abderrahim Lakehal, Adel Alti, **Sébastien Laborie**, Philippe Roose. Ontology-based Context-aware Recommendation Approach For Dynamic Situations Enrichment. 13th International Workshop on Semantic and Social Media Adaptation and Personalization, Sep 2018, Zaragoza, Spain. pp.81-86. <10.1109/SMAP.2018.8501880>

- Khallouki Hajar, Bahaj Mohamed, Philippe Roose, **Sébastien Laborie**. SMPMA : Semantic multimodal Profile for Multimedia documents adaptation. The 5th International Workshop on Codes, Cryptography and Communication Systems (IWCCCS'14), Nov 2014, El Jadida, Morocco.
- Adel Alti, **Sébastien Laborie**, Philippe Roose. Context-aware Quality Adaptation Using Rich Explicit Constraints in E-health System. 8th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP'13), Dec 2013, Bayonne, France. pp.47-52. [⟨10.1109/SMAP.2013.11⟩](#)
- **Sébastien Laborie**, Hajar Khallouki, Philippe Roose. Bridging the gap between quantitative and qualitative constraints in profiles. 7th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP'12), Dec 2012, Luxembourg. pp.121-125.
- Cédric Dromzee, **Sébastien Laborie**, Philippe Roose. Towards an interoperable device profile containing rich user constraints. International Workshop on Interoperability of User Profiles in Multi-Application Web Environments (MultiA-Pro 2012), in conjunction with the World Wide Web Conference (WWW 2012), Apr 2012, Lyon, France. pp.977-986.
- **Sébastien Laborie**, Jérôme Euzenat, Nabil Layaïda. Semantic multimedia document adaptation with functional annotations. Proc. of the 4th international workshop on Semantic Media Adaptation and Personalization (SMAP'09), Dec 2009, San Sebastián, Spain. pp.44-49. [⟨10.1109/SMAP.2009.19⟩](#)
- **Sébastien Laborie**, Antoine Zimmermann. A framework for media adaptation using the web and the semantic web. Second International Workshop on Semantic Media Adaptation and Personalization (SMAP), Dec 2007, Londres, United Kingdom. pp.32-37, **Best Paper Award** [⟨10.1109/SMAP.2007.1⟩](#)
- **Sébastien Laborie**, Jérôme Euzenat, Nabil Layaïda. Multimedia document summarization based on a semantic adaptation framework. International workshop on Semantically aware document processing and indexing, ACM, May 2007, Montpellier, France. [⟨10.1145/1283880.1283896⟩](#)
- **Sébastien Laborie**, Jérôme Euzenat. Adapting the Hypermedia Structure in a Generic Multimedia Adaptation Framework. The First International Workshop on Semantic Media Adaptation and Personalization, Dec 2006, Athènes, Greece. pp.62-67, [⟨10.1109/SMAP.2006.6⟩](#)

Publications dans des ateliers nationaux (5)

- Jorge Andrés Larracochea, Philippe Roose, Sergio Ilarri, **Sébastien Laborie**, Yudit Cardinale and Mauricio Jacobo González. Modeling Energy Consumption in SOA : Requirements and Current Status. Atelier "Évolution des SI" INFORSID 2021, June 2021, Dijon, France.

- Cédric Dromzee, **Sébastien Laborie**, Philippe Roose. Décrire avec RDF des profils pour applications pervasives multimédias. 8èmes journées francophones Mobilité et Ubiquité - Ubimob, Jun 2012, Anglet, France. pp.1-12.
- **Sébastien Laborie**, Jérôme Euzenat, Nabil Layaïda. Adaptation spatio-temporelle et hypermédia de documents multimédia. Atelier Représentation et Raisonnement sur le Temps et l'Espace (RTE), Jun 2008, France. pp.1-13.
- **Sébastien Laborie**, Jérôme Euzenat, Nabil Layaïda. Adaptation sémantique de documents SMIL. Actes journées de travail interdisciplinaire sur autour des documents structurés, Oct 2006, Giens, France. pp.1-5.
- **Sébastien Laborie**, Jérôme Euzenat, Nabil Layaïda. Adapter temporellement un document SMIL. Actes atelier plate-forme AFIA 2005 sur Connaissance et document temporel, May 2005, Nice, France. pp.47-58.

Direction d'ouvrages (5)

- Tayeb Lemlouma, Yevgeniya Kovalchuk, **Sébastien Laborie**, Abderrezak Rachedi. Special Issue on e-health Pervasive Wireless Applications and Services (e-HPWAS'22). A special issue of Computers (ISSN 2073-431X), 12(2-3), 2023.
- Tayeb Lemlouma, **Sébastien Laborie**, Abderrezak Rachedi, Antonio Santos, Athanasios Vasilakos. Special Issue on Selected Papers from e-Health Pervasive Wireless Applications and Services 2017. Information, 10 (2), 52p, 2019.
- Max Chevalier, **Sébastien Laborie**, Cassia Trojahn dos Santos, Antoine Zimmermann. Impact des open et/ou linked data dans les SI. Revue des Sciences et Technologies de l'Information - Série ISI : Ingénierie des Systèmes d'Information, 23 (3-4), Lavoisier, 120p, 2019.
- **Sébastien Laborie**, Phivos Mylonas, Philippe Roose, Manolis Wallace. International Journal of Virtual Communities and Social Networking (IJVCSN), Special Issue on Social Media and Networks for Multimedia Content Management (Part 2). France. 7 (3), IGI Global, 2015, 1942-9010.
- **Sébastien Laborie**, Phivos Mylonas, Philippe Roose, Manolis Wallace. International Journal of Virtual Communities and Social Networking (IJVCSN), Special Issue on Social Media and Networks for Multimedia Content Management. France. 7 (2), IGI Global, 2015.

Actes (13)

- Tayeb Lemlouma, Yevgeniya Kovalchuk, Abderrezak Rachedi, **Sébastien Laborie**. eHPWAS'23 : Eleventh International workshop on e-health pervasive wireless applications and services 2023. In Proceeding of the 2023 IEEE 19th International

- Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2023).
- Tayeb Lemlouma, Yevgeniya Kovalchuk, Abderrezak Rachedi, **Sébastien Laborie**. eHPWAS'22 : Tenth International workshop on e-health pervasive wireless applications and services 2022. In Proceeding of the 2022 IEEE 18th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2022).
 - Tayeb Lemlouma, Yevgeniya Kovalchuk, Abderrezak Rachedi, **Sébastien Laborie**. eHPWAS'21 : Ninth International workshop on e-health pervasive wireless applications and services 2021. Oct 2021, Virtual. In Proceeding of the 2021 IEEE 17th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2021).
 - Tayeb Lemlouma, Yevgeniya Kovalchuk, Abderrezak Rachedi, **Sébastien Laborie**. eHPWAS'20 : Eighth International workshop on e-health pervasive wireless applications and services 2020. Oct 2020, Thessaloniki, Greece. In Proceeding of the 2020 IEEE 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2020).
 - Tayeb Lemlouma, Abderrezak Rachedi, **Sébastien Laborie**, Yevgeniya Kovalchuk. eHPWAS'19 : Eighth International workshop on e-health pervasive wireless applications and services 2019. Oct 2019, Barcelona, Spain. In Proceeding of the 2019 IEEE 15th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2019).
 - Tayeb Lemlouma, Abderrezak Rachedi, **Sébastien Laborie**, Yevgeniya Kovalchuk. IEEE eHPWAS'18 : Sixth international IEEE workshop on e-health pervasive wireless applications and services 2018. IEEE. Oct 2018, 2018, Proceeding of the 2018 IEEE 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob).
 - Tayeb Lemlouma, Abderrezak Rachedi, **Sébastien Laborie**, Santos António, Athanasios V. Vasilakos. IEEE eHPWAS'17 : Fifth international IEEE workshop on e-health pervasive wireless applications and services 2017. IEEE. France. 2017, Proceeding of the 2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob).
 - Tayeb Lemlouma, Abderrezak Rachedi, **Sébastien Laborie**, Santos António. IEEE eHPWAS'16 : Fourth international IEEE workshop on e-health pervasive wireless applications and services 2016. Oct 2016, New York, United States. In Proceeding of the 2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2016). [10.1109/WiMOB.2016.7763167](https://doi.org/10.1109/WiMOB.2016.7763167)
 - Tayeb Lemlouma, Abderrezak Rachedi, **Sébastien Laborie**, Antonio Santos. e-Health Pervasive Wireless Applications and Services Workshop (eHPWAS'15). WiMob 2015 : Wireless and Mobile Computing, Networking and Communications, Oct 2015, Abu Dhabi, United Arab Emirates. 2015.

- Tayeb Lemlouma, Abderrezak Rachedi, **Sébastien Laborie**, Antonio Santos. e-Health Pervasive Wireless Applications and Services Workshop (eHPWAS'14). Wi-Mob 2014 : Wireless and Mobile Computing, Networking and Communications, Oct 2014, Larnaca, Cyprus. IEEE, 2014.
- **Sébastien Laborie**, Philippe Roose. 8th International Workshop on Social and Semantic Media Adaptation and Personalization. Dec 2013, Bayonne, France. IEEE Computer Society, pp.109, 2013, 978-0-7695-5132-6. [⟨10.1109/SMAP2013.1⟩](#)
- Tayeb Lemlouma, Abderrezak Rachedi, **Sébastien Laborie**, Santos António, Philippe Roose. e-Health Pervasive Wireless Applications and Services Workshop. France. IEEE Press, pp.43, 2013, 978-1-4577-2014-7.
- **Sébastien Laborie**, Florence Le Ber. SIXIÈME ATELIER : Représentation et raisonnement sur le temps et l'espace (RTE 2011). 2011, Chambéry, France. 2011.

Bibliographie

- [1] ACHOURI, M., ADEL, A., DERDOUR, M., LABORIE, S., AND ROOSE, P. A New Two-Layered Architecture for Efficient Situations Management in Smart Environments. In *The 9th International ACM Conference on Management of Digital EcoSystems (MEDES 2017)* (Bangkok, France, Nov. 2017), ACM Press, pp. 44 – 59.
- [2] ACHOURI, M., ALTI, A., MAKHLOUF, D., LABORIE, S., AND ROOSE, P. Smart fog computing for efficient situations management for smart health environments. *Journal of Information and Communication Technology (JICT)* 17, 4 (Oct. 2018), 537–567.
- [3] ADEL, A., ROOSE, P., AND LABORIE, S. Multimedia Documents Adaptation Based on Semantic Multi-Partite Social Context-Aware Networks. *International Journal of Virtual Communities and Social Networking (IJVCSN)* 9, 3 (July 2017), 44 – 59.
- [4] ADIDA, B., BIRBECK, M., MCCARRON, S., AND HERMAN, I. *RDFa Core 1.1 - Third Edition*. W3C, <https://www.w3.org/TR/rdfa-core/>, 2015. W3C Recommendation.
- [5] ALTI, A., LABORIE, S., AND ROOSE, P. A framework for managing and optimizing the adaptation process qualities. In *The 7th IEEE GCC Conference and Exhibition (GCC'13)* (Doha, Qatar, Nov. 2013), Proceedings of the 7th IEEE GCC Conference and Exhibition (GCC'13), IEEE, pp. 125–130.
- [6] ALTI, A., LABORIE, S., AND ROOSE, P. Automatic Adaptation of Multimedia Documents. In *International Symposium on Frontiers in Ambient and Mobile Systems (FAMS-2013) in conjunction with the 4th International Conference on Ambient Systems, Networks and Technologies (ANT-2013)* (Halifax, Canada, June 2013), Elsevier.
- [7] ALTI, A., LABORIE, S., AND ROOSE, P. Context-aware Quality Adaptation Using Rich Explicit Constraints in E-health System. In *8th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP'13)* (Bayonne, France, Dec. 2013), P. R. Sébastien Laborie, Ed., IEEE Computer Society, pp. 47–52.
- [8] ALTI, A., LABORIE, S., AND ROOSE, P. Towards a Context-Aware Service and Quality Multimedia Adaptation for Healthcare Applications. In *The International Conference on Digital Information Processing, E-Business and Cloud Computing (DIPECC2013)* (Dubai, United Arab Emirates, Oct. 2013), pp. 113–122.

- [9] ALTI, A., LABORIE, S., AND ROOSE, P. Dynamic semantic-based adaptation of multimedia documents. *Transactions on emerging telecommunications technologies* 25, 2 (Feb. 2014), 239–258.
- [10] ALTI, A., LABORIE, S., AND ROOSE, P. Enrich the Expressiveness of Multimedia Document Adaptation Processes. In *Semantic Multimedia Analysis and Processing*, E. Spyrou, D. Lakovidis, and P. Mylonas, Eds., Digital Imaging and Computer Vision. CRC Press, June 2014, pp. 185–217.
- [11] ALTI, A., LABORIE, S., AND ROOSE, P. Enrichir l’expressivité des profils et des processus d’adaptation de contenus multimédias. In *Variabilité et adaptation de l’accès à l’information*, C. Soulé-Dupuy, Ed., Document Numérique. Lavoisier, July 2014, pp. 101–124.
- [12] ALTI, A., LABORIE, S., AND ROOSE, P. Semantic Context-aware Adaptation Platform Architecture. In *The 5th International Conference on Ambient Systems, Networks and Technologies (ANT-2014)* (Hasselt, Belgium, June 2014), vol. 32, Procedia Computer Science, pp. 959–964.
- [13] ALTI, A., LABORIE, S., AND ROOSE, P. A Community-Based Semantic Social Context-Aware Driven Adaptation for Multimedia Documents. *International Journal of Virtual Communities and Social Networking (IJVCSN)* 7, 2 (Apr. 2015), 31–49.
- [14] ALTI, A., LABORIE, S., AND ROOSE, P. Cloud Semantic-based Dynamic Multimodal Platform for Building mHealth Context-aware Services. In *Proceedings of the 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2015)* (Abu Dhabi, United Arab Emirates, Oct. 2015), IEEE, pp. 357–364.
- [15] ALTI, A., LABORIE, S., AND ROOSE, P. USSAP : Universal Smart Social Adaptation Platform. In *The 6th International Conference on Ambient Systems, Networks and Technologies (ANT-2015)* (London, United Kingdom, June 2015), vol. 52 of *Proceedings of The 6th International Conference on Ambient Systems, Networks and Technologies (ANT-2015)*, Elsevier, pp. 670–674.
- [16] ALTI, A., LABORIE, S., ROOSE, P., AND HAMZA, R. Dynamic Quality Management of Emergency Systems via Rich Explicit Constraints. In *The 5th International Conference on Web and Information Technologies (ICWIT’13)* (Hammamet, Tunisia, May 2013), pp. 531–540.
- [17] ALTI, A., LAKEHAL, A., LABORIE, S., AND ROOSE, P. Autonomic Semantic-Based Context-Aware Platform for Mobile Applications in Pervasive Environments. *Future Internet* 8, 4 (2016).
- [18] ALTI, A., ROOSE, P., SAFFIDINE, R., AND LABORIE, S. Towards an Automatic Adaptation of Heterogeneous Multimedia Mobile Applications. In *NOTERE / CFIP 2012* (Oct. 2012), Cépadués Editions, pp. 79–87.
- [19] ALTI, A., WASSIM, B., LABORIE, S., AND ROOSE, P. CSDMAP : Cloud Semantic-based Dynamic Multimodal Adaptation Platform. In *The 7th International*

- Conference on Management of computational and collective Intelligence in Digital EcoSystems (MEDES'15)* (Caraguatatuba/Sao Paulo, Brazil, Oct. 2015), The 7th International Conference on Management of computational and collective Intelligence in Digital EcoSystems (MEDES'15), ACM, pp. 161–168.
- [20] ANGSUCHOTMETEE, C., CHBEIR, R., AND CARDINALE, Y. MASN-Onto : An ontology-based approach for flexible event processing in Multimedia Sensor Networks. *Future Generation Computer Systems* 108 (2020), 1140–1158.
- [21] ARCAINI, P., RICCOBENE, E., AND SCANDURRA, P. Modeling and Analyzing MAPE-K Feedback Loops for Self-Adaptation. In *IEEE/ACM 10th International Symposium on Software Engineering for Adaptive and Self-Managing Systems* (2015), pp. 13–23.
- [22] ASADI, M. K., AND DUFOURD, J. Knowledge-Based and Semantic Adaptation of Multimedia Content. In *Knowledge-Based Media Analysis for Self-Adaptive and Agile Multi-Media, Proceedings of the European Workshop for the Integration of Knowledge, Semantics and Digital Media Technology, EWIMT* (may 2004).
- [23] BANERJEE, S., MISHRA, A., AND DASGUPTA, R. Semantic exploration of sensor data. In *Proceedings of the 5th International Workshop on Web-Scale Knowledge Representation Retrieval & Reasoning* (New York, NY, USA, 2014), Web-KR'14, Association for Computing Machinery, pp. 55—58.
- [24] BERHE, G., BRUNIE, L., AND PIERSON, J.-M. Planning-based multimedia adaptation services composition for pervasive computing. In *Fifth International Conference on Signal Image Technology and Internet Based Systems* (2009), pp. 326–331.
- [25] BERNERS-LEE, T., HENDLER, J., AND LASSILA, O. The Semantic Web. *Scientific American* 284, 5 (May 2001), 34–43.
- [26] BILASCO, I. M., GENSEL, J., AND VILLANOVA-OLIVER, M. STAMP : a model for generating adaptable multimedia presentations. *Multimedia Tools and Applications, Special Issue on Metadata and Adaptability in Web-based Information Systems* 25, 3 (2005), 361–375.
- [27] BOLL, S. MM4U - a framework for creating personalized multimedia content. In *Proceeding of the International Conference on Distributed Multimedia Systems* (2003).
- [28] BOUKADI, K., GHEDIRA, C., CHAARI, S., VINCENT, L., AND BATAINEH, E. CWSC4EC : How to employ context, Web service and Community in Enterprise collaboration. In *Proceedings of the 8th International Conference on New Technologies in Distributed Systems* (2008).
- [29] BRAY, T., PAOLI, J., SPERBERG-MCQUEEN, C. M., MALER, E., AND YERGEAU, F. *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. W3C, <https://www.w3.org/TR/xml/>, 2008. W3C Recommendation.
- [30] BRICKLEY, D., AND GUHA, R. *RDF Schema 1.1*. W3C, <https://www.w3.org/TR/rdf-schema/>, 2014. W3C Recommendation.

- [31] BUCHHOLZ, S., HAMANN, T., AND HUBSCH, G. Comprehensive structured context profiles (CSCP) : design and experiences. In *IEEE Annual Conference on Pervasive Computing and Communications Workshops* (2004), pp. 43–47.
- [32] BULTERMAN, D., JANSEN, J., CESAR, P., MULLENDER, S., HYPHE, E., DEMEGLIO, M., QUINT, J., KAWAMURA, H., WECK, D., PAÑEDA, X. G., MELENDI, D., CRUZ-LARA, S., HANCLIK, M., ZUCKER, D. F., AND MICHEL, T. *Synchronized Multimedia Integration Language (SMIL 3.0)*. W3C, <http://www.w3.org/TR/SMIL3/>, 2008. W3C Recommendation.
- [33] CALBIMONTE, J.-P., JEUNG, H., CORCHO, O., AND ABERER, K. Enabling query technologies for the semantic sensor web. *International Journal on Semantic Web and Information Systems* 8, 1 (2012), 43–63.
- [34] CAROTHERS, G., AND SEABOURNE, A. *RDF 1.1 N-Triples*. W3C, <http://www.w3.org/TR/n-triples/>, 2014. W3C Recommendation.
- [35] CHARBEL, N. *Semantic Representation of a Heterogeneous Document Corpus for an Innovative Information Retrieval Model : Application to the Construction Industry*. PhD thesis, Université de Pau et des Pays de l'Adour, Dec. 2018.
- [36] CHAUMIER, J. Les ontologies : Antécédents, aspects techniques et limites. *Revue Documentaliste-Sciences de l'Information* 44, 1 (2007), 81–83.
- [37] CYGANIAK, R., WOOD, D., AND LANTHALER, M. *RDF 1.1 Concepts and Abstract Syntax*. W3C, <https://www.w3.org/TR/rdf11-concepts/>, 2014. W3C Recommendation.
- [38] DAVID, J., EUZENAT, J., SCHARFFE, F., AND TROJAHN DOS SANTOS, C. The Alignment API 4.0. *Semantic Web* 2, 1 (jan 2011), 3—10.
- [39] DE BRA, P., SMITS, D., AND STASH, N. The Design of AHA! In *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia* (New York, NY, USA, 2006), Association for Computing Machinery, pp. 171—195.
- [40] DONGO, I. *Anonymisation de descriptions RDF*. PhD thesis, Université de Pau et des Pays de l'Adour, Dec. 2017.
- [41] DROMZEE, C., LABORIE, S., AND ROOSE, P. Décrire avec RDF des profils pour applications pervasives multimédias. In *8èmes journées francophones Mobilité et Ubiquité - Ubimob* (June 2012), Cépaduès, pp. 1–12.
- [42] DROMZEE, C., LABORIE, S., AND ROOSE, P. Profil générique sémantique pour l'adaptation de documents multimédias. In *30ème congrès INFORSID (INFormatique des ORganisations et Systèmes d'Information et de Décision) - INFORSID'12* (Montpellier, France, May 2012), p. 16.
- [43] DROMZÉE, C., LABORIE, S., AND ROOSE, P. A Semantic Generic Profile for Multimedia Documents Adaptation. In *Intelligent Multimedia Technologies for Networking Applications : Techniques and Tools*, D. Kanellopoulos, Ed. IGI Global, Jan. 2013, pp. 225–246.

- [44] DROMZEE, C., LABORIE, S., AND ROOSE, P. Adapting Multimedia Documents with a Semantic Generic Profile. In *Intelligent Multimedia Technologies for Networking Applications : Techniques and Tools*. IGI Global, Jan. 2013, pp. 225–246.
- [45] ETCHEVERRY, P., LABORIE, S., MARQUESUZAÀ, C., NODENOT, T., AND LUONG, T. N. Conception d’applications web géographiques guidée par les contenus et les usages : cadre méthodologique et opérationnalisation avec l’environnement WINDMash. *Journal d’Interaction Personne-Système* 3, 1 (July 2015), 1–42.
- [46] EUZENAT, J., AND SHVAIKO, P. *Ontology Matching, Second Edition*. Springer, 2013.
- [47] FARIA, C., SERRA, I., AND GIRARDI, R. A domain-independent process for automatic ontology population from text. *Science of Computer Programming* 95 (2014), 26–43. Special Issue on Systems Development by Means of Semantic Technologies.
- [48] FARRELL, J., AND LAUSEN, H. *Semantic Annotations for WSDL and XML Schema*. W3C, <https://www.w3.org/TR/sawSDL/>, 2007. W3C Recommendation.
- [49] FENSEL, D., KERRIGAN, M., AND ZAREMBA, M., Eds. *WSMO and WSMML*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 43–65.
- [50] FERNÁNDEZ, J. D., MARTÍNEZ-PRieto, M. A., GUTIÉRREZ, C., POLLERES, A., AND ARIAS, M. Binary RDF representation for publication and exchange (HDT). *Journal of Web Semantics* 19 (2013), 22–41.
- [51] FREIRE, J., KUMAR, B., AND LIEUWEN, D. WebViews : Accessing Personalized Web Content and Services. In *Proceedings of the 10th International Conference on World Wide Web (WWW’01)* (New York, NY, USA, 2001), ACM, pp. 576—586.
- [52] GANDON, F., AND SCHREIBER, G. *RDF/XML 1.1 XML Syntax*. W3C, <http://www.w3.org/TR/rdf-syntax-grammar/>, 2014. W3C Recommendation.
- [53] GEURTS, J., VAN OSSENBRUGGEN, J., AND HARDMAN, L. Application-specific constraints for multimedia presentation generation. In *Proceedings of the International Conference on Multimedia Modeling* (2001), pp. 247–266.
- [54] GUARINO, N., OBERLE, D., AND STAAB, S. *What Is an Ontology?* Handbook on Ontologies. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 1–17.
- [55] GUENNOUNI, N. *ISEE : un système pour l’explication des événements dans les environnements hybrides - mise en œuvre sur des données de capteurs associées à des corpus documentaires*. PhD thesis, Université de Pau et des Pays de l’Adour, June 2022.
- [56] GUENNOUNI, N., LABORIE, S., SALLABERRY, C., CHBEIR, R., AND MANSOUR, E. ISEEapp : An Event Explanation Prototype bridging the gap between sensor network and document corpora data. In *18th International Conference on Intelligent Environments (IE 2022)* (Biarritz, France, June 2022), Proc. of the 18th International Conference on Intelligent Environments (IE 2022), IEEE.
- [57] GUENNOUNI, N., SALLABERRY, C., LABORIE, S., AND CHBEIR, R. A Novel Framework for Event Interpretation in a Heterogeneous Information System. In *MEDES*

- '20 : 12th International Conference on Management of Digital EcoSystems (Virtual Event United Arab Emirates, France, Nov. 2020), ACM, pp. 140–148.
- [58] GUENNOUNI, N., SALLABERRY, C., LABORIE, S., CHBEIR, R., AND MANSOUR, E. ISEE : A heterogeneous information system for event explainability in smart connected environments. *Internet of Things* (Sept. 2021).
- [59] GÓMEZ-PÉREZ, A., AND CORCHO, O. Ontology languages for the semantic web. *IEEE Intelligent Systems* 17, 1 (2002), 54–60.
- [60] HAI, Q. P., LABORIE, S., AND ROOSE, P. On-the-fly Multimedia Document Adaptation Architecture. In *Pervasive Environments (SUPE'2012), in conjunction with the 9th International Conference on Mobile Web Information Systems (MobiWIS'12)* (Aug. 2012), Procedia Computer Science, Elsevier Science, pp. 1–6.
- [61] HAJAR, K., MOHAMED, B., ROOSE, P., AND LABORIE, S. SMPMA : Semantic multimodal Profile for Multimedia documents adaptation. In *The 5th International Workshop on Codes, Cryptography and Communication Systems (IWCCCS'14)* (El Jadida, Morocco, Nov. 2014), IEEE.
- [62] HALLER, A., JANOWICZ, K., COX, S., PHUOC, D. L., TAYLOR, K., AND LEFRANÇOIS, M. *Semantic Sensor Network Ontology*. W3C, <https://www.w3.org/TR/vocab-ssn/>, 2017. W3C Recommendation.
- [63] HAMBORG, F., LACHNIT, S., SCHUBOTZ, M., HEPP, T., AND GIPP, B. Giveme5W : Main Event Retrieval from News Articles by Extraction of the Five Journalistic W Questions. In *Transforming Digital Worlds* (2018), Springer International Publishing, pp. 356–366.
- [64] HAYES, J., AND GUTIERREZ, C. Bipartite Graphs as Intermediate Model for RDF. In *The Semantic Web – ISWC 2004* (2004), Springer Berlin Heidelberg, pp. 47–61.
- [65] HE, J., GAO, T., HAO, W., YEN, I.-L., AND BASTANI, F. A flexible content adaptation system using a rule-based approach. *IEEE Transactions on Knowledge and Data Engineering* 19, 1 (2007), 127–140.
- [66] HERRERA, J.-M., HOGAN, A., AND KÄFER, T. BTC-2019 : The 2019 Billion Triple Challenge Dataset. In *The Semantic Web – ISWC 2019* (2019), Springer International Publishing, pp. 163–180.
- [67] HITZLER, P. A Review of the Semantic Web Field. *Communications of the ACM* 64, 2 (February 2021), 76–83.
- [68] HORROCKS, I., PATEL-SCHNEIDER, P. F., BOLEY, H., TABET, S., GROSOE, B., AND DEAN, M. *SWRL : A Semantic Web Rule Language Combining OWL and RuleML*. W3C, <https://www.w3.org/Submission/SWRL/>, 2004. W3C Member Submission 21 May 2004.
- [69] KLYNE, G., AND CARROLL, J. J. *Resource Description Framework (RDF) : Concepts and Abstract Syntax*. W3C, <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>, 2004. W3C Recommendation.

- [70] KLYNE, G., REYNOLDS, F., WOODROW, C., OHTO, H., HJELM, J., BUTLER, M. H., AND TRAN, L. *Composite Capability/Preference Profiles (CC/PP) : Structure and Vocabularies 1.0*. W3C, <https://www.w3.org/TR/CCPP-struct-vocab/>, 2004. W3C Recommendation.
- [71] LABORIE, S. *Adaptation sémantique de documents multimédia*. Thèse en informatique, Université Joseph Fourier, Grenoble, 2008.
- [72] LABORIE, S., KHALLOUKI, H., AND ROOSE, P. Bridging the gap between quantitative and qualitative constraints in profiles. In *SMAP 2012* (Luxembourg, Dec. 2012), IEEE, Ed., IEEE Computer Society, pp. 121–125.
- [73] LAKEHAL, A., ALTI, A., LABORIE, S., AND ROOSE, P. Ontology-based Context-aware Recommendation Approach For Dynamic Situations Enrichment. In *13th International Workshop on Semantic and Social Media Adaptation and Personalization* (Zaragoza, Spain, Sept. 2018), IEEE, pp. 81–86.
- [74] LAKEHAL, A., ALTI, A., LABORIE, S., AND ROOSE, P. A Semantic Agile Approach for Reconfigurable Distributed Applications in Pervasive Environments. *International Journal of Ambient Computing and Intelligence* 11, 2 (Apr. 2020), 48–67.
- [75] LAMOLLE, M., GOMEZ, J., AND EXPOSITO, E. MODA : une architecture multimédia dirigée par les ontologies pour des systèmes multimédia en réseau. *Revue des Nouvelles Technologies de l'Information 4e Conférence francophone sur les Architectures Logicielles, RNTI-L-5* (2010), 92–104.
- [76] LARRACOECHEA, J. A., ROOSE, P., ILARRI, S., CARDINALE, Y., AND LABORIE, S. Profilage de services pour la gestion de l'énergie dans les architectures orientées services. In *INFORSID 2022* (Dijon, France, May 2022).
- [77] LARRACOECHEA, J. A., ROOSE, P., ILARRI, S., CARDINALE, Y., LABORIE, S., AND VARA, O. P. Behavior-Based Consumption Profiles for the Approximation of the Energy Consumption of Services. In *Proc. of the 30th International Conference on Information Systems Development (ISD 2022)* (Cluj-Napoca, Romania, Aug. 2022).
- [78] LARRACOECHEA, J. A., ROOSE, P., ILARRI, S., LABORIE, S., AND CARDINALE, Y. RADIANCE : A tool for software behavior design and energy consumption categorization. In *INFORSID 2023* (La Rochelle, France, May 2023). Demo paper.
- [79] LARRACOECHEA, J. A., ROOSE, P., ILLARI, S., CARDINALE, Y., LABORIE, S., AND GONZALEZ, M. J. Towards Services Profiling for Energy Management in Service-Oriented Architectures. In *17th International Conference on Web Information Systems and Technologies (WEBIST)* (Online Streaming, France, Oct. 2021). Best Student Paper Award.
- [80] LARRACOECHEA, J. A., ROOSE, P., ILLARI, S., LABORIE, S., CARDINALE, Y., AND GONZALEZ, M. J. Modeling Energy Consumption in SOA : Requirements and Current Status. In *Atelier "Évolution des SI" (INFORSID 2021)* (Dijon, France, June 2021).

- [81] LEI, Z., AND GEORGANAS, N. Context-based media adaptation in pervasive computing. In *Canadian Conference on Electrical and Computer Engineering* (2001), vol. 2, pp. 913–918.
- [82] LEMLOUMA, T., LABORIE, S., ROOSE, P., RACHEDI, A., AND ABDELAZIZ, K. mHealth Contents and Services Delivery and Adaptation Challenges for Smart Environments. In *mHealth Multidisciplinary Verticals*, S. Adibi, Ed. CRC Press/-Taylor & Francis, Nov. 2014, pp. 295–314. Chapitre 17.
- [83] LEMLOUMA, T., AND LAYAÏDA, N. The negotiation of multimedia content services in heterogeneous environments. In *In Proceedings of the 8th International Conference on Multimedia Modeling* (2001), pp. 187–206.
- [84] LEMLOUMA, T., AND LAYAÏDA, N. Adapted content delivery for different contexts. In *Symposium on Applications and the Internet* (2003), pp. 190–197.
- [85] LEOPOLD, K., JANNACH, D., AND HELLWAGNER, H. A knowledge and component based multimedia adaptation framework. In *IEEE Sixth International Symposium on Multimedia Software Engineering* (2004), pp. 10–17.
- [86] LONGLEY, D., AND SPORNY, M. *RDF Graph Normalization*. <https://json-ld.org/spec/ED/rdf-graph-normalization/20111016/>, 2011.
- [87] LUONG, T. N. *Modélisation centrée utilisateur final appliquée à la conception d'applications interactives en géographie : une démarche basée sur les contenus et les usages*. PhD thesis, Université de Pau et des Pays de l'Adour, Dec. 2012.
- [88] LUONG, T. N., LABORIE, S., AND NODENOT, T. A Framework with Tools for Designing Web-based Geographic Applications. In *The 11th ACM Symposium on Document Engineering* (Sept. 2011), ACM, Ed., DocEng '11, pp. 33–42.
- [89] LUONG, T. N., MARQUESUZAÀ, C., ETCHEVERRY, P., NODENOT, T., AND LABORIE, S. Facilitating the Design/Evaluation Process of Web-based Geographic Applications : a Case Study with WINDMash. In *The 2nd International Conference on Future Data and Security Engineering 2015 (FDSE)* (Ho Chi Minh City, Vietnam, Nov. 2015), vol. 9446, Springer LNCS, pp. 259–271.
- [90] MANSOUR, E., CHBEIR, R., AND ARNOULD, P. HSSN : An Ontology for Hybrid Semantic Sensor Networks. In *Proceedings of the 23rd International Database Applications & Engineering Symposium* (New York, NY, USA, 2019), IDEAS'19, Association for Computing Machinery.
- [91] MANSOURI, K., ALTI, A., ROOSE, P., AND LABORIE, S. Dynamic semantic-based green bio-inspired approach for optimizing energy and cloud services qualities. *Transactions on emerging telecommunications technologies* 29, 5 (May 2018).
- [92] MARQUESUZAÀ, C., ETCHEVERRY, P., LABORIE, S., NODENOT, T., AND LUONG, T. N. User-Centered Design of Geographic Interactive Applications : From High-Level Specification to Code Generation, from Prototypes to Better Specifications. *Transactions on Large-Scale Data- and Knowledge-Centered Systems LNCS 10140* (Dec. 2016), 1–29.

- [93] MCGUINNESS, D. L., AND VAN HARMELEN, F. *OWL Web Ontology Language*. W3C, <http://www.w3.org/TR/owl-features/>, 2004. W3C Recommendation.
- [94] MISHRA, A., AND JAIN, S. K. A survey on question answering systems with classification. *Journal of King Saud University - Computer and Information Sciences* 28, 3 (2016), 345–361.
- [95] MOUNIR, A., ADEL, A., MAKHLOUF, D., LABORIE, S., AND ROOSE, P. A New Two-Level Clustering Approach for Situations Management in Distributed Smart Environments. *International Journal of Ambient Computing and Intelligence* 10, 2 (Apr. 2019), 91–111.
- [96] NOËL, D., VILLANOVA-OLIVER, M., GENSEL, J., AND LE QUÉAU, P. Design patterns for modelling life trajectories in the semantic web. In *Web and Wireless Geographical Information Systems* (2017), D. Brosset, C. Claramunt, X. Li, and T. Wang, Eds., Springer International Publishing, pp. 51–65.
- [97] NOUREDDINE, H., RAY, C., AND CLARAMUNT, C. A hierarchical indoor and outdoor model for semantic trajectories. *Transactions in GIS* 26, 1 (2022), 214–235.
- [98] PAPAEFSTATHIOU, I., AND HATZOPOULOS, A. *Heterogeneous Cyber Physical Systems of Systems*. River Publishers, 2021.
- [99] PIETERS, S. *HTML5 Differences from HTML4*. W3C, <https://www.w3.org/TR/html5-diff/>, 2014. W3C Working Group Note 9 December 2014.
- [100] PRUD’HOMMEAUX, E., AND CAROTHERS, G. *RDF 1.1 Turtle : Terse RDF Triple Language*. W3C, <http://www.w3.org/TR/turtle/>, 2014. W3C Recommendation.
- [101] PUTTASWAMY, K., NANDAGOPAL, T., AND KODIALAM, M. Frugal storage for cloud file systems. In *EuroSys’12 : Proceedings of the 7th ACM european conference on Computer Systems* (Apr. 2012), pp. 71–84.
- [102] RAVAT, F., AND ZHAO, Y. Metadata Management for Data Lakes. In *23rd East-European Conference on Advances in Databases and Information Systems (ADBIS 2019)* (Sept. 2019), pp. 37–44.
- [103] RIBEIRO DE ALMEIDA, D., DE SOUZA BAPTISTA, C., AND DE ANDRADE, F. G. Similarity search on semantic trajectories using text processing. *ISPRS International Journal of Geo-Information* 11, 7 (2022).
- [104] ROOSE, P., ILARRI, S., LARRACOECHEA, J. A., CARDINALE, Y., AND LABORIE, S. Towards an Integrated Full-Stack Green Software Development Methodology. In *Proc. of the 29th International Conference on Information Systems Development (ISD 2021)* (Valencia, Spain, Sept. 2021).
- [105] ROUSSEAU, F., GARCÍA-MARCÍAS, J. A., VALDENI DE LIMA, J., AND DUDA, A. User Adaptable Multimedia Presentations for the World Wide Web. *International Journal of Computer and Telecommunications Networking* 31, 11-16 (may 1999), 1273—1290.

- [106] SABOU, M., BIFFL, S., EINFALT, A., KRAMMER, L., KASTNER, W., AND EKAPUTRA, F. Semantics for Cyber-Physical Systems : A cross-domain perspective. *Semantic Web 11* (December 2019), 1–10.
- [107] SAIGHI, A. *Qualité de service pour les adaptations multimédia dans un environnement mobile*. PhD thesis, Université Badji Mokhtar - Annaba & Université de Pau et des Pays de l'Adour, 2018.
- [108] SAIGHI, A., LABOUDI, Z., ROOSE, P., LABORIE, S., AND GHOUALMI-ZINE, N. On Using Multiple Disabilities Profiles to Adapt Multimedia Documents. *International Journal of Information Technology and Web Engineering* 15, 3 (July 2020), 34–60.
- [109] SAIGHI, A., ROOSE, P., GHOUALMI, N., LABORIE, S., AND LABOUDI, Z. Ontological Reasoning for the Personalization of Multimedia Adaptation Services in Smart Home Domain. In *Proc. of The Second International Symposium on Informatics and its Applications (ISIA 2016)* (M'sila, Algeria, Nov. 2016).
- [110] SAIGHI, A., ROOSE, P., GHOUALMI, N., LABORIE, S., AND LABOUDI, Z. HaMA : A Handicap-based Architecture for Multimedia Document Adaptation. *International Journal of Multimedia Data Engineering and Management (IJMDEM)* 8, 3 (July 2017), 55–96.
- [111] SAIGHI, A., ROOSE, P., GHOULAMI, N., LABORIE, S., AND LABOUDI, Z. Multimedia Presentation Adaptation Architecture Assisted by Ontological Reasoning. In *The 2nd International Conference on Pattern Analysis and Intelligent Systems (PAIS 2016)* (Khenchela, Algeria, Nov. 2016).
- [112] SALTON, G., FOX, E. A., AND WU, H. Extended boolean information retrieval. *Communications of the ACM* 26, 11 (nov 1983), 1022—1036.
- [113] SAUERMAN, L., AND CYGANIAK, R. *Cool URIs for the Semantic Web*. W3C, <https://www.w3.org/TR/cooluris/>, 2008. W3C Interest Group Note 03 December 2008.
- [114] SAWADOGO, P., AND DARMONT, J. On data lake architectures and metadata management. *Journal of Intelligent Information Systems* 56 (2021), 97–120.
- [115] SCHATZKI, T. R. *The Trajectories of a Life*. Springer International Publishing, 2022, pp. 19–34.
- [116] SCHREIBER, G., AND RAIMOND, Y. *RDF 1.1 Primer*. W3C, <https://www.w3.org/TR/rdf11-primer/>, 2014. W3C Working Group Note 24 June 2014.
- [117] SHARMA, A., AND KUMAR, S. Semantic Web-Based Information Retrieval Models : A Systematic Survey. In *Data Science and Analytics* (Singapore, 2020), U. Batra, N. R. Roy, and B. Panda, Eds., Springer Singapore, pp. 204–222.
- [118] SHETH, A., HENSON, C., AND SAHOO, S. S. Semantic sensor web. *IEEE Internet Computing* 12, 4 (2008), 78–83.
- [119] SLIMANI, S., ALTI, A., LABORIE, S., AND ROOSE, P. An Advanced Adaptation Approach for Pervasive Applications. In *The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016)* (Madrid, Spain, May 2016),

- Proc. of the 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016).
- [120] SOBOROFF, I. The better cross-language datasets. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York, NY, USA, 2023), SIGIR'23, Association for Computing Machinery, pp. 3047—3053.
- [121] SPORNY, M., KELLOGG, G., AND LANTHALER, M. *JSON-LD 1.1 : A JSON-based Serialization for Linked Data*. W3C, <http://www.w3.org/TR/json-ld/>, 2014. W3C Recommendation.
- [122] STAAB, S., SCHERP, A., ARNDT, R., TRONCY, R., GRZEGORZEK, M., SAATHOFF, C., SCHENK, S., AND HARDMAN, L. *Semantic Multimedia*. Springer Berlin Heidelberg, 2008, pp. 125–170.
- [123] THE W3C SPARQL WORKING GROUP. *SPARQL 1.1 Overview*. W3C, <https://www.w3.org/TR/sparql11-overview/>, 2013. W3C Recommendation.
- [124] THOMA, M., MEYER, S., SPERNER, K., MEISSNER, S., AND BRAUN, T. On IoT-services : Survey, Classification and Enterprise Integration. In *2012 IEEE International Conference on Green Computing and Communications* (2012), pp. 257–260.
- [125] TICONA-HERRERA, R., TEKLI, J., CHBEIR, R., LABORIE, S., DONGO, I., AND GUZMAN, R. Toward RDF Normalization. In *Proceedings of the International Conference on Conceptual Modeling (ER 2015)* (2015), LNCS 9381, Springer International Publishing, pp. 261–275.
- [126] TICONA HERRERA, R. P. *Towards RDF Normalization*. PhD thesis, Université de Pau et des Pays de l'Adour, July 2016.
- [127] USCHOLD, M., AND GRÜNINGER, M. Ontologies : Principles, methods and applications. *Knowledge Engineering Review* 11, 2 (1996), 93–136.
- [128] VILLANOVA-OLIVER, M., GENSEL, J., MARTIN, H., AND LOPEZ-VELASCO, C. Adaptabilité à l'utilisateur dans le contexte des services web. *Revue des Nouvelles Technologies de l'Information Extraction des connaissances : Etat et perspectives (Ateliers de la conférence EGC'2005)*, RNTI-E-5 (2005), 153–158.
- [129] VILLCA-ROCHA, A., ZHENG, M., DUAN, C., AND WANG, H. Towards semantic search in building sensor data. In *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (New York, NY, USA, 2021), BuildSys '21, Association for Computing Machinery, pp. 164—167.
- [130] VRANDEČIĆ, D., DENGLER, F., RUDOLPH, S., AND ERDMANN, M. RDF syntax normalization using XML validation. In *Proceedings of the Workshop Semantics for the Rest of Us at ISWC 2009* (October 2009), L. Kagal, O. Lassila, and T. Finin, Eds., CEUR-WS.
- [131] W3C OWL WORKING GROUP. *OWL 2 Web Ontology Language*. W3C, <http://www.w3.org/TR/owl-overview>, 2012. W3C Recommendation.

- [132] WANG, W., ZHAO, D., ZOU, L., WANG, D., AND ZHENG, W. Extracting 5W1H Event Semantic Elements from Chinese Online News. In *Web-Age Information Management* (Berlin, Heidelberg, 2010), Springer Berlin Heidelberg, pp. 644–655.
- [133] WEI, W., AND BARNAGHI, P. Semantic annotation and reasoning for sensor data. In *Smart Sensing and Context (EuroSSC 2009). Lecture Notes in Computer Science, vol 5741*. (2009), Springer Berlin Heidelberg, pp. 66–76.
- [134] WIRELESS APPLICATION PROTOCOL FORUM. *WAG UAProf*. <https://www.openmobilealliance.org/tech/affiliates/wap/wap-248-uaprof-20011020-a.pdf>, 2001.
- [135] WU, H., CHELMIS, C., SORATHIA, V., ZHANG, Y., PATRI, O. P., AND PRASANNA, V. K. Enriching employee ontology for enterprises with knowledge discovery from social networks. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (New York, NY, USA, 2013), ASONAM '13, Association for Computing Machinery, pp. 1315—1322.

Table des figures

1.1	Mon implication dans des travaux de thèse depuis 2010	5
2.1	La pyramide du Web Sémantique	9
2.2	Un triplet RDF	10
2.3	Un exemple de description RDF	11
2.4	Un exemple d'un réseau de données liées	12
2.5	Un exemple de schéma RDF	14
2.6	Des exemples de requêtes SPARQL et leurs patrons de graphe	16
3.1	Des descriptions RDF au sujet d'une même ressource	21
3.2	Des exemples de sérialisation en RDF/XML	23
3.3	Architecture de notre proposition de normalisation de descriptions RDF .	27
3.4	Performance de notre processus de normalisation	28
4.1	Un exemple d'hétérogénéité de dispositifs mobiles.	32
4.2	Un exemple d'adaptation d'un document multimédia.	33
4.3	Un exemple de profil CC/PP sérialisé en RDF/XML	36
4.4	Une autre vision du profil à base de services	39
4.5	Comblent le fossé entre valeurs quantitatives et qualitatives des profils . .	40
4.6	Une ontologie pour guider le processus d'adaptation.	42
4.7	Un processus d'adaptation exploitant le mécanisme d'inférence.	43
4.8	Déterminer une chaîne de services d'adaptation sur Android.	45
5.1	Digitalisation et investissement dans les environnements connectés . . .	49
5.2	Recherche d'explications dans les environnements hybrides	50
5.3	Vue d'ensemble de notre approche ISEE	53
5.4	Un exemple simplifié illustrant concepts et instances	54
5.5	Notre proposition ISEE à base de filtrage successifs	55
5.6	ISEEapp : une explication d'un événement déclenché	59
5.7	Résultats de l'expérimentation d'ISEEapp en IUT	59
6.1	Le Web Sémantique : un pont entre données massives et décisions	65
6.2	Estimation de la demande mondiale en électricité pour les centres de données	66

TABLE DES FIGURES

7.1 Mon parcours d'enseignant-chercheur à l'UPPA. 69

Liste des tableaux

2.1	Un exemple de primitives des langages RDF & RDFS	13
2.2	Une vue générale de mes contributions dans la cadre du Web Sémantique	18
3.1	Les besoins en faveur d'une normalisation de descriptions RDF	26
4.1	Un exemple de règle SWRL.	44
5.1	Observations collectées par les 5 types de capteurs	58
7.1	Type et classement de mes publications.	72
7.2	Thèses codirigées et co-encadrées.	74
7.3	Projets de recherche menés.	74