

UNIVERSITÉ NICE SOPHIA ANTIPOLIS
ÉCOLE DOCTORALE STIC
SCIENCES ET TECHNOLOGIES DE L'INFORMATION
ET DE LA COMMUNICATION

THÈSE

pour obtenir le titre de

Docteur en Sciences

de l'Université Nice Sophia Antipolis (FRANCE) & de
l'Université Gaston Berger - Saint-Louis (SENEGAL)

Mention : INFORMATIQUE

Présentée et soutenue par

Papa Fary DIALLO

Aspects Socioculturels et Temporels dans les Ontologies pour les Communautés Virtuelles

préparée à l'INRIA Sophia Antipolis, Projet WIMMICS

soutenue le 16 Septembre 2016

Jury :

<i>Co-directeur de thèse :</i>	Isabell MIRBEL Moussa LO	Maître de Conférences HDR UNS Professeur des Universités UGB
<i>Co-encadrement de thèse :</i>	Olivier CORBY Seydina NDIAYE	Chercheur CR INRIA Chargé d'enseignement UGB
<i>Président :</i>	Andrea TETTAMANZI	Professeur des Universités UNS
<i>Rapporteurs :</i>	Ollivier HAEMMERLÉ Sylvie DESPRES	Professeur des Universités Toulouse Professeur des Universités Paris 13
<i>Examineur :</i>	Béatrice MARKHOFF	Maître de conférences HDR Tours

Remerciements

Je remercie Pr Moussa Lo et Dr Seydina Ndiaye de l'Université Gaston Berger de Saint-Louis d'avoir initié le projet WestAfricapedia et d'avoir eu confiance en moi et de m'avoir permis de participer à ce projet. Je leur remercie également pour leur encadrement durant ces années.

Mes remerciements vont particulièrement aussi à Isabelle Mirbel et Olivier Corby pour l'aide précieuse qu'ils m'ont apporté tout au long de cette thèse. Sans leur aide, le travail présenté dans cette thèse n'existerait pas. Je leur témoigne toute ma gratitude, merci.

Je remercie également Sylvie Despres, Professeur à l'Université Paris 13 et Olivier Haemmerlé, Professeur à l'Université Toulouse Jean Jaurès, qui m'ont fait l'honneur d'être les rapporteurs de cette thèse. J'adresse mes remerciements à Béatrice Markhoff maître de conférences à l'Université de Tours d'avoir accepté d'être examinatrice de cette thèse et Andrea Tettamanzi, Professeur à l'Université de Nice, d'avoir accepté de présider le jury de cette thèse.

Mes remerciements vont également aux membres de l'équipe WIMMICS, en particulier à Fabien Gandon et Alain Giboin pour leur collaboration dans ce travail et à travers eux tous les permanents de l'équipe. Mention spéciale aux doctorants de l'équipe, anciens comme actuels, en particuliers Oumy Seye et famille, Émilie Palagi, Amel BenOthmane, Amosse Édouard, Abdoul Karim Macina, Maxime Lefrançois, Rakeb Hasan et mes anciens collègues de bureau Christophe Desclaux et Julien Cojan. Un remerciement spécial aussi à Christine Foggia pour son assistance dans les démarches administratives.

Mes remerciements vont aussi à mes amis de SANAR, en particulier, Mansour Seck, Moustapha Diakhaté, Mamadou Djigo, Ciré Sall Ba, Amadou Ndiaye Mbaye et famille, Moustapha Kane et Baïdy Diallo.

Je remercie aussi les institutions qui ont financé ce travail : Ministère de l'enseignement supérieur du Sénégal, l'AUF, l'état Français à travers le service culturel de l'ambassade de France à Dakar et le LIRIMA de l'INRIA.

Last but not the least, je remercie toute ma famille, en particulier, mes parents et mes deux grands frères qui n'ont ménagé aucun effort pour ma réussite.

Résumé

Les travaux que nous présentons dans cette thèse concernent la modélisation des aspects socioculturels et temporels pour permettre aux communautés sénégalaises de partager et de co-construire leur connaissances socioculturelles. En effet, avec la mondialisation la nouvelle génération africaine et particulièrement sénégalaise a de moins en moins de connaissances sur les aspects socioculturels de leur environnement. Ainsi pour rafraîchir la mémoire de nos concitoyens et redonner vie aux nombreux récits qui accompagnent la création et la vie au quotidien des différents terroirs sénégalais, nous avons initié la mise en place d'une application en ligne pour permettre aux communautés sénégalaises de partager et de co-construire leur patrimoine socioculturel.

Nos propositions s'appuient sur les technologies du Web social et du Web sémantique. En effet, le Web social propose un cadre qui permet à tout utilisateur de participer à la création de contenu dans le Web. De ce fait, le Web social facilite la co-construction de corpus. Les technologies du Web sémantique permettent d'annoter et de raisonner sur ce corpus. Le Web sémantique rend accessible les connaissances d'un domaine aux agents logiciels pour une meilleure recherche d'informations. La combinaison de ces deux technologies permet aux communautés sénégalaises de partager et de co-construire leur patrimoine culturel dans un environnement collaboratif et sémantique.

Nos contributions consistent à (i) proposer des ontologies pour annoter des ressources socioculturelles et (ii) proposer un cadre permettant aux communautés de partager et de co-construire leur connaissances. Les ontologies représentent le socle du Web sémantique et permettent de caractériser un domaine en décrivant les concepts fondamentaux et les relations entre eux. Ainsi, nous avons défini deux ontologies : 1) une ontologie socioculturelle reposant sur la théorie historico-culturelle de l'activité et 2) une ontologie temporelle pour annoter les ressources socioculturelles. Nous avons aussi proposé un type de communauté virtuelle appelé communauté de co-élaboration de connaissances socioculturelle qui est une adaptation des communautés de co-élaboration de connaissances dans le domaine culturel.

Dans ce manuscrit, nous présentons (1) un état de l'art, d'une part, sur la capitalisation de connaissances qui couvre à la fois la théorie historico-culturelle de l'activité et les typologies des communautés virtuelles et, d'autres part, sur l'annotation temporelle des ressources du Web sémantique, (2) une méthodologie de construction d'une ontologie socioculturelle et les caractéristiques des communautés de co-élaboration de connaissances culturelles, (3) une ontologie pour l'annotation temporelle des ressources du Web sémantique particulièrement les expressions humaines du temps et (4) la mise en œuvre d'un système qui intègre les contributions scientifiques de ce travail et qui permet aux communautés sénégalaises de partager et de co-construire leur patrimoine culturel.

Mots clés : Web sémantique, Web social, Ontologies, Communautés Virtuelles, Société, Culture, Annotation temporelle, Co-construction et Partage.

Abstract

This thesis is set in a research effort that aims to model sociocultural and temporal aspects to allow Senegalese communities to share and to co-construct their sociocultural knowledge. Indeed, with the globalization it is very common to meet African youth and particularly Senegalese youth knowing more about the geography of the West than their own countries. Thus, to refresh the memory of our fellow citizens and revive the many stories that accompany the creation and daily life of the different Senegalese territories, we initiated the establishment of an online application that allows Senegalese communities to share and co-construct their cultural heritage.

Our proposals are based on social and semantic web technologies. indeed, social web proposes a framework where value is created by the aggregation of many individual user contributions. Thereby, social web makes easier corpus co-construction. The semantic web enables to find, to combine and to share resources, not only between humans but also between machines. The combination of these two technologies enables Senegalese communities to share and co-construct their cultural heritage in a collaborative and semantic environment .

Our contributions include to (i) propose ontologies to annotate sociocultural resources and (ii) provide a framework for communities to share and co-construct their knowledge. Ontologies are backbone of the semantic web and allow to characterize a domain by describing the basic concepts and the relations between them. Thus, we have defined two ontologies : 1) a sociocultural ontology based on cultural-historical activity theory and 2) a temporal ontology to annotate temporally sociocultural resources. We also proposed a virtual community called cultural knowledge-building community which is an adaptation of the knowledge-building community in the cultural field.

In this dissertation, we present (1) a state of the art, one hand, on the capitalization of knowledge that covers both the cultural-historical activity theory and typologies of virtual communities and, other hand, the temporal annotation of semantic Web resources, (2) a methodology to build a sociocultural ontology and characteristics of the cultural knowledge-building communities, (3) an ontology to annotate temporally semantic web resources particularly human time expressions and (4) the implementation of a system that integrates the scientific contributions of this work and that allows Senegalese communities to share and co-construct their cultural heritage.

Keywords : Semantic web, Social web, Ontologies, Virtual Communities, Society, Culture, Temporal Annotation, Co-construction and Share.

Contents

1	Introduction Générale	1
1.1	Introduction	1
1.2	Contexte de recherche	2
1.3	Problématique et questions de recherche	3
1.3.1	Web Sémantique	4
1.3.2	Web social	5
1.4	Contributions	6
1.4.1	Des ontologies pour annoter des ressources socioculturelles	7
1.4.2	Web social	8
1.4.3	Validation des contributions	8
1.5	Organisation du manuscrit	9
1.5.1	Première partie : état de l'art	9
1.5.2	Deuxième partie : contributions	10
I	État de l'art	11
2	Capitalisation de Connaissances	13
2.1	Introduction	13
2.2	Théorie historico-culturelle de l'activité	14
2.2.1	La Première génération de la CHAT	15
2.2.2	La Seconde génération de la CHAT	16
2.2.3	La Troisième Génération de la CHAT	17
2.2.4	Utilisation de la CHAT comme modèle	18
2.3	Les Communautés Virtuelles	20
2.3.1	Définition d'une Communauté virtuelle	21
2.3.2	Typologie des Communautés Virtuelles	24
2.3.3	Animation des communautés	32
2.4	Conclusion	35
3	Ontologies Temporelles au sein du Web Sémantique	37
3.1	Introduction	37
3.2	Définition de quelques notions	38
3.3	Modélisation d'expressions temporelles	40
3.3.1	Ontologies de haut niveau	40
3.3.2	Ontologies de domaine	44
3.4	Annotation temporelle de données	53
3.4.1	Approches du W3C	53
3.4.2	Approches spécifiques	56
3.5	Conclusion	60

4	Positionnement	63
4.1	Introduction	63
4.2	Capitalisation de connaissances	63
4.3	Ontologies Temporelles	67
4.4	Conclusion	69
II	Contributions	71
5	Capitalisation de Connaissances : Ontologie Socioculturelle et Communauté de Co-élaboration de Connaissances Culturelles	73
5.1	Introduction	73
5.2	Une ontologie socioculturelle pour le partage et la co-construction de connaissances	74
5.2.1	Extension du modèle de Vygotsky	75
5.2.2	Ontologie de haut-hiveau	75
5.2.3	Alignement de schémas	78
5.2.4	Cas d'utilisation de USCO	80
5.3	Communauté de co-élaboration de connaissances culturelles	82
5.3.1	Introduire les CKBC aux membres	85
5.3.2	Technologie utilisée : MediaWiki	86
5.3.3	Animation des communautés utilisatrices	88
5.4	Synthèse	89
6	HuTO : Human Time Ontology (une ontologie temporelle narra- tive)	91
6.1	Introduction	91
6.2	Modélisation d'expressions temporelles	93
6.2.1	Généralités	93
6.2.2	Date, temps calendaire et unités temporelles	94
6.2.3	Instant, intervalle et durée	97
6.2.4	HuTO et les relations d'Allen	100
6.3	Annotation temporelle des données	101
6.4	Règles temporelles	106
6.4.1	Normalisation de la représentation temporelle	106
6.4.2	Implications et inférences	110
6.5	Requêtes-types	112
6.5.1	Requête-type sur les ressources	112
6.5.2	Requêtes-types sur les éléments temporels	113
6.6	Transformations	116
6.6.1	Transformation RDF2Text	117
6.7	Synthèse	121
6.7.1	Comparaison de HuTO avec les approches de l'état de l'art	123

7	Implémentation et Validation	127
7.1	Introduction	127
7.2	Architecture de la plateforme	128
7.2.1	Couche IHM	129
7.2.2	Couche persistance	132
7.2.3	Couche sémantique	134
7.3	Ontologie socioculturelle du Sénégal	135
7.4	Concepts temporels de la plateforme	137
7.5	Prototype	141
7.5.1	Outils installés	142
7.5.2	Implémentation	143
7.5.3	Maquette	149
7.6	Synthèse	154
8	Conclusion générale et perspectives	157
8.1	Introduction	157
8.2	Bilan	157
8.3	Publications	161
8.4	Perspectives	162
A	Annexe : HuTO	165
A.1	Concepts généraux de HuTO	165
A.2	Règles	165
A.3	Requêtes Types	173
B	Annexe : Installation des Outils	183
	Bibliography	187

List of Figures

2.1	Triangle de médiation de Vygotsky [17].	16
2.2	Modèle du système d'Activité d'Engeström [16].	18
2.3	Caractérisation des communautés virtuelles [46].	26
2.4	Typologie des communautés virtuelles [44].	27
3.1	Les 13 relations d'Allen.	39
3.2	Représentation en relations n-aires.	55
3.3	Représentation avec la réification RDF.	56
3.4	Représentation en Temporal RDF.	57
3.5	Représentation en 4D-Fluents.	59
5.1	Ontologie Socioculturelle de haut niveau (USCO).	76
5.2	Relations entre les concepts.	77
5.3	Alignement avec Schema.org et DBpedia.	79
5.4	Cas d'utilisation de l'ontologie USCO.	81
6.1	Principaux sous concepts et propriétés du concept <i>TemporalPosition</i>	94
6.2	Les intervalles et leurs propriétés dans HuTO.	97
6.3	Modélisation temporelle qualitative dans HuTO.	104
6.4	Résultats de clauses SELECT (à gauche) et DESCRIBE (à droite).	117
6.5	Exemple de résultats de transformations sur une requête temporelle.	121
7.1	Architecture de la plateforme WestAfricapedia.	128
7.2	Ontologie Socioculturelle du Sénégal.	138
7.3	Transformation du résultat de la requête en un format JSON Exhibit.	148
7.4	Code de la page Exhibit Simple.	149
7.5	Page d'accueil de la plateforme.	150
7.6	Page de la Région de Louga	151
7.7	Création ou mise à jour d'une page de type Infrastructure	152
7.8	Création ou mise à jour d'une page de type Activité	152
7.9	Page Exhibit Simple sur les Infrastructure	153
7.10	Page Exhibit Map sur les Infrastructure	153
A.1	Principaux concepts de HuTO.	166

List of Tables

4.1	Comparaison de l'approche de Vygotsky (socioculturelle) et celle utilisée dans la plateforme.	65
4.2	Comparaison de la Communauté de Co-élaboration de Connaissances Éducationnelles - C3E (Educational Knowledge-building Community - EKBC et de la Communauté de Co-élaboration de Connaissances Culturelles - C4 (Cultural Knowledge-building Community - CKBC)	66
4.3	Comparaison des différentes approches dans le domaine de la modélisation d'expressions temporelles (✓ signifie que l'ontologie intègre cette caractéristique sinon ×).	68
4.4	Comparaison des différentes approches dans le domaine de l'annotation temporelle des données (✓ signifie que l'ontologie intègre cette caractéristique sinon ×).	69
6.1	Comparaison de HuTO avec les trois approches les plus significatives dans le domaine de la modélisation d'expressions temporelles (✓ signifie que l'ontologie intègre cette caractéristique sinon ×).	123
6.2	Comparaison de HuTO avec les approches majeures dans le domaine de l'annotation temporelle des données (✓ signifie que l'ontologie intègre cette caractéristique sinon ×).	124
7.1	Correspondance entre SMW et OWL.	131
7.2	Comparaison entre les triplestores utilisés par SMW (M signifie million et B signifie milliard. À titre indicatif, DBpedia compte moins de 440M de triplets).	133

Introduction Générale

Sommaire

1.1	Introduction	1
1.2	Contexte de recherche	2
1.3	Problématique et questions de recherche	3
1.3.1	Web Sémantique	4
1.3.2	Web social	5
1.4	Contributions	6
1.4.1	Des ontologies pour annoter des ressources socioculturelles . .	7
1.4.2	Web social	8
1.4.3	Validation des contributions	8
1.5	Organisation du manuscrit	9
1.5.1	Première partie : état de l'art	9
1.5.2	Deuxième partie : contributions	10

1.1 Introduction

En Afrique, la méconnaissance de nos territoires s'aggrave de jour en jour avec l'expansion de la télévision (mondialisation des contenus avec les chaînes câblées de plus en plus nombreuses) et l'émergence de l'Internet dans les pays de l'Afrique, qui poussent de plus en plus les jeunes à se tourner vers l'Occident au détriment de la découverte de leurs espaces nationaux. Il est très courant de rencontrer de jeunes africains connaissant mieux la géographie de la France que celle de leur pays, connaissant mieux les histoires et légendes de telle ou telle ville de France ou d'Europe, alors qu'à quelques kilomètres de chez eux existe une ville qui, bien que regorgeant de récits palpitants, leur est quasiment inconnue. Ils arriveront à peine à la situer sur une carte et encore moins à citer une seule histoire de cette ville. Ce fossé d'information se retrouve aussi sur l'Internet où il existe peu d'information sur la culture des pays africains.

Comme le montre l'UNESCO¹ [1], la connaissance et l'information ont un impact significatif sur la vie des gens. Le partage de connaissances et de l'information,

¹United Nations Educational, Scientific and Cultural Organization (Organisation des Nations unies pour l'éducation, la science et la culture)

notamment par le biais des Technologies de l'Information et de la Communication (TIC) a le pouvoir de transformer les économies et les sociétés mais également le développement humain. Pour atteindre cet objectif, l'UNESCO identifie certaines solutions parmi lesquelles l'accès libre à l'information via des plateformes ouvertes et à distance [1]. Ceci doit permettre une maîtrise de l'information par les populations et créer de nouvelles opportunités pour améliorer la qualité de la vie.

Cela devient plus facile avec l'apparition du Web social où l'utilisateur est actif. En effet, avec les technologies du Web social, l'utilisateur peut participer à la rédaction du contenu du Web, ce qui peut lui permettre de partager ses connaissances avec d'autres utilisateurs. Toujours dans cette optique, le Web sémantique [2] peut aussi faciliter ce partage de connaissances en rendant les ressources du Web social plus aisément accessibles aussi bien pour l'homme que pour les agents logiciels, grâce à la représentation sémantique de leurs contenus et à l'échange de données entre les applications au travers du Web.

Cette thèse cherche à proposer des éléments de solution afin de permettre aux communautés sénégalaises de partager leurs connaissances socioculturelles. Cette solution ambitionne d'utiliser à la fois les technologies du Web social et celles du Web sémantique.

1.2 Contexte de recherche

Cette thèse se place dans un contexte ouest africain où la problématique principale est celle du Développement économique et social et des moyens pour y accéder. Après l'échec et la remise en cause de plusieurs modèles de développement ces dernières décennies, la recherche théorique se tourne vers des approches du développement basées sur les connaissances.

Ainsi pour entretenir la mémoire des citoyens et redonner vie aux nombreux récits qui accompagnent la création et la vie au quotidien des différents terroirs africains, la mise en place d'une encyclopédie socioculturelle en ligne a été imaginée au sein du projet WestAfricapedia. La place du savoir, des sciences et des technologies dans la dynamique actuelle de la croissance donne lieu à une intense réflexion au sein du champ économique. Dans les pays de la zone UEMOA² les systèmes nationaux d'information sont précaires et les capacités technologiques des entreprises locales sont faibles. Engager un processus durable d'accumulation de connaissances susceptible d'accélérer le développement des capacités productives dans ces pays est nécessaire. Toute stratégie dans ce domaine doit viser non seulement à renforcer le socle de connaissances endogènes, mais aussi à favoriser le transfert et la bonne assimilation de la technologie. Les savoirs traditionnels, informels, dont sont porteurs les individus jouent un rôle crucial dans divers secteurs, notamment la culture, l'agriculture, la santé et les industries de la création. Les politiques de modernisation des capacités technologiques doivent

²UEMOA : Union Économique et Monétaire Ouest-Africaine.

valoriser ces savoirs locaux et les intégrer aux technologies.

Le projet WestAfricapedia répond à un besoin de vulgarisation et de pérennisation de la culture africaine et sénégalaise en particulier à travers un partage des coutumes et de l'histoire des différentes localités du Sénégal. Il peut être assimilé à la construction d'une plateforme qui serait à cheval entre une « mémoire d'entreprise » (ou plus exactement une « mémoire d'organisation ») et un « réseau social », mais appliquée au contexte d'un pays.

L'objectif général de cette thèse est de soutenir l'approche du développement socio-économique fondée sur les connaissances en mettant en œuvre un cadre durable permettant aux communautés d'identifier et de partager leurs ressources (en termes de connaissances, de savoirs, de potentiels, de projets, etc.) et leurs besoins (manque d'infrastructures, de moyens, etc.). Ceci a pour but d'avoir une vue d'ensemble des opportunités et problèmes au niveau de tout le territoire et de rassembler des informations venant de la base (des acteurs au quotidien dans les différentes communautés), plutôt que d'éléments globaux éloignés de la réalité du terrain. Dans le cadre de cette thèse, nous proposons une plateforme en ligne enrichie par une couche sémantique permettant l'accès aux différentes informations. Cette plateforme sera mise à jour par les acteurs et les citoyens de ces territoires, qui partageront ainsi leur histoire et leur patrimoine à travers leur plateforme de type «Web social».

L'objectif de cette thèse est de permettre aux communautés sénégalaises de partager leurs connaissances au moyen des technologies du Web social et du Web sémantique. En effet, la combinaison de ces deux approches produit des « systèmes de connaissances collectives » [3] (**collective knowledge systems**). Les technologies du Web social permettent aux utilisateurs de co-construire un corpus, une « intelligence collective » [3] (**collective intelligence**). L'utilisation des technologies du Web sémantique au travers la représentation et le raisonnement permettent de créer de nouvelles connaissances et offrent de meilleurs moyens d'apprendre et de partager à travers le Web.

1.3 Problématique et questions de recherche

Le travail réalisé pendant cette thèse s'inscrit sur deux domaines. Premièrement, dans le domaine du Web sémantique où la problématique de la thèse est de permettre l'annotation des ressources d'un environnement socioculturel. Deuxièmement, dans le domaine du Web social où la problématique de la thèse est de proposer un cadre permettant aux communautés sénégalaises de créer une « intelligence collective » sémantiquement annotée pour produire un « système de connaissances collectives » dans le domaine socioculturel.

1.3.1 Web Sémantique

« Le Web sémantique repose sur deux choses : un ensemble de formats communs pour l'intégration et la combinaison de données provenant de diverses sources, là où le Web originel est concentré principalement sur l'échange de documents. C'est aussi une façon de représenter à l'aide de langages comment les données se rapportent à des objets du monde réel. Cela permet à une personne, ou une machine, à partir d'une source de données, d'accéder à un ensemble de données qui sont reliées non pas par des fils mais par des thématiques. Ainsi, le Web sémantique fournit un cadre commun qui permet aux données d'être partagées et réutilisées par des applications, des entreprises et des communautés. »³

Tim Berners-Lee et ses collaborateurs présentent leur vision du Web sémantique comme un vaste espace d'échange de ressources entre êtres humains et machines permettant une exploitation, qualitativement supérieure, de grands volumes d'informations et de services variés [2]. En effet, le Web actuel est essentiellement syntaxique, dans le sens où la structure des documents est bien définie, leurs contenus restent difficilement accessibles aux traitements automatiques par les machines. Seuls les humains peuvent interpréter ces contenus. Le Web sémantique a pour ambition de lever cette difficulté. L'idée est que la connaissance devrait être accessible aussi bien à l'homme qu'à la machine, permettant aux programmes d'effectuer des recherches intelligentes en utilisant diverses sources d'informations. Ainsi les données seraient plus facilement accessibles, grâce à la représentation sémantique de leur contenu. Cette représentation sémantique repose sur des **ontologies** qui sont des **spécifications explicites de conceptualisations** [4]. Une conceptualisation rend compte du sens des termes; dans un domaine donné elle définit quelles sont les entités (concepts) qui le caractérisent. Une spécification explicite signifie que les relations qui existent entre les concepts, les propriétés, les fonctions et les axiomes sont définis explicitement.

Les ontologies sont apparues avant le Web sémantique. En effet, les ontologies sont nées durant le projet DARPA Knowledge Sharing Effort de 1991 et la première conférence scientifique dédiée à l'ingénierie ontologique a été initiée en 1996 [5]. L'objectif principal de cette rencontre a été de fixer un ensemble de critères et de règles de bonne pratique pour construire des ontologies. En effet, jusqu'en 1996, les premières ontologies ont été développées de façon artisanale, sans suivre de méthode prédéfinie. Les premiers projets Enterprise Ontology [7] et Toronto Virtual Enterprise (TOVE) [6] ont permis de proposer des listes de recommandations constituant des ébauches de méthodes ou de cadres méthodologiques. Les deux projets étaient dans le domaine de la modélisation d'entreprise. Ainsi, à partir de 1997, nous avons assisté à la naissance de cadres méthodologiques plus élaborés inspirés des méthodes de l'ingénierie de connaissances (METHONTOLOGY [8] et SENSUS [9]) et du domaine de la linguistique (TERMINAE [10]). Gómez-Pérez et ses collègues répertorient plus d'une dizaine d'approches dans différents domaines

³<http://www.w3.org/2001/sw/>, vu le 23 Mars 2016.

[5]. Le résultat de cette étude montre qu'il n'y a pas de lignes directrices communes et il n'existe aucune méthodologie standard pour le développement des ontologies [11].

Une approche répandue de l'ingénierie ontologique s'appuie sur la réutilisation d'ontologies existantes. Une étude sur les processus de développement des ontologies dans le domaine socioculturel a été menée [12]. Cette étude a montré qu'il n'existait pas d'ontologie dans ce domaine. En outre, la manipulation de données socioculturelles fait intervenir de nombreuses notions temporelles parmi lesquelles des notions temporelles partagées à travers le monde, ainsi que des notions plus spécifiques à la culture d'un pays ou d'une région.

Aussi, le premier travail dans le domaine du Web sémantique est de proposer une méthodologie pour identifier les caractéristiques qui représentent les aspects sociaux d'une communauté, à savoir les communautés, leurs activités et les informations générales autour de leur existence. Nous entendons par méthodologie les procédures de travail, les étapes, qui décrivent le pourquoi et le comment de la conceptualisation puis de la construction de l'artefact.

Les approches décrites par Gómez-Pérez et ses collègues [5] comme METHONTOLOGY [8] et SENSUS [9] ont en commun le fait qu'elles ont commencé à partir de l'identification de l'objet de l'ontologie et la nécessité de la connaissance du domaine. Ainsi comme nous nous sommes intéressés aux aspects socioculturels des communautés sénégalaises, nous avons choisi de nous appuyer sur la théorie historico-culturelle de l'activité (Cultural Historical Activity Theory - CHAT) [17, 18] qui est un cadre conceptuel qui émane des travaux de psychologues russes pour étudier et expliquer les activités humaines dans un processus d'apprentissage. Cette approche nous a servi de *méta-modèle* pour la conception de l'ontologie socioculturelle proposée.

Le deuxième travail dans le domaine du Web sémantique a été de proposer une ontologie temporelle pour annoter temporellement les ressources socioculturelles. Plus précisément, l'objectif principal a été de proposer une façon d'annoter les expressions humaines du temps. En effet comme le montre Hall [13], le temps humain est lié à la culture dans la mesure où c'est est un agrégat de concepts, de phénomènes et de rythmes recouvrant une très large réalité.

1.3.2 Web social

Le Web social, parfois appelé Web 2.0⁴, est une nouvelle vision du Web où l'utilisateur n'est plus seulement un consommateur mais aussi un producteur du contenu du Web. Cette nouvelle vision considère le Web comme un espace de socialisation qui consiste à mettre en relation les utilisateurs, à leur faciliter l'interaction et à le rendre plus riche et plus productif. Selon Gruber [3], le Web social est un écosystème de la participation, où la valeur est créée par le regroupement de

⁴Le terme a été inventé par Dale Dougherty de la société O'Reilly Media lors d'un brainstorming avec Craig Cline de MediaLive en octobre 2004.

plusieurs contributions des utilisateurs individuels. Ainsi le Web social est un espace dont le but est de faire interagir les utilisateurs afin d'assurer une production continue de contenus.

Les technologies du Web 2.0 (blogs, wikis, services de partage de contenus, ...) facilitent le partage et la collaboration entre individus, dans l'objectif de faire émerger une « intelligence collective » ou une « sagesse de groupe ». Elles introduisent de nouvelles problématiques en termes de relations et d'échanges entre les utilisateurs. Les différents types de relations et d'échanges entre les utilisateurs sont assujettis aux types de technologies utilisées. Ainsi, les technologies du Web 2.0 peuvent être classées en trois catégories :

1. Les Wikis dont l'avantage est de permettre aux utilisateurs de se corriger entre eux pour obtenir de meilleurs résultats : rien n'est jamais parfait mais il y a une amélioration après chaque participation. Les utilisateurs sont complémentaires et le travail est collaboratif.
2. Les blogs, diminutif de Web log. Un blog "est un site [Web personnel] présentant sur sa page d'accueil un ensemble de billets (posts) consistant en des notes ou articles plus ou moins longs et ordonnés de manière antichronologique. L'usage d'hyperliens (internes et externes) y étant abondant" [14]. Contrairement au Wiki, un blog est souvent la propriété d'un acteur - blogueur - qui met à jour son "journal du web".
3. Les services de partage de contenus : ce sont des sites de partage de contenus comme Youtube.com (vidéo), Flickr.com (photo), SlideShare.com (présentation), Twitter.com (microblogage) et Facebook.com (relation ou connexion). Les utilisateurs échangent et se retrouvent autour de ces objets particuliers pour y former des réseaux.

L'objectif dans le contexte du Web social est de permettre aux utilisateurs de créer un corpus de manière collaborative. Ainsi, la technologie pour favoriser cela est celle des Wikis où le travail est collaboratif, ce qui est essentiel dans le domaine culturel puisque la connaissance culturelle est partagée au sein d'une communauté. Ainsi, le travail dans le domaine social est de positionner les communautés utilisatrices par rapport aux types de communautés dans le domaine du Web social.

1.4 Contributions

Comme nous l'avons décrit précédemment, l'objectif de cette thèse est de fournir un cadre permettant aux communautés de partager et de co-construire leur patrimoine culturel. Les contributions de cette thèse s'inscrivent à la fois dans le domaine du Web sémantique et dans le domaine du Web social.

1.4.1 Des ontologies pour annoter des ressources socioculturelles

Dans le domaine du Web sémantique, cette thèse présente deux contributions : (i) une proposition de méthodologie de construction d'une ontologie socioculturelle et (ii) une ontologie temporelle.

1. Le processus de développement proposé repose sur l'extension de l'approche socioculturelle de la théorie historico-culturelle de l'activité (Cultural Historical Activity Theory - CHAT) à une communauté. Pour cela, nous avons identifié trois principaux concepts de l'ontologie **USCO** (*Upper-level SocioCultural Ontology*) proposée: *usco:Artefact* et *usco:Object* qui sont hérités de l'approche socioculturelle et le concept *usco:Community*. Puisque conceptualiser un domaine aussi large que la culture est compliqué, nous avons aligné l'ontologie USCO avec deux vocabulaires du Linked Open Data (LOD). En effet, le LOD est un jeu de données ouvert et lié qui couvre de très nombreuses thématiques. Dans le LOD, nous avons choisi les vocabulaires de DBpedia⁵ et Schema.org⁶. Schema.org a été choisi car il propose une large collection de schémas pour des données structurées. En plus c'est un travail collaboratif entre les géants de l'Internet (Google, Microsoft, Yahoo et Yandex). Quant à DBpedia, il fournit une ontologie et une base de connaissances extraites de l'encyclopédie collaborative Wikipedia. Les alignements réalisés nous ont permis de bénéficier des descriptions faites dans ces vocabulaires. Ainsi, l'ontologie USCO que nous avons proposée permet à des communautés de partager et de co-construire leur patrimoine socioculturel. Cela se fait à travers les descriptions des événements socioculturels organisés par ces communautés, les descriptions des ressources à la disposition de ces communautés mais également les descriptions de ces communautés elles-mêmes.
2. L'ontologie temporelle **HuTO**⁷ (*Human Time Ontology*) : HuTO est une ontologie narrative pour annoter temporellement des ressources du Web sémantique. Cette ontologie apporte deux contributions dans l'annotation temporelle.
 - (a) En ce qui concerne la modélisation d'expressions temporelles, HuTO permet une modélisation de l'expression humaine du temps. C'est-à-dire une modélisation qualitative qui permet de décrire des notions temporelles relatives à la culture ou à la position géographique et qui donne la possibilité d'utiliser une ressource comme un marqueur temporel (la ressource est utilisée comme une date). L'ontologie HuTO permet aussi d'étendre ses unités temporelles pour intégrer d'autres unités temporelles comme celles du calendrier Maya ou Aztèque.

⁵<http://dbpedia.org/About/>

⁶<http://schema.org>

⁷<http://ns.inria.fr/huto>

- (b) En matière d'annotation temporelle des ressources, la contribution de l'ontologie HuTO permet de garder la sémantique des triplets RDF tout en ajoutant une dimension temporelle et de ne pas étendre les standards du Web sémantique.
- (c) Pour compléter les deux premières contributions, nous avons proposé des requêtes type qui permettent d'interroger une base de connaissances annotée avec HuTO. Nous avons aussi proposé des transformations de RDF en langage naturel (RDF2Text) afin de rendre les résultats des requêtes mieux compréhensibles par un utilisateur humain.

1.4.2 Web social

L'importance croissante du Web social et des communautés virtuelles nous a amené à positionner nos communautés utilisatrices résultant de l'extension de la théorie historico-culturelle de l'activité (Cultural Historical Activity Theory - CHAT) à une communauté. Nous avons d'abord observé les communautés auxquelles se destine la plateforme. Il s'agit des communautés de co-élaboration de connaissances (Knowledge-building Community - KBC) [50, 52] dans le domaine socioculturel. Ces communautés, de co-élaboration de connaissances culturelles (Cultural Knowledge-building Community - CKBC), rassemblent des individus qui participent à un processus social et cognitif d'appropriation, de transmission, de production et de création de connaissances caractérisant leur communauté. Ce processus permet (i) la construction de connaissances à partir de sources (ressources) sociales et culturelles, et (ii) l'intégration de cette connaissance à des connaissances pré-existantes (connaissances des générations précédentes). Nous avons proposé des caractérisations d'une CKBC et des types de membres qui participent à la co-élaboration de connaissances.

La contribution dans ce domaine est l'adaptation de la notion des KBC dans le domaine socioculturel (CKBC) en proposant 4 principes pour une bonne co-élaboration de connaissances socioculturelles et de proposer un environnement technologique qui permet de supporter le travail collaboratif des CKBC.

1.4.3 Validation des contributions

Nous avons également déployé une plateforme qui intègre les résultats de nos contributions scientifiques. Cette plateforme s'inscrit dans le contexte sénégalais. De ce fait, l'ontologie USCO a été contextualisée pour les phénomènes socioculturels du Sénégal en ajoutant des concepts spécifiques comme *usco:ReligiousEvent*, *usco:ReligiousCommunity*, etc. De même, des notions temporelles qualitatives relatives au Sénégal ont été ajoutées à l'ontologie HuTO. La plateforme développée repose sur Semantic MediaWiki (SMW) qui apporte une couche sémantique à MediaWiki (MW). Ainsi, les ressources utilisées dans la plateforme sont annotées par rapport au vocabulaires de HuTO et de USCO. En effet, SMW permet d'importer des vocabulaires. Le Triple Store Virtuoso est utilisé pour sauvegarder les ontolo-

gies (USCO et HuTO) et les données annotées dans la plateforme. Puisque USCO est aligné avec DBpedia, la base de connaissances (BC) est partiellement mise à jour avec les données qui concernent le Sénégal dans DBpedia. Pour une recherche d'information conviviale, l'outil Exhibit du projet SIMILE est utilisé. En effet, Exhibit permet d'afficher sur une carte Google Map toutes les ressources géolocalisées dans la plateforme. Ces ressources sont filtrées par type et par localisation.

La plateforme développée permet aux communautés sénégalaises de partager et de co-construire leur patrimoine culturel en décrivant les événements socioculturels qu'elles organisent, les ressources à leur disposition (infrastructures, localités, etc.), le tout de façon collaborative.

1.5 Organisation du manuscrit

L'organisation de ce manuscrit s'articule en deux parties principales. Une première partie est consacrée à l'étude bibliographique abordant les différents champs de recherche qui rentrent dans le contexte des sujets que nous traitons dans cette thèse. La deuxième partie présente nos contributions qui visent essentiellement à permettre aux communautés de partager et de co-construire leur patrimoine culturel.

1.5.1 Première partie : état de l'art

Cette partie est composée de trois chapitres. Le chapitre 2 intitulé **Capitalisation de connaissances** présente les deux principaux axes de recherche dans lesquels s'inscrit le travail sur les communautés de co-élaboration de connaissances socioculturelles. Dans un premier temps, nous présentons les trois étapes de la création de la théorie historico-culturelle de l'activité (CHAT) sur laquelle s'appuie le travail. Nous présentons également différents travaux qui utilisent la CHAT comme modèle conceptuel pour l'apprentissage collaboratif, la co-construction de connaissances et l'Interaction Homme-Machine (IHM). Dans un deuxième temps, nous présentons les communautés virtuelles de la littérature. Ensuite nous présentons les différentes typologies et nous terminons par présenter les différents modes d'animation de ces communautés.

Le chapitre 3 présente les travaux sur la modélisation temporelle dans le Web sémantique. Ce chapitre commence par présenter les définitions qui ont servi comme métriques pour comparer les ontologies temporelles. Ensuite, nous présentons les ontologies temporelles de la littérature qui sont des ontologies de haut niveau ou des ontologies de domaine. La dernière partie du chapitre est consacrée à l'annotation temporelle des ressources et présente à la fois les approches préconisées par le W3C et les approches spécifiques.

Le dernier chapitre de l'état de l'art, le chapitre 4, présente une synthèse des chapitres 2 et 3. Dans ce chapitre nous expliquons comment la théorie de la CHAT pourra être utilisée comme méta-modèle dans nos travaux. Ensuite, une synthèse est faite sur les différentes typologies de communautés virtuelles et nous avons présenté la spécificité de nos communautés par rapport à la typologie la plus proche que

nous avons rencontrée dans la littérature. Une synthèse est aussi faite dans le domaine de la modélisation d'expressions temporelles en comparant les approches les plus complètes, ce qui a permis d'identifier les critères à respecter dans la modélisation. Une partie de cette synthèse porte également sur l'annotation temporelle des ressources. Ceci a permis également de mettre en évidence les caractéristiques les plus importantes que doit intégrer l'approche proposée.

1.5.2 Deuxième partie : contributions

Dans cette partie, nous présentons nos contributions visant principalement à permettre à des communautés de partager et de co-construire leur patrimoine culturel. Le chapitre 5 décrit, en premier, l'approche que nous proposons pour construire une ontologie socioculturelle, USCO (Upper-level SocioCultural Ontology). Les trois phases du processus de construction sont présentées : (i) extension de la CHAT, (ii) proposition d'une ontologie de haut-niveau développée et (iii) alignement avec deux vocabulaires du Linked Open Data (LOD). Un cas d'utilisation de l'ontologie USCO est aussi présenté. La deuxième partie du chapitre présente les caractéristiques des communautés de co-élaboration de connaissances culturelles au Sénégal. Nous avons aussi présenté la technologie utilisée pour faciliter cette co-élaboration et les type de membres au sein de ces communautés.

Le chapitre 6 présente l'ontologie temporelle narrative, HuTO (Human Time Ontology). Cette ontologie couvre les deux aspects de modélisation identifiées dans l'état de l'art c'est-à-dire la modélisation d'expressions temporelles et l'annotation temporelle des ressources. Nous présentons également les règles temporelles que nous avons proposées et qui font partie de la spécification de HuTO⁸. Pour une utilisation plus conviviale, nous avons également proposé des requêtes type pour interroger une base de connaissances annotée avec HuTO. Nous présentons aussi nos transformations RDF2Text⁹ qui permettent un rendu plus lisible et compréhensible des requêtes SPARQL sur des données annotées grâce à HuTO. La dernière partie du chapitre présente des comparaisons de l'ontologie HuTO avec les meilleures approches identifiées dans l'état de l'art.

Enfin, le chapitre 7 décrit le prototype que nous avons développé et qui permet de mettre en œuvre les contributions scientifiques de cette thèse. Plus précisément, dans la première partie du chapitre, nous présentons l'architecture générale et fonctionnelle de la plateforme. Ensuite, nous montrons comment les deux ontologies développées dans cette thèse, à savoir USCO et HuTO, sont utilisées dans le contexte sénégalais. En dernier lieu, nous décrivons les outils utilisés, l'implémentation faite et la maquette de la plateforme.

⁸<http://ns.inria.fr/huto/rule/>

⁹<http://ns.inria.fr/huto/rdf2text/>

Part I

État de l'art

Capitalisation de Connaissances

Sommaire

2.1	Introduction	13
2.2	Théorie historico-culturelle de l'activité	14
2.2.1	La Première génération de la CHAT	15
2.2.2	La Seconde génération de la CHAT	16
2.2.3	La Troisième Génération de la CHAT	17
2.2.4	Utilisation de la CHAT comme modèle	18
2.3	Les Communautés Virtuelles	20
2.3.1	Définition d'une Communauté virtuelle	21
2.3.2	Typologie des Communautés Virtuelles	24
2.3.3	Animation des communautés	32
2.4	Conclusion	35

2.1 Introduction

Un des objectifs de cette thèse est de proposer un cadre qui permet aux communautés sénégalaises de partager et de co-construire leurs connaissances socioculturelles. Ce cadre de partage est sous-tendu par une couche sémantique qui est modélisée grâce à des ontologies (cf. chapitre 1). Aussi, dans ce chapitre nous présentons deux domaines de recherche connexes. Le premier domaine concerne la théorie historico-culturelle de l'activité (CHAT¹) qui émane des travaux de psychologues pour l'étude sur les transformations et les processus d'apprentissage dans les activités et les organisations. L'objectif de la première partie est ainsi de proposer un cadre qui permet aux membres des communautés sénégalaises, grâce à un processus de partage et de co-construction, d'apprendre sur leur environnement socioculturel. De ce fait, cette théorie a permis de disposer d'un cadre formel pour modéliser l'ontologie socioculturelle qui sous tend la plateforme.

La deuxième partie de ce chapitre concerne les travaux sur les communautés virtuelles. Ces dernières émanent de l'utilisation et de la diffusion des Technologies de l'Information et de la Communication (TIC) qui offrent la possibilité de redéfinir

¹Cultural Historical Activity Theory

et de repenser la notion de *communauté* à la fois en termes de délimitation des frontières et en termes de modes de communication utilisés entre les membres. Avant l'avènement des TIC, la notion de communauté était surtout caractérisée par la proximité géographique et des rencontres physiques. Avec l'apparition des communautés virtuelles, ces deux contraintes n'existent plus pour créer une communauté.

Dans nos travaux, la CHAT a permis d'avoir un cadre scientifique pour modéliser l'ontologie socioculturelle proposée. La modélisation de cette ontologie fait apparaître une communauté utilisatrice qu'il est nécessaire de positionner dans la littérature d'où l'étude des communautés virtuelles.

Ainsi, ce chapitre est composé de deux grandes parties. En premier nous présentons la CHAT au travers de ses trois phases d'élaboration. Nous présentons également les travaux qui utilisent cette théorie comme cadre de modélisation dans le domaine de l'apprentissage et de l'interaction homme-machine. La deuxième partie de ce chapitre présente les travaux sur les communautés virtuelles. En premier, nous donnons des définitions de la littérature et ensuite nous présentons les typologies des communautés virtuelles. Ensuite, nous présentons certaines caractéristiques qui permettent d'évaluer le succès des communautés virtuelles. Nous terminerons par une conclusion.

2.2 Théorie historico-culturelle de l'activité

La théorie historico-culturelle de l'activité (CHAT) a été initiée par un groupe de psychologues russes autour des travaux de Lev S. Vygotsky dans les années 1920 et 1930. L'objectif principal de cette théorie est d'expliquer comment le fonctionnement mental individuel (appropriation de nouvelles connaissances) est lié au contexte culturel, institutionnel et historique de son environnement. Ainsi, la théorie met l'accent sur le lien entre l'apprentissage et le développement et comment la théorie améliore les interactions sociales. Par conséquent, un accent est mis sur la perspective socioculturelle qui concerne les rôles au sein des interactions sociales et des activités culturelles organisées pour influencer le développement mental (appropriation de nouvelles connaissances) [15].

La construction de la CHAT a connu trois étapes importantes. Engeström décrit trois périodes distinctes de la CHAT [16]. La première génération concerne sa création par Vygotsky [17] et ses collègues principalement A. R. Luria and A. N. Leont'ev. La seconde génération a été développée après le décès de Vygotsky par ses anciens collègues et particulièrement Leont'ev [18] qui marque le caractère collectif de l'activité humaine. Enfin, la dernière génération concerne les travaux d'Engeström qui a développé des outils conceptuels pour comprendre le dialogue et les multiples points de vue dans l'activité humaine [16, 19]. Les modèles élaborés pendant ces trois périodes s'inscrivent dans une continuité qui marque l'évolution de la CHAT et la participation graduelle de la communauté où l'activité est étudiée.

Ainsi la CHAT n'est pas une méthodologie, encore moins une théorie prédictive;

elle est un cadre théorique ou un outil se référant à une approche interdisciplinaire pour étudier l'apprentissage humain et le développement. Notons qu'il y a eu un débat animé au cours des dernières années pour étendre les travaux de Vygotsky [17] sur l'aspect communautaire (collectif) [20]. Ainsi comme le montre Cole [20]; il existe deux positions par rapport à l'extension des travaux de Vygotsky et de ses collègues. Celle qui suit les travaux de Vygotsky et qui s'articule autour d'un processus de médiation comme unité de base de l'analyse [21, 22, 23]. Cet axe de recherche est souvent désigné comme la recherche dans le domaine socioculturel [20]. Il y a aussi ceux qui suivent les travaux de Leont'ev qui considèrent l'activité comme unité de base de l'analyse [16, 19, 24]. Leur domaine de recherche est parfois appelé Théorie de l'Activité au sein de la CHAT. Notons qu'il existe plusieurs variantes de la théorie de l'activité mais celles développées par Leont'ev et Engeström sont les plus utilisées [25, 24]. Aussi, nous nous intéressons seulement à ces deux variantes.

Dans la suite, nous allons présenter les différentes générations de la CHAT et comment cette dernière peut être utilisée dans le domaine des technologies de l'information et de la communication (TIC).

2.2.1 La Première génération de la CHAT

La théorie de la CHAT est née de la volonté de Vygotsky d'identifier des méthodes objectives pour étudier et expliquer les activités humaines [26]. Ainsi avec ses collègues, ils décrivent un espace métaphorique représentant le cadre du développement cognitif, un espace occupé par des sujets, des experts et d'autres agents capables de contribuer au développement. Ce cadre considère l'action humaine (activité) comme étant en relation avec des objets tels que des outils qui transportent avec eux l'histoire culturelle de l'humanité. Dans cette théorie, le développement individuel ne peut pas être compris sans faire référence au contexte social dans lequel il est intégré. Les humains n'interagissent pas directement avec les objets de l'environnement. Cette interaction est faite grâce à des moyens culturels, des outils et des signes. L'action humaine a une structure tripartite (Fig. 2.1); les pontillés signifient qu'il n'existe pas de relation directe entre *Sujet* et *Objet*.

Vygotsky et ses collègues introduisent un objet intermédiaire (Artefact) entre l'action de l'homme (activité) et son environnement. L'artefact est un concept pour expliquer le processus sémiotique qui permet le développement de la conscience humaine à travers l'interaction avec son milieu [26]. Dans cette interaction, l'homme n'est pas un acteur passif mais grâce à ses activités, il participe à la transformation de son environnement. Ainsi dans la Figure 2.1, *Sujet* représente un individu engagé dans un processus (activité) et la notion d'*Artefact* représente des outils psychologiques produits par l'évolution socioculturelle à laquelle les humains ont accès en étant actifs dans leur communauté et non dans l'isolement. Comme l'a fait remarqué Yamagata-Lynch [26], il existe plusieurs significations du mot russe traduit par *Objet*. Ainsi, *Objet* peut-être le but de l'activité, les motivations pour

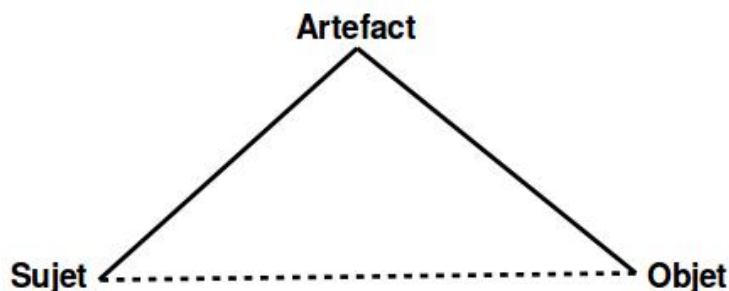


Figure 2.1: Triangle de médiation de Vygotsky [17].

participer à une activité ou les matériels produits ou utilisés durant une activité.

Pour mettre en pratique sa théorie, Vygotsky [17] l'applique au potentiel d'apprentissage des enfants avec la collaboration d'un adulte ou d'un pair pour la résolution d'une tâche. Ainsi, il définit la zone proximale de développement (ZPD) comme:

«the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers»
([17] p. 86).

Ainsi, la ZPD est un environnement qui permet d'élaborer comment les individus interagissent avec leur environnement, y compris avec les objets sociaux et d'autres ressources. Par conséquent, la ZPD est un lieu d'apprentissage qui permet à l'individu d'améliorer ses capacités grâce à ses interactions avec les autres membres. Ainsi, selon Vygotsky le développement individuel provient de sources sociales. Comme les enfants participent à des activités et travaillent ensemble, ils acquièrent plus de connaissances sur le monde et ses nombreuses cultures.

Dans le processus décrit par Vygotsky, seules les connaissances du *Sujet* sont analysées et non celles du groupe de façon globale, c'est-à-dire le gain pour le groupe à partir de ces différents échanges.

2.2.2 La Seconde génération de la CHAT

La seconde génération de la CHAT est souvent appelée théorie de l'activité. Cette théorie est un cadre conceptuel issu des travaux de Vygotsky et de ses collègues. La seconde génération de la CHAT est venue après le décès de Vygotsky et concerne les travaux de Leont'ev [18] appelés *théorie de l'activité*. Les travaux de Leont'ev sont centrés sur l'explication et la compréhension d'un point de vue psychologique de comment les besoins et les actes de l'individu peuvent être considérés comme une seule unité d'analyse et comment l'interaction entre les deux peuvent avoir un impact

à la fois sur l'individu et sur l'environnement. Ainsi selon Leont'ev, la théorie de l'activité implique une interaction entre le sujet, l'objet, la motivation, l'action, les objectifs, le contexte historico-socioculturel et les conséquences de l'activité [26]. Le modèle de Vygotsky est souvent considéré comme un processus et celui de Leont'ev comme une série de processus.

Ainsi dans [18], Leont'ev explicite la différence cruciale entre une action individuelle et une activité collective. Dans son célèbre exemple de *primeval collective hunt* [18] (p. 87), Leont'ev explique comment une action collective pourrait se dérouler. Dans cet exemple, il explique l'action d'un rabatteur dans une chasse dont le rôle se limite à faire peur à un troupeau et à guider les proies vers les autres chasseurs cachés en embuscade. La seconde phase de cette chasse sera l'activité des chasseurs. Ainsi pour le rabatteur, l'*activité* est la chasse et l'*action* est de faire peur. Les chasseurs ont la même *activité* que le rabatteur mais leur *action* est de tuer la (les) proie(s). L'objectif final (tuer une proie) de cette activité est d'avoir de quoi manger. Cette activité est scindée en deux étapes : 1) l'action du rabatteur (faire peur) et 2) celle des chasseurs (tuer la proie). Cet exemple montre une autre différence avec l'approche de Vygotsky dont l'unité d'analyse est l'action individuelle et non l'action collective.

Ainsi Leont'ev propose un modèle d'activité à trois niveaux reliés. Le niveau le plus élevé est celui de l'activité qui est défini par les objectifs (*avoir de la nourriture* dans l'exemple); le niveau moyen est celui de l'action individuelle ou collective pour atteindre les objectifs (*faire peur* pour le rabatteur et *tuer la proie* pour les chasseurs); et le bas niveau est celui des opérations automatiques liées aux conditions et aux outils à portée de main (*cuisiner et manger*).

Cependant, Leont'ev n'a pas défini de formalisme clair de la médiation sociale entre l'homme et le monde pour tenir compte des relations sociales.

2.2.3 La Troisième Génération de la CHAT

La troisième génération est liée aux travaux menés par Engeström [16, 19]. Engeström propose un modèle d'activité collective comme Leont'ev et son modèle est le résultat d'une prolongation en deux étapes du concept original d'activité de Leont'ev. Le modèle de Leont'ev est compris comme une interaction *Sujet - Objet* dans le cas d'une activité collective [24]. Alors que le modèle d'Engeström propose trois niveaux de médiation entre *Sujet - Objet - Communauté*.

Le modèle d'Engeström [16, 19] est représenté comme un diagramme triangulaire (Fig. 2.2). Le premier niveau de médiation qui est le triangle supérieur (Fig. 2.2) reprend le modèle de médiation de Vygotsky; *Instruments* représente les objets sociaux, les outils et les symboles qui mettent en relation le *Sujet* qui représente un individu ou groupe d'individu et *Objet* qui représente l'objectif ou la motivation de l'activité. *Règles* représentent la médiation au travers des conventions, les habitudes, les normes qui régulent les interactions et les actions entre *Sujet* et *Communauté*. *Division du Travail*, la dernière médiation entre *Communauté* et *Objet*, représente la répartition des actions (tâches) et des rôles au sein de la communauté pour atteindre

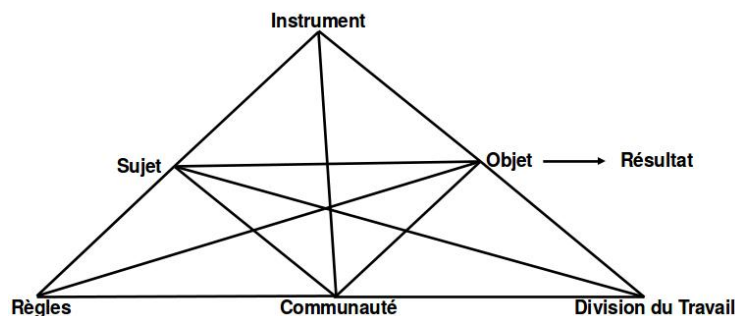


Figure 2.2: Modèle du système d'Activité d'Engeström [16].

l'objectif de l'activité. En outre, le modèle inclut le *Résultat* qui est le résultat de l'activité dans son ensemble. C'est une transformation de l'objet produit par l'activité en question par rapport au résultat recherché, qui peut être utilisé par d'autres systèmes d'activité.

Comme l'a montré Kaptelinin [27], la différence entre l'approche de Leont'ev et celle d'Engeström est très subtile et porte sur quatre points. Les deux premières différences concernent le niveau de la portée et de l'exécution des activités de l'individu chez Leont'ev bien qu'elles puissent être réalisées par la communauté. Chez Engeström elles sont exclusivement associées à la communauté. Le troisième point porte sur le but de l'activité qui est lié aux motivations/besoins chez Leont'ev alors que chez Engeström il est lié au résultat (*Résultat*). Le dernier point est lié au domaine d'application qui est la psychologie dans l'approche de Leont'ev et l'ensemble des organisations qui peuvent subir des changements dans l'approche d'Engeström. Ce point est un principe essentiel pour l'approche d'Engeström dans laquelle les systèmes d'activité se développent constamment et le résultat d'une activité (*Résultat*) peut être utilisé par une autre activité. Bien qu'il existe une certaine différence entre le modèle de Leont'ev et celui de Engeström, dans la pratique les deux modèles sont confondus et on y fait référence comme le modèle Leont'ev-Engeström [25, 27].

2.2.4 Utilisation de la CHAT comme modèle

La CHAT a été utilisée dans différents domaines pour modéliser l'apprentissage collaboratif, la co-construction de connaissances et l'Interaction Homme-Machine (IHM). Ainsi, en fonction de l'objectif soit l'approche de Vygotsky est utilisée soit celle de Leont'ev-Engeström.

John-Steiner et Mahn [23] utilisent trois principes centraux de l'approche de Vygotsky pour examiner la relation entre l'apprentissage et le développement: (a) les sources sociales du développement individuel, (b) la sémiotique (les signes, les symboles, la langue, les techniques mnémoniques, etc), la médiation dans le développement humain, et (c) l'analyse du développement. À partir de ses trois axes, ils

présentent des travaux où l'approche de Vygotsky est utilisée dans le domaine de l'éducation (apprentissage, alphabétisation) et expliquent comment la CHAT pourrait aider les enseignants à donner des cours dans une classe où il existe une diversité culturelle et linguistique.

Dans ses recherches sur la *Mémoire Humaine*, Brown et ses collègues élaborent et mettent en œuvre des programmes éducatifs basés sur le concept de l'apprentissage en utilisant l'approche de Vygotsky [28, 29]. Dans ces études, ils considèrent une salle de classe comme une ZPD (Zone Proximale de Développement) qui est composée de personnes avec différents degrés d'expertise (adultes et enfants) mais aussi des artefacts comme des livres, des vidéos, des équipements scientifiques, etc. Cette ZPD leur a permis de déterminer deux caractéristiques de l'apprentissage collaboratif: 1) l'enseignement réciproque (*Reciprocal Teaching*) qui est un processus répétitif qui permet de renforcer le discours de l'étudiant qui est appelé à jouer le rôle du professeur dans la classe; 2) la méthode *jigsaw* qui est une méthode qui affecte aux étudiants un sujet qu'ils traitent et ensuite qu'ils enseignent à d'autres via l'enseignement réciproque. Ces études montrent comment l'approche socioculturelle permet la co-construction de connaissances.

La CHAT est aussi utilisée dans l'enseignement à distance. Teemant a développé un système, *ProfessorsPlus*, qui intègre l'interaction sociale entre les participants et une assistance interactive qui adapte le contenu du cours en fonction du groupe d'apprenants [30]. Les résultats de cette étude ont montré comment un programme de formation à distance adoptant les perspectives socioculturelles peut augmenter les connaissances, les compétences et les dispositions des enseignements au profit des étudiants qui ont l'anglais comme deuxième langue [30].

Cole présente quatre exemples de recherche dans le cadre de la CHAT [20]. Ces exemples étudient les changements qualitatifs qui se produisent dans l'interaction selon différents niveaux de contrainte dans différents contextes sociaux (enseignement dyadique, une leçon en classe, tissage instruction dans un village mexicain et la lecture dans de petits groupes). Ces quatre cas d'études montrent comment la CHAT pourrait promouvoir ou restreindre l'apprentissage grâce à l'interaction sociale dans le développement de la résolution de problèmes (*problem solving*).

Palincsar [31] montre comment l'approche socioculturelle peut être utilisée pour aider des individus à constituer une communauté de discours dans le processus de re-acculturation. L'acculturation est l'ensemble des phénomènes qui résultent de l'interaction sociale entre des groupes d'individus de cultures différentes. Dans cette étude, la perspective socioculturelle a permis de comprendre les processus et les résultats de l'apprentissage collaboratif dans des contextes d'apprentissage à partir de problèmes (*problem-based learning*) [31].

L'approche de Leont'ev-Engeström est plutôt utilisée dans le domaine de l'Interaction Homme-Machine (IHM) [24, 25, 32, 33, 34].

L'impact de la CHAT sur l'IHM et la conception d'interaction porte, pour l'essentiel, sur trois points [24]. En premier, la théorie a offert quelques aperçus théoriques généraux qui proposent un cadre conceptuel plus riche. Ainsi, Kuutti

[25] explique quelques problèmes de l'IHM de l'époque et propose une approche basée sur la CHAT comme un moyen pour améliorer la situation. Plus récemment, Bjørkli et ses collègues [33] adoptent quelques uns des principes décrits par Kuutti [25] pour élaborer un cadre de soutien dans un processus de conception d'équipement de navigation pour des patrouilleurs de la marine norvégienne.

Le deuxième apport porte sur un cadre analytique pour la conception et l'évaluation des systèmes interactifs. Koszalka et Chun-Ping [32] proposent une approche interactive, basée sur la CHAT, qui met l'accent sur les relations dynamiques entre les individus, les objectifs, les outils, les membres de la communauté et les facteurs médiateurs. Ainsi, ils appliquent ce processus au projet IDEAS² de la NASA³. Ce processus leur a permis, avec des données à l'appui, d'améliorer leur technique d'enseignement et d'apprentissage en répondant mieux aux besoins des astronautes [32].

Le dernier apport a été mis en évidence ces dernières années, avec la création de nouveaux systèmes et la mise en œuvre d'activités centrées sur l'informatique. Bødker et Klokrose dans le cadre d'une société de production de PDA⁴ et de logiciels [34] en sont un exemple. Leur objectif était de concevoir un nouvel artefact portatif pour des cartes géographiques numériques. Pour ce faire, un modèle d'interaction *homme-artefact* a été utilisé pour structurer l'analyse de la carte et les possibilités de conception.

La CHAT est aussi utilisée dans d'autres domaines de recherche: les systèmes d'information, des sciences de l'information, le développement de bibliothèque numérique ou la recherche d'information interactive. Wilson présente des exemples dans chacun de ces domaines où la CHAT a été appliquée avec succès [35].

Cette sous section nous a permis de montrer le dynamisme qui existe autour des travaux sur la CHAT, une approche de modélisation qui peut être adaptée en fonction du domaine d'étude. Il faut noter que l'approche de Vygotsky est plus appropriée pour les processus d'apprentissage et de co-construction de connaissances bien que le modèle de Leont'ev-Engeström puisse être utilisée dans le cadre de l'apprentissage [32]. Cette dernière est plus utilisée dans le domaine de l'IHM.

Malgré les critiques de la CHAT [36], celle-ci a connu une croissance plutôt spectaculaire entre 2000 et 2005 (fréquence des citations dans des revues internationales [37]). Sannino et ses collègues ont montré comment la CHAT est à la fois une théorie fondée solidement dans la pratique et orientée vers l'avenir [38].

2.3 Les Communautés Virtuelles

Récemment avec le développement des Technologies de l'Information et de la Communication (TIC), les individus utilisent de plus en plus l'Internet pour collaborer

²Initiative to Develop Education through Astronomy and Space Science

³National Aeronautics and Space Administration

⁴Assistant Numérique Personnel

à travers des systèmes comme les groupes de discussion, les listes de diffusion, les réseaux sociaux, les systèmes de chat, etc. Grâce à l'Internet, ces systèmes créent un nouveau phénomène appelé **communauté en ligne** ou **communauté virtuelle**⁵. Le mot communauté vient du mot latin **Communitas** où **cum** signifie *ensemble* et **munus** *devoir*. Le mot *en ligne* ou *virtuelle* est utilisé en opposition avec la forme traditionnelle des communautés où les rencontres se font face-à-face. Cependant dans la littérature, il n'existe quasiment pas de consensus pour définir ou classifier les communautés virtuelles. Ces nombreuses définitions résultent des différentes perspectives de recherches des chercheurs en sociologie, anthropologie, psychologie. Les différentes classifications sont le résultat des différents axes de classification: démographique, en fonction des caractéristiques des membres, environnement technologique. Même si les termes sont différents d'une définition à une autre, le sens reste parfois le même.

L'objectif principal de cette sous section est de situer les communautés utilisatrices du projet WestAfricapedia au sein de la littérature. Comme présenté dans l'introduction, l'ambition du projet est de fournir un cadre de partage à ces communautés pour qu'elles puissent vulgariser, pérenniser mais également partager leur patrimoine socioculturel à travers le Web. Aussi, après avoir fait une revue de la littérature sur les classifications des communautés virtuelles, nous situons nos communautés de partage de connaissances socioculturelles dans les classifications de la littérature. Nous présentons aussi les différentes caractéristiques qui poussent les individus à rejoindre une communauté virtuelle afin de justifier les services que propose notre plateforme pour aider les communautés à se constituer et à exister.

Cette sous section est organisée de la façon suivante. Un état de l'art est fait sur les définitions des communautés virtuelles en montrant l'évolution qu'il y a eu concernant cette notion avec l'évolution des TIC. Ensuite, nous présentons un état de l'art des typologies sur les communautés : les typologies générales, les typologies en fonction de l'objectif de la communauté, une typologie en fonction de la transmission de savoir au sein de la communauté et une typologie en fonction de la structure formelle des communautés. En dernier lieu, nous présentons les caractéristiques qui permettent d'évaluer le succès des communautés virtuelles.

2.3.1 Définition d'une Communauté virtuelle

L'utilisation et la diffusion des Technologies de l'Information et de la Communication (TIC) offrent la possibilité de redéfinir et de repenser la *communauté* à la fois en termes de délimitation des frontières et en termes de modes de communication utilisés entre les membres. Les TIC offrent la possibilité d'une communication généralisée pour le public et relativement facile d'accès. De même, les TIC ont permis un nouveau phénomène appelé communauté virtuelle basé sur des ressources réparties géographiquement et permettant à ses membres des échanges et une production d'information sans qu'il y ait de contrainte de proximité géographique pour

⁵Dans la littérature on rencontre d'autres appellations comme communauté électronique. Dans la suite du document, nous utiliserons l'expression *communauté virtuelle*.

ses membres (alors que c'est le cas des communautés traditionnelles). Ainsi, des systèmes comme les forums de discussion, les listes de diffusion, les réseaux sociaux, les systèmes de chat sont des exemples d'outils permettant à un ensemble de personnes de collaborer sans qu'elles aient besoin de se rencontrer physiquement. Ainsi la proximité géographique des communautés traditionnelles n'est plus une contrainte pour ce nouveau phénomène appelé *communauté virtuelle ou en ligne* qui ne s'articule que sur la proximité des intérêts (même centres d'intérêt).

Bien qu'il soit relativement facile de reconnaître une communauté virtuelle, il n'existe pas de consensus pour la définir. Ces différences de point de vue émanent surtout des pluralités disciplinaires dans les études des communautés et également de l'évolution des TIC. En effet, les premières définitions mettaient l'accent sur les activités coopératives assistées par ordinateur (CSCW⁶) pour l'interaction entre les utilisateurs mais avec l'apparition de nouvelles technologies comme les *smartphones* ces définitions sont devenues caduques.

2.3.1.1 Définition de la littérature

White [39] répertorie une dizaine de définitions d'une communauté virtuelle qui proviennent de différentes perspectives : commerciale, sociologique, anthropologique et psychologique. Parmi ces définitions, celles de Rheingold [40], Hagel et Armstrong [41] et de Jones et Rafaeli [42] sont les plus populaires dans la littérature [43]. Le premier à utiliser le terme de communauté virtuelle a été Rheingold [40] dans son livre *The virtual community: Homesteading on the electronic frontier* où il définit les communautés virtuelles comme des :

« social aggregations that emerge from the Net when enough people carry on those public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace. »

La définition de Rheingold met l'accent sur l'utilisation de l'Internet (*Net*) pour ainsi différencier les communautés virtuelles des communautés traditionnelles et il définit l'Internet (*Net*) comme un réseau d'ordinateurs inter-connectés en utilisant les communications médiatisées par ordinateur (CMC) pour relier les individus partout dans le monde. Il met aussi en évidence le fait qu'il y ait une discussion publique entre les membres autour d'un intérêt commun et que le contenu soit généré par les membres et non par l'administrateur de la communauté (*webmaster*).

Pour Hagel & Armstrong [41] les communautés virtuelles :

« ... are computer-mediated spaces where there is a potential for an integration of content and communication with an emphasis on member-generated content. »

⁶CSCW: Computer Supported Cooperative Work.

Dans leur livre sur les affaires (*business*) liées aux communautés virtuelles, *Net Gain* [41], Hagel et Armstrong proposent cinq éléments qui définissent une communauté virtuelle : 1) un sujet de discussion spécifique qui identifie une communauté virtuelle, 2) la publication de contenu qui est intégrée avec la communication entre les membres, 3) la diffusion et la valorisation du contenu, 4) l'ouverture à une information plurielle (chaque membre a la possibilité de participer à la construction de l'information) grâce à un grand nombre de ressources de haute qualité et 5) une orientation de plus en plus commerciale. Ces cinq éléments sont assistés par les CMCs (computer-mediated communication).

Quant à Jones & Rafaeli [42], ils utilisent l'expression *public virtuel* au lieu de communauté virtuelle :

« Virtual Publics are symbolically delineated computer-mediated spaces whose existence is relatively transparent and open, that allow groups of individuals to attend and contribute to a similar set of computer-mediated interpersonal interactions. »

Le terme *public virtuel* est utilisé pour éviter la confusion qui entoure le terme *communauté* qui est selon eux *évasif* et *vague*. Un public virtuel, qui n'est pas un sous ensemble d'une communauté virtuelle, s'inscrit dans une approche différente pour catégoriser les cyberespaces⁷ (liste de diffusion, bulletin électronique⁸, vidéo conférence, etc.). Comme pour les deux définitions précédentes, les publics virtuels sont formés par un regroupement d'individus qui contribuent à une série d'interactions assistées par ordinateur (CMC).

En préconisant toujours l'utilisation des CMCs comme moyen d'interaction, Lee et ses collègues [43] définissent une communauté virtuelle comme :

« a cyberspace supported by computer-based information technology, generate centered upon communication and interaction of participants to member-driven contents, resulting in a relationship being built up. »

D'après ces quatre définitions le média utilisé et l'interaction sont les caractéristiques principales d'une communauté virtuelle. Un aspect important des communautés virtuelles réside dans le fait que le contenu ou le sujet de discussion soit créé par les participants grâce à l'interaction entre eux ce qui les distingue des services d'information en ligne. Les communautés virtuelles sont également caractérisées par l'utilisation des CMCs pour soutenir les activités en leur sein.

⁷D'après le Petit Robert, un cyberspace est un « ensemble de données numérisées constituant un univers d'information et un milieu de communication, lié à l'interconnexion mondiale des ordinateurs. »

⁸Bulletin board

Bien que ces définitions intègrent plusieurs aspects des communautés virtuelles comme la production de contenu par les membres, elles n'intègrent pas l'évolution des TIC. En effet, toutes ces définitions mettent l'accent sur l'utilisation des ordinateurs pour assister l'interaction entre utilisateurs alors qu'avec l'évolution des TIC, des outils comme les *smartphones* influent sur les types de contenu et la façon de les créer et de les faire évoluer. De plus comme le montre Porter [44], la plupart des définitions s'orientent vers les communautés d'individus alors que dans le domaine du marketing on parle plutôt de relations entre individus et/ou organisations. Ainsi la définition qui semble prendre tous ces aspects en considération est celle de Porter [44] qui définit une communauté virtuelle comme :

« an aggregation of individuals or business partners who interact around a shared interest, where the interaction is at least partially supported and/or mediated by technology and guided by some protocols or norms. »

Dans cette définition, Porter ajoute l'aspect commercial (*business*) des communautés virtuelles mais inclut aussi les situations où la proximité permet des rencontres face-à-face entre les membres. Elle ajoute aussi le respect de certaines normes qui n'apparaît pas dans les définitions précédentes et qui est intrinsèque à toute communauté pour sa bonne marche. Contrairement aux définitions précédentes, elle n'utilise pas le terme CMC mais le mot technologie qui peut englober tous les nouveaux outils des TIC utilisés pour l'interaction des membres au sein d'une communauté virtuelle.

2.3.2 Typologie des Communautés Virtuelles

Comme pour la définition, il n'existe pas de typologie consensuelle sur les communautés virtuelles. Les différentes propositions reflètent des opinions divergentes en fonction des principes sous-jacents, des objectifs visés ainsi que des perspectives de recherche : sociologique, psychologique ou anthropologique.

2.3.2.1 Typologies générales

★ Typologie de Lazar & Preece (1998)

Lazar & Preece [45] proposent une classification des communautés virtuelles basée sur quatre axes :

1. Leurs caractéristiques (*attributes*) : ce sont les caractéristiques que possède la communauté comme par exemple l'objectif ou l'intérêt commun des membres, les activités partagées par les membres, l'accès aux ressources partagées, le soutien entre les membres, les conventions sociales et la taille de la communauté.
2. Le mode d'adhésion (*boundedness*) : il définit la cohésion sociale au sein d'une communauté. Si c'est une communauté dans laquelle les liens sont étroits

comme dans un intranet par exemple, les relations sociales et la communication prennent place parmi les membres. Si c'est une communauté dans laquelle les membres ont peu de liens comme par exemple sur le Web, alors les membres de la communauté ont plus de liens sociaux avec des individus qui sont en dehors de la communauté.

3. La relation aux communautés physiques : cette caractéristique décrit les rencontres physiques, si elles existent, entre les membres de la communauté. Ainsi, on peut avoir des communautés virtuelles basées sur des communautés physiques qui sont généralement des communautés virtuelles géographiquement concentrées. Il y a aussi les communautés virtuelles qui ne sont pas entièrement basées sur une zone géographique où les membres peuvent se réunir physiquement périodiquement. Enfin, il y a des communautés virtuelles entièrement en ligne dont les membres sont dispersés géographiquement ou souhaitent garder leur anonymat.
4. Le support logiciel : cette caractéristique porte sur la technologie logicielle utilisée pour maintenir l'interaction entre les membres d'une communauté virtuelle. La communication au sein d'une communauté virtuelle peut être maintenue grâce par exemple à des listes de diffusions, des bulletins électroniques, des discussions relayées par Internet (Internet Relay Chat - IRC), etc.

★ Typologie de Markus (2002)

Markus [46] propose une classification sans chevauchement basée sur l'orientation sociale, l'orientation professionnelle et l'orientation commerciale des communautés virtuelles. Comme le montre plus précisément la figure 2.3, le premier niveau de distinction est basé sur l'orientation des communautés virtuelles c'est-à-dire l'objectif principal des membres en rejoignant la communauté. L'orientation sociale est historiquement le premier critère de classification de communauté virtuelle à avoir été identifié [46]. L'orientation sociale est subdivisée en deux sous catégories : 1) création de relation (*relationship building*) - dont le premier objectif est d'établir un lien social avec les membres de la communauté virtuelle et 2) divertissement (*entertainment*) qui répond à un besoin personnel plus qu'à celui d'un groupe bien que des relations puissent être construites au sein d'une communauté virtuelle de divertissement. Le deuxième critère de distinction des communautés virtuelles est relative à l'orientation professionnelle. Certaines communautés regroupent des professionnels qui veulent échanger des informations à leur travail au sein de leur organisation. Les communautés virtuelles professionnelles sont subdivisées en deux sous catégories : 1) les réseaux d'apprentissage (*learning networks*) pour acquérir des connaissances sur un nouveau sujet ou dans un domaine d'intérêt; nous les rencontrons surtout dans le domaine de l'éducation et de la formation et 2) les réseaux d'experts (*expert networks*) pour étendre, développer et documenter les connaissances existantes. La dernière catégorie de communauté virtuelle est celle à orientation commerciale dont l'objectif principal est de réaliser un profit ou un avantage financier. Cette catégorie est subdivisée en deux sous catégories : 1) entre

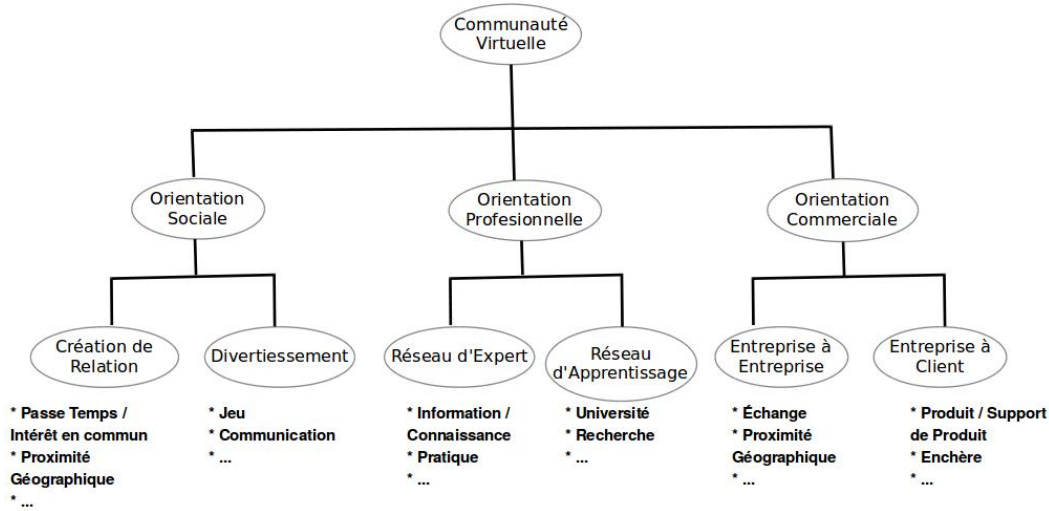


Figure 2.3: Caractérisation des communautés virtuelles [46].

entreprises (*business-to-business: B2B*). Cette catégorie regroupe des entreprises qui veulent collaborer pour élaborer des stratégies communes et trouver des partenaires commerciaux appropriés et 2) entreprise-à-client (*business-to-consumer: B2C*). Cette catégorie est utilisée par les entreprises pour faire la promotion d'un produit ou d'une marque et les membres (clients) adhèrent à ce type de communauté pour avoir des informations sur les produits mais aussi pour accéder aux commentaires des utilisateurs (autres membres) sur ces produits.

★ Typologie de Porter [44]

Pour caractériser une communauté virtuelle, Porter suggère l'utilisation de cinq caractéristiques appelées les cinq P des communautés virtuelles [44] :

1. Objectif (contenu de l'interaction - *Purpose*) : décrit le contenu central de la communication au sein de la communauté. Cette caractéristique fournit la raison d'être d'une partie de la communauté [45].
2. Place (*Place*) : définit comment se passe l'interaction et où elle se produit c'est-à-dire complètement ou partiellement en ligne. C'est le même critère que celui de la relation aux communautés physiques dans [45].
3. Plateforme (*Platform*) : fait référence au support technique de l'interaction au sein de la communauté virtuelle c'est-à-dire est-ce que les supports permettent la communication synchrone, asynchrone ou les deux ?
4. Population (modèle d'interaction - *Population*) : fait référence au schéma d'interaction entre les membres de la communauté, c'est-à-dire à la structure du groupe et aux types de liens sociaux. La structure du groupe représente la

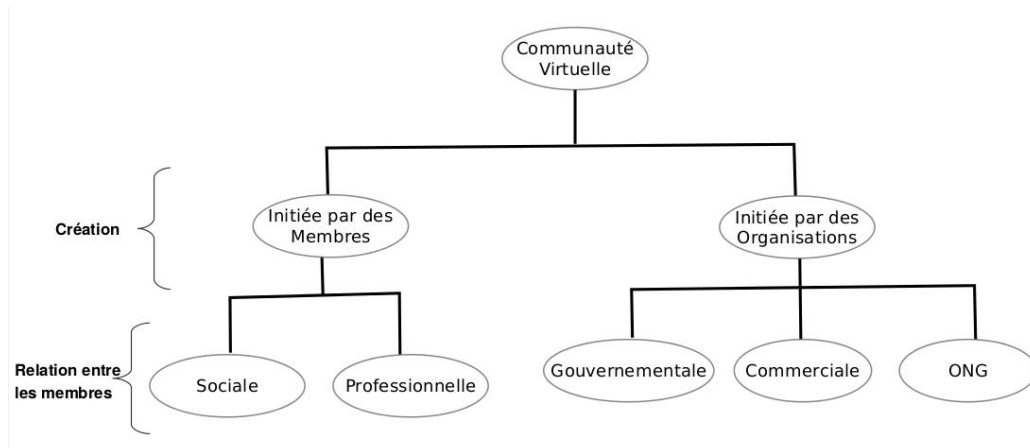


Figure 2.4: Typologie des communautés virtuelles [44].

taille de la communauté, appelé *effet de seuil* dans [45], qui motive certains utilisateurs à adhérer à une communauté en fonction de sa taille et de liens sociaux. Le type de lien social décrit la force des relations entre les membres (exemple : forte, faible, stressante).

5. Modèle économique (*Profit Model*) : s'applique le plus souvent aux communautés commerciales. Il est possible que d'autres types de communautés puissent engendrer une valeur économique. Cela représente le gain économique que les membres peuvent avoir en participant aux activités de la communauté.

Porter [44] propose une typologie (figure 2.4) basée sur une revue de la littérature et sur les observations de la conception et de l'activité dans les communautés virtuelles réelles. Ainsi, les communautés virtuelles au sein d'une catégorie donnée sont susceptibles de partager des caractéristiques communes, mais aucune communauté individuelle ne possède toutes les caractéristiques décrites précédemment. Ainsi, la typologie proposée par Porter utilise le mode de création et l'orientation de la relation des communautés virtuelles qu'elle évalue ensuite avec les cinq P.

La typologie est composée de deux niveaux ; le premier où la catégorisation est faite suivant le mode de création de la communauté. Les communautés *Member-Initiated* sont celles créées et maintenues par les membres contrairement aux communautés *Organization-Sponsored* créées par des organismes (gouvernement, ONG, sociétés, etc.). Le deuxième niveau est formé par les sous catégories de : 1) *Member-Initiated* qui sont les communautés virtuelles à orientation sociale et professionnelle et 2) *Organization-Sponsored* qui sont parrainées par une organisation commerciale, non-commerciale ou gouvernementale.

2.3.2.2 Typologies fonction de l'objectif

* Typologie de Hagel & Armstrong (1997)

Hagel et Armstrong [41] se concentrent sur les facteurs qui interviennent au niveau de l'interprétation sociale et proposent une classification dans une perspective sociologique. Autrement dit, ils se concentrent sur l'opinion que les gens ont sur les autres et la satisfaction des besoins des membres. Ainsi, ils classent les communautés virtuelles suivant les quatre besoins fondamentaux de leurs membres :

1. Fantaisiste/imaginaire : une communauté virtuelle fantaisiste/imaginaire consiste en une communauté où les membres interagissent dans le but de se divertir ; Ultima Online⁹ est un bon exemple de communauté virtuelle animée sur Internet.
2. Intérêt : une communauté virtuelle d'intérêt est articulée autour d'une thématique et permet le partage d'information/connaissances. Cependant une bonne partie de ses membres ne participent qu'épisodiquement.
3. Relationnel : pour les communautés virtuelles relationnelles, l'objectif principal est d'établir un lien entre les membres ayant vécu des expériences similaires ; l'engagement envers le groupe est beaucoup plus fort que pour les autres types de communautés.
4. Transactionnel : une communauté virtuelle transactionnelle est une communauté où les membres échangent des renseignements pour faciliter l'achat ou la vente de produits ou de services. Ils subdivisent la catégorie transactionnelle en deux sous catégories : B2C et B2B comme celle de Markus [46].

★ **Typologie de Jones & Rafaeli (2000)**

Jones et Rafaeli [42] ont mis au point trois classifications selon des objectifs différents :

1. *Use* : cette classification peut être appliquée pour l'étude des problèmes sociaux et psychologiques dans les communautés virtuelles [43]. Ils adoptent la classification de Hagel et Armstrong [41] qui répond aux quatre besoins des utilisateurs. Cependant, ils subdivisent le B2C en trois types - géographique, démographique et topique - et le B2B en quatre sous catégories - industries verticales (par exemple, développement de logiciel), fonctionnelle (par exemple, marketing), géographique (par exemple, association d'entreprises locales), ou type de l'entreprise publics virtuels (par exemple, les petites entreprises).
2. *Social structure* qui est une classification sociologique des réseaux sociaux qui sont plus spécifiques que certaines communautés virtuelles. La classification est faite selon trois catégories - *virtual-settlements* où une distinction est faite entre réseau social et communauté, *virtual airport bars* où les interactions sociales sont plus éphémères et *cyber-Inns virtual* où la construction de communautés est possible.

⁹<http://www.uo.com/>

3. *Technologie* : cette classification est faite en fonction des caractéristiques technologiques des applications utilisées. Ainsi on retrouve les listes de diffusion, IRC, le groupe USENET, le système de bulletin électronique et d'autres technologies de CMC.

Suite à ces trois classifications, Jones et Rafaeli [42] suggèrent qu'une classification des communautés virtuelles devrait traiter au moins quatre aspects :

1. La thématique des échanges entre les membres ;
2. Les profils des utilisateurs : les informations démographiques (taille de la communauté) et géographiques ;
3. La structure sociale : la force des liens sociaux ;
4. La technologie et les fonctionnalités qui facilitent les échanges entre les membres (thématique).

2.3.2.3 Classification en fonction de la transmission de savoir

Toujours suivant l'orientation sociale au sein des communautés virtuelles et le type d'échange de savoir entre les membres, nous rencontrons dans la littérature plusieurs typologies. Bien que la nomenclature proposée change d'une proposition à une autre, les propositions peuvent être représentées de la façon suivante¹⁰ :

- Communautés d'Intérêts (CoI) [47, 48] : elles sont essentiellement thématiques. Leurs membres participent à la communauté en vue d'échanger des informations se rapportant à la vie ou à l'activité professionnelle afin d'obtenir des réponses à des questions ou à des problèmes personnels. De plus, l'objectif n'est pas la production collective ni l'usage collectif. Les membres de ces communautés ne s'imposent pas de partager leurs connaissances systématiquement. Les CoI sont des sous catégories des communautés virtuelles sociales au sens de Porter [44].
- Communautés d'Intérêts Intelligentes¹¹ (CoI2) : dans la littérature, on trouve d'autres dénominations comme Communauté d'intérêt de But¹² [47], Communauté d'intérêts basée sur les tâches¹³ [48]. Ce sont des communautés dont l'objectif est de répondre à un besoin ciblé, de résoudre un problème, de réaliser un projet. Les membres de ces communautés ont souvent des compétences diverses et collaborent pour atteindre le but fixé. Les CoI2 sont des sous catégories de communautés virtuelles professionnelles au sens de Porter [44].

¹⁰Nous avons proposé ce regroupement

¹¹Nous avons proposé cette appellation.

¹²Goal-oriented community of interest

¹³task-based community of interest

- Communautés d'Apprenants (CoA) [47, 48] : ce sont des groupes de personnes qui partagent des valeurs et convictions et qui s'engagent activement dans un processus d'apprentissage. Au sein de ce type de communauté, il existe un (des) expert(s) qui partage(nt) ses connaissances avec un ensemble d'apprenants. On les rencontre souvent dans le domaine de l'éducation. Les CoA sont des sous catégories de communautés virtuelles professionnelles au sens de Porter [44].
- Communautés de Pratiques (CoP) : ce sont la plupart du temps des communautés professionnelles qui s'orientent vers la réussite d'une activité, ce qui peut créer une connaissance qui est essentiellement le *savoir faire*. Pour de plus amples informations, se référer à [49].
- Communauté de co-élaboration de connaissances¹⁴¹⁵ : ce concept a été développé par l'école de Toronto dans [50, 51] mais s'applique seulement dans le domaine de l'éducation. Ainsi dans le contexte éducatif, les auteurs le considèrent comme une stratégie pédagogique socio-constructiviste où le développement de la connaissance ne se fait pas simultanément mais de classe en classe. Les communautés de développement des connaissances visent à favoriser la connaissance des membres collectivement, tout en soutenant la croissance individuelle dans le but de produire de *nouveaux experts* au sein du domaine de la communauté. Puisque cette communauté est au sein de l'éducation, un rôle important est donné à l'*expert* (enseignant) qui est appelé à favoriser l'apprentissage.

Mis à part le type de transmission du savoir, ces communautés virtuelles peuvent être classées en fonction de la force des liens sociaux entre leurs membres. Les liens sociaux sont les affiliations, les relations qui unissent les membres entre eux. Ainsi au sein d'une CoI, la collaboration est occasionnelle ce qui fait que les liens sociaux sont faibles. Les liens sociaux sont plus forts dans une CoI2 que dans une CoI bien qu'ils restent assez faibles ; la force par rapport à une CoI est surtout liée à l'affiliation de ses membres. De la CoA à la Communauté de développement/renforcement de connaissances l'engagement, la participation et l'entraide gagnent en intensité et les liens sociaux deviennent plus forts. Ainsi, les liens sociaux d'une communauté peuvent se développer et se préciser au fil des interactions entre ses membres.

2.3.2.4 Classification suivant la structure formelle (adhésion)

L'adhésion peut être utilisée pour classer les différents types de communautés. La conception, la ou les actions des membres de la communauté, peuvent déterminer le niveau d'accessibilité des informations au sein des communautés. La restriction peut

¹⁴Knowledge-building community.

¹⁵Une conversation entre Stéphane Allaire et Gaston Mialaret lors d'un colloque sur les communautés d'apprentissage à Barcelone avait conduit à la considération du terme Communauté de co-élaboration de connaissances en milieu francophone.

être décidée pour des questions de sécurité (entreprises), pour cibler des personnes qui vivent dans une zone géographique spécifique, etc. Ainsi Mitchell [53] propose une catégorisation axée sur le contexte organisationnel de la communauté, c'est-à-dire sur le fait que les informations échangées soient ou non accessibles aux seuls membres de la communauté. Ainsi il propose trois types de communautés :

1. Communauté centralisée : la frontière (adhésion) est clairement définie. L'appartenance et la connexion (login) sont liées. L'information est centralisée et n'est accessible qu'aux membres de la communauté d'où le nom de communauté centralisée. L'activité sociale se développe grâce à la discussion, l'échange de connaissances et le réseautage au sein de la communauté. Au sein d'une communauté centralisée il y a l'existence d'un animateur (*facilitator*) qui veille au respect des règles établies par la communauté.
2. Communauté décentralisée : elle a des limites similaires à la communauté centralisée mais elles sont poreuses. L'information peut circuler hors de la communauté. En effet, ce modèle offre aux utilisateurs la possibilité de participer à certaines activités d'une communauté définies par l'animateur sans qu'il ait besoin de se connecter.
3. Communauté distribuée : c'est le modèle que nous voyons de plus en plus sur le Web. L'animateur a moins de pouvoir par rapport aux deux types de communautés précédents et les membres ont la possibilité de créer leur propre réseau (groupe) thématique autour de leurs centres d'intérêt (par exemple Facebook). Ainsi, la frontière de la communauté est définie par une marque, un concept, c'est une association théorique représentée par un mot-clé qui peut être partagé entre les différents services et utilisé par les membres.

En résumé, chacun des différents schémas de classification présentés dans cette sous-section offre une perspective différente. Dans la littérature, l'aspect technique (support logiciel) est un peu négligé bien que Jones et Rafaeli [42] l'abordent un peu. Mais, avec le développement des TIC et l'apparition de nouvelles technologies, cet axe de classification mériterait d'être revu. Les services proposés sur la plateforme qui héberge la communauté délimitent la portée et la forme des interactions et ont un impact significatif sur le renforcement de la communauté. Ainsi les types de services (système de chat, mailing liste, etc) façonnent les types d'interactions et la cohésion de la communauté, ce qui peut être une source de classification des communautés virtuelles. De même la classification proposée par Mitchell [53] est utile car elle permet de comprendre le type de politique à mener au sein d'une communauté pour la gestion de l'information. De ce fait, chacune de ces classifications représente une compréhension spécifique dans différentes situations. En fonction de la perspective adoptée, une classification peut être plus appropriée par rapport à une autre.

2.3.3 Animation des communautés

Pour comprendre le fonctionnement d'une communauté virtuelle il faut savoir d'abord ce qui pousse les personnes à appartenir à cette communauté virtuelle, c'est-à-dire évaluer le succès des communautés virtuelles. Preece [54] propose deux grands critères pour cela : la sociabilité et la convivialité. Ridings et Gefen [55] ont également fait une enquête sur près de quatre cent personnes en leur posant la question : « *pourquoi avez-vous rejoint une communauté virtuelle ?* » À partir de cette enquête, ils définissent quatre objectifs que les utilisateurs cherchent à atteindre en rejoignant une communauté : l'accès à l'information, le soutien social, l'amitié et le divertissement.

2.3.3.1 Critères de Preece [54]

★ Sociabilité

Avec l'augmentation du nombre et de la diversité des communautés virtuelles, il ne suffit plus de concevoir une plateforme proposant un ensemble de services aux membres d'une communauté mais il faut aussi comprendre comment la technologie peut soutenir l'interaction entre les membres de la communauté. Ainsi la sociabilité d'une communauté virtuelle est fonction des interactions entre les membres de la communauté au travers des services technologiques mis en œuvre sur la plateforme (c'est-à-dire les interactions humain-humain) [54] ; c'est la façon dont les membres interagissent au sein d'une communauté et la façon dont ils s'organisent autour de leur objectif en commun. Bien que la sociabilité soit importante pour toute communauté virtuelle, au sein des communautés virtuelles de partage de connaissances, une bonne sociabilité peut permettre aux membres d'apprendre les uns des autres et ainsi favoriser le partage de connaissances. Cela nécessite également une orientation cohérente dans les interactions entre les membres pour que les connaissances partagées restent dans le domaine défini. Preece [54] définit trois points essentiels pour le bon fonctionnement d'une communauté virtuelle :

1. Thématique (*purpose*) : une communauté virtuelle s'articule toujours autour d'une thématique qui est le centre d'intérêt de ses membres et qui fournit la raison d'être de cette communauté. Ainsi l'engouement que cette thématique suscite auprès de ses membres est évalué (mesuré) grâce à quatre métriques [54] : 1) le nombre de messages des membres ou un certain type de membres montre comment les membres sont engagés pour atteindre leur objectif. L'idée pourrait être étendue pour examiner le nombre de messages au cours d'une certaine période de temps. 2) L'inter-activité implique que les membres soient à l'aise pour s'engager dans des échanges interpersonnels dans l'espace technologique. Cela va affiner leurs expertises et par conséquent les pousser à exploiter leurs connaissances. 3) La réciprocité est le ratio de ce qu'un membre reçoit de la communauté (en fonction de ses besoins) par rapport à ce qu'il contribue (les réponses qu'il fournit par rapport aux demandes des autres membres). 4) La qualité des contributions est un peu difficile à évaluer car

dépendante du type de communauté virtuelle. Elle peut être vue comme la valeur de confiance que les membres donnent aux contributions d'un membre.

2. Participants : la mesure la plus utilisée pour évaluer l'attractivité d'une communauté est le nombre de participants au sein de cette communauté [54]. Cependant, cette mesure peut être précisée en fonction des types de membres. Ainsi dans la littérature, il existe différents types de membres en fonction des communautés virtuelles : 1) le modérateur est un membre dont le rôle principal est de veiller à ce que l'interaction entre les membres de la communauté se conforme à un ensemble de politiques et de règles. Au sein d'une communauté de partage de connaissances, un modérateur joue un rôle clé en veillant à ce que les échanges des membres soient pertinents pour le domaine de connaissance cible. 2) le membre actif est un membre qui contribue régulièrement aux interactions au sein de la communauté et qui suit aussi régulièrement les évolutions dans la communauté. 3) le membre rôdeur (*lurker*) est un membre qui intervient de temps à autre dans les activités de la communauté. 4) le visiteur qui est un utilisateur non enregistré au sein de la plateforme mais qui lit les informations. Ainsi, pour évaluer l'activité d'une communauté, il faut savoir le nombre de fois où les membres actifs participent dans l'espace de travail mais également le nombre de fois où les informations de la communauté sont accédées par les utilisateurs de la plateforme. Cette mesure pourrait être évaluée régulièrement. Au début de la création de la communauté, si la communauté est active, le nombre de participations des membres actifs doit être beaucoup plus important parce que les membres actifs doivent beaucoup interagir pour construire la mémoire collective. Une fois cette mémoire collective bien mature, évaluer l'impact de la communauté sur un domaine consiste à observer le nombre de visites concernant ce domaine.
3. Réglementation (*policies*): la réglementation guide l'interaction sociale au sein d'une communauté. Ainsi elle concerne deux aspects, à savoir, la confiance dans les contributions (qui doivent être véridiques) et le code de comportement des utilisateurs qui doit être respecté pour garantir une bonne interaction entre les membres.

★ Convivialité

La convivialité est liée à la façon dont les utilisateurs interagissent grâce à la plateforme. C'est le sentiment d'efficacité et de facilité d'utilisation que les utilisateurs ont vis-à-vis de la plateforme. La sociabilité est étroitement liée à la convivialité et elle peut être considérée comme une partie de la convivialité [54]. Ainsi Preece [54] propose quatre critères pour évaluer la convivialité d'un système :

- Dialogue et appui à l'interaction sociale : la rapidité et la réaction des membres par rapport à l'interaction sociale, c'est-à-dire par exemple la facilité d'un utilisateur à accomplir une tâche.

- Présentation de l'information : la qualité de la présentation de l'information. Par exemple, combien de temps faut-il pour lire et comprendre des informations dans une fenêtre, un message ou toute autre information associée à la communauté.
- Navigation : la facilité des utilisateurs à naviguer et à trouver ce qu'ils recherchent dans la communauté.
- Accès : les moyens technologiques nécessaires aux utilisateurs (PC, laptop, etc) pour utiliser une plateforme de communauté virtuelle. Par exemple, ont-ils besoin d'une grande bande passante pour un chargement rapide des différents services de la plateforme ?

Pour évaluer ces critères, Preece [54] propose cinq mesures. 1) la vitesse d'apprentissage qui détermine le temps que prend un utilisateur pour accomplir une tâche et se familiariser avec les outils proposés par une plateforme de communauté virtuelle. 2) la satisfaction des utilisateurs qui mesure leur satisfaction après avoir accompli une tâche. Habituellement, elle est évaluée par des questionnaires de satisfaction. 3) la rétention représente la capacité des utilisateurs à se souvenir des fonctionnements des différents services proposés par la plateforme lors de leur prochaines visites. 4) la productivité dénote la durée qu'il faut aux utilisateurs pour effectuer une ou plusieurs tâches. 5) l'erreur représente le nombre d'erreurs qu'un utilisateur fait en accomplissant une tâche.

2.3.3.2 Critères de Ridings et Gefen [55]

Ridings et Gefen [55] ont mené une enquête dans des communautés virtuelles utilisant la technologie des bulletins électroniques (*Bulletin Board*). Ainsi, les messages d'un membre peuvent être consultés par toute la communauté et parfois même être accessibles au public. L'enquête a été faite sur vingt sept communautés virtuelles. Les 399 participants pouvaient choisir parmi plusieurs réponses pour répondre à la question « *pourquoi avez-vous rejoint une communauté virtuelle ?* » Après avoir étudié l'état de l'art, Ridings et Gefen [55] proposent quatre catégories pour ces réponses. Deux personnes (juges) étaient chargées de classer les réponses dans les quatre catégories définies mais avaient la possibilité de créer d'autres catégories si elles n'arrivaient pas à classer une réponse parmi les quatre prédéfinies. Ainsi, à partir de cette enquête, le résultat sur la motivation des individus à rejoindre une communauté virtuelle sont :

- Accès à l'information : les résultats de l'enquête montrent que les premières raisons qui poussent les individus à rejoindre une communauté virtuelle sont d'obtenir et de donner des informations sur un sujet, sensibiliser à propos d'un sujet et apprendre de nouvelles choses. Il s'agit donc de raisons liées à la recherche et au partage de l'information.

- Soutien Social : la seconde raison qui fait que les individus rejoignent une communauté virtuelle est le soutien social que leur fournissent les autres membres de la communauté. Ainsi le soutien social dans une communauté est le degré de satisfaction des besoins sociaux qu'un membre a reçu grâce aux échanges qu'il a eu avec les autres membres de sa communauté.
- Amitié (*friendship*) : une autre raison possible pour laquelle les personnes rejoignent des communautés virtuelles est la recherche d'amitié. Dans ce contexte, l'amitié est le sentiment d'être ensemble [55] et le partage d'une même vision.
- Divertissement : trouver des personnes avec qui on partage le même type de divertissement est une autre raison pour laquelle les gens rejoignent des communautés virtuelles. Habituellement, le but ultime est de se divertir (jouer) avec des individus et ainsi former des communautés de joueurs en ligne.

Mis à part ces quatre objectifs qui expliquent pourquoi les individus rejoignent des communautés virtuelles, l'enquête révèle que les caractéristiques techniques de la plateforme motivent aussi les individus à rejoindre une communauté.

D'après les critères de Preece [54] et de Ridings et Gefen [55], on peut dire que les propositions de Ridings et Gefen [55] traitent de la sociabilité au sens de Preece [54] et particulièrement de la thématique qui est la raison d'être de toute communauté. L'enquête de Ridings et Gefen [55] montre aussi l'importance des caractéristiques techniques de la plateforme ce qui a été traité par Preece [54] par la convivialité.

2.4 Conclusion

Dans la première partie de ce chapitre, nous avons présenté la théorie historico-culturelle de l'activité (CHAT). C'est une approche qui permet de comprendre le processus de l'apprentissage et du développement (mental). Elle privilégie une approche historico-culturelle de l'apprentissage grâce à la médiation [17] ou à travers l'action individuelle et l'action collective [18][16]. Elle considère que le développement humain (apprentissage) se fait dans l'environnement social et grâce aux interactions sociales multiples avec les pairs.

La CHAT n'est pas une méthodologie, encore moins une théorie prédictive; c'est un cadre philosophique ou un outil de clarification pour étudier les pratiques humaines comme un processus de développement. Cette théorie considère que l'individu construit sa connaissance grâce à l'apprentissage de son environnement à travers les interactions sociales. Elle offre, au moins en principe, la possibilité de conceptualiser des processus méta-cognitifs [56]. Ainsi, l'objectif principal de cette théorie est de surmonter le fossé entre, d'une part, l'être humain, et d'autre part, la culture et la société [24].

La CHAT a été utilisée comme cadre conceptuel dans le domaine de l'apprentissage collaboratif, la co-construction de connaissances

[23, 28, 29, 30, 20, 31]. La CHAT a été aussi utilisée dans domaine de l'IHM en proposant 1) un cadre conceptuel riche [25, 33]; 2) un cadre analytique pour la conception et l'évaluation des systèmes interactifs [32] et 3) la création de nouveaux systèmes et la mise en œuvre d'activités centrées sur l'informatique [34].

Dans la deuxième partie de ce chapitre nous avons passé en revue les différentes définitions et typologies des communautés virtuelles. Ces définitions et typologies diffèrent en fonction du cadre de recherche mais également des objectifs visés pour la classification.

Bien qu'il soit relativement facile de reconnaître une communauté virtuelle, il existe une multitude de définitions. Cela est principalement dû à deux choses. Premièrement à l'évolution des TIC où de nouvelles technologies apparaissent qui facilitent l'interaction des utilisateurs. Deuxièmement aux besoins des utilisateurs qui font apparaître de nouveaux domaines où ces communautés virtuelles peuvent être créées. Cependant, la définition la plus large et qui prend en compte plusieurs aspects est celle de [44]. De même pour les typologies il n'existe pas de consensus. Ainsi il existe des typologies générales qui regroupent plusieurs caractéristiques pour élaborer une typologie [45, 46, 44]. Il existe aussi des typologies plus détaillées par rapport aux objectifs des membres de la communauté. Elles constituent des sous catégories des typologies générales [41, 42]. Il existe aussi des typologies par rapport au type de transmission de savoir (connaissance) au sein de la communauté. La dernière typologie présentée s'articule autour de la structure de la communauté (adhésion).

Nous avons aussi présenté quelques caractéristiques qui favorisent l'utilisation des communautés virtuelles par les individus qui sont particulièrement la sociabilité au sein de la communauté et la convivialité de la plateforme.

Ontologies Temporelles au sein du Web Sémantique

Sommaire

3.1	Introduction	37
3.2	Définition de quelques notions	38
3.3	Modélisation d'expressions temporelles	40
3.3.1	Ontologies de haut niveau	40
3.3.2	Ontologies de domaine	44
3.4	Annotation temporelle de données	53
3.4.1	Approches du W3C	53
3.4.2	Approches spécifiques	56
3.5	Conclusion	60

3.1 Introduction

Les phénomènes temporels ont de nombreuses facettes qui sont étudiées par différentes communautés. Ainsi en physique, le temps est défini par sa mesure (mesure du temps écoulé). De ce fait, les physiciens étudient l'écoulement du temps et sa relation aux objets physiques et aux événements. La dimension temporelle des données est aussi étudiée dans le domaine de l'informatique où il y a un besoin croissant de modéliser des systèmes calendaires, des événements répétitifs et des faits qui sont vrais pour un certain temps et faux par ailleurs. C'est le cas des systèmes d'information qui doivent faire face au problème des données obsolètes. En Intelligence Artificielle, des modèles abstraits ont été proposés pour pouvoir raisonner sur des concepts temporels. Dans ce domaine, Allen [57, 58] a présenté un modèle de calcul entre les intervalles de temps qui a influencé les travaux sur la modélisation du temps. Les travaux de Allen ont été étendus aux intervalles non-convexes (intervalles répétitifs) par Ladkin [59, 60]. Dans le Traitement Automatique des Langues Naturelles (TALN) les modèles développés cherchent à extraire les expressions temporelles mais aussi leur sémantique en langue naturelle. Ainsi, un défi important dans le domaine de l'informatique est la représentation et le raisonnement sur des informations temporelles. L'intérêt de ce travail est de plus en plus important

maintenant avec l'émergence du Web sémantique où de gros volumes de données hétérogènes sont manipulés.

Dans le domaine du Web sémantique, sont produites des données comportant à la fois des notions temporelles informelles, semi-formelles et formelles qui doivent être *comprises* par les agents logiciels. Ainsi, nous distinguons deux axes de travail dans le domaine du Web sémantique : *la modélisation d'expressions temporelles* et *l'annotation temporelle des données*. La modélisation d'expressions temporelles permet de modéliser une date, un intervalle, des notions temporelles répétitives (*chaque mercredi John joue au foot*), relatives ou absolues, etc. La problématique est la représentation de l'information mais également la capacité à raisonner. L'annotation temporelle des données permet la représentation de notions temporelles de façon à annoter des connaissances (exprimées sous forme de triplet en RDF) et cela en conservant l'évolution des données (changement de valeur) dans le temps. Par exemple le triplet *address(John, Nice)* peut avoir une durée de validité dans l'intervalle de temps (t_1, t_2) et pour un autre intervalle (t_3, t_4) c'est le triplet *address(John, Dakar)* qui est valide. La problématique ici est de pouvoir représenter l'information de façon à avoir dans une même base de connaissances ces deux données et de pouvoir répondre correctement à une requête suivant l'intervalle de temps choisi. Pour cela, le Web sémantique repose sur des ontologies qui sont *une spécification explicite et formelle d'une conceptualisation partagée* [61].

Ainsi, ce chapitre est composé de trois grandes parties. En premier, nous définissons certaines notions temporelles qui nous ont permis d'évaluer les ontologies présentées. Ensuite, nous présentons les ontologies dans le domaine de *la modélisation d'expressions temporelles*. Cette sous section est scindée en deux parties : 1) présenter la partie temporelle des ontologies de haut niveau et 2) présenter les ontologies de domaine temporelles. La troisième partie présente les ontologies dans le domaine de *l'annotation temporelle des données*. Cette sous section est aussi scindée en deux parties: 1) présenter les approches du W3C qui peuvent être adoptées pour l'annotation temporelle et 2) présenter les ontologies spécifiques du domaine temporel. Enfin, nous terminons ce chapitre par une conclusion.

3.2 Définition de quelques notions

Pour une meilleure compréhension de ce chapitre, nous allons présenter quelques notions temporelles qui nous permettront d'évaluer les différentes ontologies présentées dans ce chapitre:

1. Modélisation du temps calendaire et des heures : qui permet d'exprimer des dates précises comme par exemple *heure, mois, jour, week-end, année bissextile*, etc.
2. Modélisation des instants et des intervalles que ce soit des intervalles continus (convexes [59, 60]) ou des intervalles discontinus (non-convexes [59, 60]).

3. Modélisation du temps absolu et relatif : on parle de temps absolu quand il est relié à un fait c'est-à-dire si le fait est donné avec une date précise. Par exemple, **la seconde guerre mondiale a débuté en Septembre 1939** est un temps absolu. Par contre si un fait est daté grâce à un autre fait qui a une date absolue, on parlera alors de temps relatif pour le premier décrit. Par exemple, **la guerre d'Indochine a eu lieu après la seconde guerre mondiale**.
4. Modélisation des relations entre intervalles : les relations (fig. 3.1) les plus connues sont celles définies par Allen [57, 58] pour les intervalles convexes. Cependant il existe une infinité de relations entre intervalles non-convexes comme l'a montré Ladkin [59, 60].

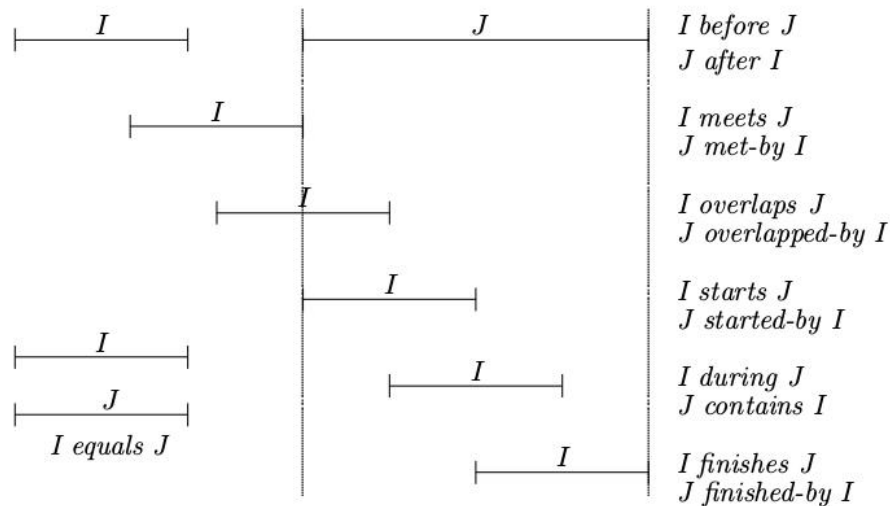


Figure 3.1: Les 13 relations d'Allen.

5. Modélisation des intervalles fermés et infinis : un intervalle fermé est un intervalle dont on connaît les deux bornes (date de début et de fin). Par contre, un intervalle infini est un intervalle dont on ne connaît que l'une de ses bornes.
6. Utilisation de concept comme l'horodatage (*timestamp*) : ceci permet de considérer le concept comme une référence temporelle. Ainsi par exemple, au lieu d'utiliser l'expression *durant le mois de Juillet 2014*, on utilisera l'expression *durant la coupe du monde de football au Brésil*.
7. Modélisation de différentes unités (granularités) de temps: les unités représentent les différentes échelles de temps modélisées qui peuvent être des *secondes*, *minutes*, ..., *décennies*, *siècles*, ... ou même des concepts horodatés.
8. Modélisation de notion de temps imprécis : cela permet de donner des dates qui ne sont pas un point fixe dans le temps. Cela sert à modéliser des dates

que l'on ne connaît pas précisément mais dont on connaît la période. Par exemple, *au début du 18ème siècle, à la fin des années 80*.

9. Modélisation du temps déictique : le temps déictique forme une relation spécifique avec le temps du discours; par exemple : *aujourd'hui, demain, la semaine dernière*.
10. Modélisation qualitative : permet de définir des notions temporelles dont la modélisation dépend de la culture et/ou de la position géographique. Par exemple, la notion d'*été* existe dans l'hémisphère sud et dans l'hémisphère nord mais l'intervalle de temps n'est pas le même.

Les dix notions présentées ci-dessus serviront de métriques pour comparer les différentes ontologies que nous allons présenter.

3.3 Modélisation d'expressions temporelles

La représentation d'expressions temporelles sert à modéliser des informations temporelles concernant un objet (une donnée). Ainsi la plupart des ontologies temporelles de la littérature traitent de cet aspect qui consiste à caractériser la dimension temporelle d'un objet par rapport à son environnement. Dans la littérature, il existe des ontologies de haut niveau qui traitent de l'aspect temporel des objets comme GFO [62], OpenCyc [63, 64], CIDOC CRM [65], SUMO [66] et DOLCE [67]. Mais il existe aussi des ontologies dédiées uniquement à la modélisation du temps comme celles qu'on peut trouver dans Linked Open Vocabularies¹ (LOV) ou Clinical Narrative Temporal Relation Ontology (CNTRO) [68, 69]. Il y a aussi TimeML [70] qui est surtout utilisé dans le TALN.

3.3.1 Ontologies de haut niveau

Une ontologie de haut-niveau est une ontologie qui décrit les concepts (classes) généraux et les propriétés (relations et attributs) dont le sens et la signification sont partagés à travers le monde dans un domaine aussi large que possible. Ainsi, les ontologies présentées ici couvrent des domaines assez larges et ne se contentent pas seulement de modéliser l'aspect temporel des objets.

3.3.1.1 Le temps dans l'ontologie GFO

General Formal Ontology (GFO) [62] est une ontologie de haut niveau pour la modélisation conceptuelle, qui est développée et maintenue par l'équipe Onto-Med de l'Université de Leipzig. Elle comprend des catégories comme des objets, des processus, le temps et l'espace, des propriétés, des relations, des rôles, des fonctions, des faits et des situations.

¹<http://lov.okfn.org/dataset/lov/vocabs?tag=Time>.

Dans sa modélisation temporelle, GFO considère deux entités de base: les *Chronoid* qui sont des intervalles connexes et les *TimeBoundary* qui sont des instants (*time-point*). Chaque *Chronoid* a deux *TimeBoundary* comme extrémité et une infinité de *TimeBoundary* entre ces points. Les *TimeBoundary* dépendent des *Chronoid* c'est-à-dire qu'ils n'ont pas d'existence indépendante et qu'ils peuvent coïncider. Il y a aussi le concept de *TimeRegion* qui est un ensemble de *Chronoid* non-connectés, c'est l'équivalent des intervalles non-convexes.

Par rapport aux relations d'Allen, GFO modélise seulement 4 relations et ne modélise aucune relation entre les *TimeRegion*. GFO ne modélise pas non plus les instants comme des entités indépendantes de même que les dates calendaires et les heures. Il considère seulement les intervalles fermés à l'aide des *Chronoid*. L'ontologie proposée par GFO ne modélise également pas les temps relatifs ni l'utilisation d'un concept comme *timestamp*.

3.3.1.2 Le temps dans l'ontologie OpenCyc

Cyc est un projet qui cherche à développer une ontologie globale et une base de connaissance générale [63, 64]. Sa dernière version open source, OpenCyc (4.0), a été publiée en Juin 2012². Cette version qui couvre plusieurs domaines comprend environ 239000 termes et 2093000 triplets³. Certains de ces termes sont alignés avec d'autres ontologies/taxonomies externes comme DBpedia et WordNet.³

Dans sa modélisation temporelle⁴, le concept général de OpenCyc est *TemporalThing* qui représente une collection de toute chose qui a une partie ou une localisation temporelle, toute chose pour laquelle on peut poser raisonnablement la question «Quand ?». Le concept *TemporalThing* peut correspondre entre autres à des objets physiques, des événements et des intervalles de temps. Ses principaux sous concepts sont : 1) *SomethingExisting* dont les instances sont plus ou moins statiques (personnes, lacs, étoiles, etc), 2) *Situation* qui est disjoint de *SomethingExisting* et dont chaque instance est un état ou un événement consistant en un ou plusieurs objets ayant certaines propriétés ou portant certaines relations avec un autre objet et 3) *TimeInterval* dont chaque instance est un objet temporel immatériel qui est entièrement caractérisé par sa partie temporelle. Il a comme sous concepts *CalendarYear*, *CalendarMonth* et *FiscalQuarter*.

Ainsi OpenCyc permet la modélisation du temps absolu et définit différentes unités de temps (granularités). Il permet aussi la modélisation des instants et des intervalles fermés et intègre les 13 relations d'Allen. Cependant il ne considère ni les intervalles infinis, ni les intervalles non-convexes, ni le temps relatif et ne modélise pas les relations entre intervalles non-convexes.

²<http://sourceforge.net/projects/opencyc/files/OpenCyc%04.0/>

³<http://www.cyc.com/platform/opencyc/>

³

⁴<http://sw.opencyc.org/2012/05/10/concept/en/TemporalThing>.

3.3.1.3 Le temps dans l'ontologie CIDOC CRM

CIDOC⁵ CRM (Conceptual Reference Model) est une ontologie pour l'échange d'informations et l'intégration entre les sources d'information hétérogènes du patrimoine culturel des musées, des bibliothèques et des archives [65]. Elle vise à fournir des définitions sémantiques et des clarifications nécessaires à la transformation des sources d'information disparates localisées dans une ressource globale et cohérente, que ce soit au sein d'une grande institution, dans les intranets ou sur Internet. La dernière version de CIDOC CRM, la 6.0 publiée en Janvier 2015, compte 93 concepts et 165 propriétés [65].

CIDOC CRM modélise le temps à partir de 4 concepts principaux [65]:

- *E2 TemporalEntity* : est un concept abstrait (il n'existe pas d'instance directe pour ce concept) qui regroupe tous les concepts qui ont une composante temporelle et une existence limitée dans le temps (*Perdurant*). Ses deux sous concepts sont:
 1. *E4 Period* : qui comprend l'ensemble des phénomènes ou manifestations culturelles limitées dans le temps. Typiquement *E4 Period* est utilisé pour décrire les périodes préhistoriques ou historiques telles que le *Néolithique* ou *La Dynastie des Ming* [65]. *E5 Event* est un sous concept de *E4 Period* et a comme sous concepts *E63 BeginningOfExistence* et *E64 EndOfExistence*.
 2. *E3 ConditionState* : qui comprend les états d'objets caractérisés par une certaine condition sur un laps de temps.
- *E52 Time-Span* : sert à modéliser les intervalles convexes et fermés ou une durée. *E52 Time-Span* est utilisé pour définir l'intervalle de temps d'un *E4 Period*, d'un *E5 Event* ou de tout autre objet valide pendant un certain temps.
- *E61 TimePrimitive* : permet de définir des intervalles de temps fermés en utilisant la notation logique ou des validations appropriées relevant de la documentation culturelle.
- *E49 TimeAppellation* : ce concept comprend toutes les formes de noms ou de codes, tels que les périodes historiques, les dates qui sont typiquement utilisées pour se référer à un objet spécifique de *E52 Time-Span*. *E49 TimeAppellation* permet aussi de définir des temps relatifs, des granularités et des notions de temps imprécis. Il a comme sous concept *E50 Date* qui permet la représentation d'une date sous différents formats comme le format calendaire.

Bien que CIDOC CRM modélise la plupart des notions présentées dans la section 3.2, l'utilisation de certaines notions est un peu fastidieuse. Par exemple pour modéliser un intervalle ouvert, il faut d'abord créer un objet de type *E77 PersistentItem* qui représente le concept des endurants et créer une partie temporelle de cet objet

⁵International Committee for Documentation

qui est de type *E63 BeginningOfExistence* si on veut modéliser un intervalle ouvert à droite. Pour ce genre de modélisation, le raisonnement devient plus compliqué car on a différents nœuds temporels pour un même objet. Il faut aussi noter que CIDOC CRM modélise les 13 relations d'Allen.

Cependant, CIDOC CRM ne propose pas le temps déictique ni la modélisation qualitative.

3.3.1.4 Le temps dans l'ontologie SUMO

Suggested Upper Merged Ontology (SUMO) [66] est une ontologie de haut niveau cohérente, conçue à partir de certaines sources publiques pour une variété de systèmes d'information. Elle a été développée sous la direction de l'équipe IEEE Standard Upper Ontology working group et son développement a débuté en mai 2000. Sa dernière version a été publiée en mai 2010 et compte 4558 concepts et 902 propriétés (892 relations et 10 attributs) [66] ; elle a été aussi alignée avec WordNet et DBpedia.

Les concepts de base dans sa modélisation temporelle sont *TimePoint* et *TimeInterval* qui sont des sous concepts de *TimePosition*. Un *TimePoint* est un instant précis et une instance de *TimePoint* ne peut exister sans une instance de *TimeInterval*. D'ailleurs, dans leur modélisation, tout est considéré comme *TimeInterval*, même le concept *Second* est modélisé comme sous concept de *TimeInterval*. Le concept *TimePoint* est utilisé pour avoir la date exacte de début et de fin d'un *TimeInterval*. L'ontologie permet la modélisation de différentes unités de temps (granularités), la modélisation de temps relatif et les 13 relations d'Allen.

Cependant, elle ne considère pas les intervalles infinis ni les intervalles non-convexes encore moins le temps déictique. De plus la modélisation d'une date simple fait intervenir différents concepts disjoints.

3.3.1.5 Le temps dans l'ontologie DOLCE

Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) est une ontologie qui modélise les catégories ontologiques sous-jacentes au langage naturel et du sens commun [67]. Dans sa modélisation, DOLCE fait la distinction entre les *Endurants* qui sont des objets qui sont toujours présents dans le temps et les *Perdurants* qui sont des objets qui existent pour une période donnée.

Ainsi dans sa modélisation temporelle, DOLCE propose deux principaux concepts : 1) le concept *Temporal-Quality* qui est une entité qui permet de modéliser les parties temporelles d'une entité de type *Perdurant* et 2) le concept *Temporal-Region* qui est une métrique temporelle qui peut être prolongée (intervalle convexe) ou instantanée (instant) et seule une entité de type *Temporal-Quality* peut y être reliée. Le concept *Temporal-Region* a comme sous concept *Time-Interval* qui a lui même le concept *Date* comme sous concept. Les intervalles non-convexes sont modélisés grâce aux *Temporal-Quality* qui sont composés par un ensemble de *Temporal-Region*. DOLCE modélise également les relations d'Allen.

Cependant, DOLCE ne fournit aucune information par rapport aux intervalles convexes fermés ou infinis. L'ontologie ne modélise ni le temps absolu ni les références temporelles (*timestamps*) et encore moins les granularités temporelles.

3.3.2 Ontologies de domaine

Une ontologie de domaine décrit un domaine/environnement particulier du monde. Elle représente une *vision du monde* d'un domaine particulier en décrivant les concepts et les propriétés de ces concepts. Ici, nous allons présenter des ontologies du domaine temporel c'est-à-dire des ontologies qui se consacrent uniquement à la modélisation du temps et qui peuvent être utilisées par d'autres applications qui ont besoin d'intégrer l'aspect temporel.

3.3.2.1 Les ontologies temporelles du LOV

Linked Open Vocabularies (LOV) est un jeu de données qui contient la description de vocabulaires RDFS et d'ontologies OWL définis et utilisés par les jeux de données Linked Open Data (LOD). Chaque fois que c'est possible, le vocabulaire/ontologie comprend les références aux jeux de données utilisés, en particulier ceux qui figurent dans CKAN⁶. Les descriptions utilisent en particulier le vocabulaire VOA⁷ (Vocabulary of a Friend) pour indiquer les différentes façons dont ces vocabulaires/ontologies peuvent être utilisés, étendus, spécifiés, annotés ou reliés entre eux et réutilisés.

Le LOV comprend 16 vocabulaires/ontologies temporels⁸ développés dans différents projets pour lesquels la modélisation du temps a été une nécessité. La classification suivante est faite selon le domaine des vocabulaires/ontologies proposés.

RDF Calendar⁹ [71], NEPOMUK Calendar Ontology¹⁰ [72] et The Opening Hours Vocabulary¹¹ [73]: le *cal*, nom donné dans LOV à RDF Calendar, est un vocabulaire en RDF de la spécification *iCalendar*¹² afin d'intégrer les données calendaires avec d'autres données du Web sémantique. Le *cal* est plutôt utilisé pour échanger des calendriers sur l'Internet. Ainsi, le concept général est *Vcalendar* qui contient deux types de base:

⁶CKAN est un système de gestion de données puissant qui fournit des données open source. Il fournit des outils pour simplifier l'édition, le partage, la recherche et l'utilisation de ces jeux de données.

⁷VOAF est une spécification de vocabulaire permettant de décrire des vocabulaires/ontologies (RDFS/OWL) utilisés dans LOD.

⁸<http://lov.okfn.org/dataset/lov/vocabs?tag=Time>, vu le 4 Novembre 2015.

⁹<http://www.w3.org/2002/12/cal/ical>

¹⁰<http://www.semanticdesktop.org/ontologies/2007/04/02/ncal/>

¹¹<http://semweb.mmlab.be/ns/oh>

¹²Internet Calendaring and Scheduling Core Object Specification (iCalendar) est un standard pour les échanges de données calendaires. Il définit la structuration des données dans un fichier de type événement de calendrier, ce n'est donc pas un protocole.

1. *Vevent* pour la modélisation des événements. Ainsi, pour le *cal* tous les faits sont modélisés comme des événements et on peut définir des intervalles fermés grâce aux attributs *dtstart* et *dtend*. Si l'attribut *dtend* n'est pas spécifié alors la fin de l'intervalle est celle de *dtstart*. On peut aussi donner une date de type *xsd:dateTime*¹³ comme la date d'un événement. Les propriétés *rrule* et *rdate* permettent de modéliser des événements récurrents (répétitifs) comme une règle. Avec les propriétés *exrule* et *exdate* on peut exprimer des exceptions dans ces événements répétitifs.
2. *Vtimezone*: puisque l'objectif est de partager des calendriers sur l'Internet, la spécification du fuseau horaire est importante pour une compatibilité des calendriers.

Bien que le vocabulaire *ical* ne modélise pas la plupart des notions temporelles présentées précédemment, il met l'accent sur deux aspects importants dans la modélisation temporelle: 1) les exceptions dans les événements répétitifs et 2) les fuseaux horaires.

NEPOMUK Calendar Ontology [72], *ncal* dans le LOV, offre une meilleure modélisation de la spécification de *iCalendar* que celle de *RDF Calendar* en proposant de nouvelles propriétés et surtout en fixant le *domain* et le *range* de ces propriétés ce qui n'est pas le cas pour *RDF Calendar*. De même, il modélise les exceptions dans les événements répétitifs et les fuseaux horaires.

The Opening Hours vocabulary [73] est un vocabulaire pour décrire les heures d'ouverture des institutions gouvernementales publiées sur le Web à l'aide des calendriers. Les calendriers recommandés par les auteurs sont : *iCal*, *RDFCal*.

OntoMedia Date¹⁴ [75, 74] : OntoMedia (Ontology for Media) est une ontologie qui fournit un vocabulaire pour le partage et l'annotation de données multimédias existant sur le Web. Sa modélisation temporelle, OntoMedia Date, permet de définir différentes unités de temps (année, mois, etc) et des temps de type XML *dateTime*. Elle n'intègre pas la plupart des notions temporelles présentées dans la section précédente.

OWL-Time [76, 77], **Time Entry**¹⁵ et **The Timeline Ontology**¹⁶: *OWL-Time* est un Working Draft du W3C. Anciennement appelé *DAML-Time*, elle est représentée en *OWL* et fournit une description temporelle des contenus du Web mais également des propriétés temporelles des Web services. L'ontologie définit trois concepts généraux pour la modélisation du temps. Le concept *TemporalEntity* qui a deux sous concepts : 1) *Interval* qui représente les intervalles de temps qui peuvent avoir un début et une fin et 2) *Instant* qui représente un point fixe. Le début et la fin d'un *Interval* sont des *Instants* et l'ontologie permet d'exprimer des intervalles

¹³http://www.datypic.com/sc/xsd/t-xsd_dateTime.html

¹⁴<https://code.google.com/p/contextus/source/browse/trunk/ontomedia/misc/date.owl>

¹⁵<http://www.w3.org/2006/time-entry#>

¹⁶<http://motools.sourceforge.net/timeline/timeline.html>

infinis en omettant une des propriétés (*hasBeginning* et *hasEnd*). Pour modéliser les 13 relations d'Allen, le concept *ProperInterval* est défini comme sous concept de *Interval* qui représente des intervalles dont le début et la fin sont distincts. La propriété *inside* a été modélisée pour représenter la relation entre un *Interval* et un *Instant*. Un autre concept majeur de *OWL-Time* est *DurationDescription* qui permet de spécifier la durée d'un intervalle grâce à la propriété *hasDurationDescription*. Ainsi, cette durée peut être spécifiée en *years*, *months*, *weeks*, *days*, *hours*, *minutes*, et *seconds*. Le concept peut avoir des sous concepts comme *Year*, *Month*, etc. Le troisième concept majeur de *OWL-Time* est *DateTimeDescription* qui sert à décrire un *Instant* grâce à la propriété *inDateTime* et un *DateTimeInterval* qui est un sous concept de *ProperInterval* grâce à la propriété *hasDurationDescription*. Une propriété importante de ce concept est *unitType* qui permet de spécifier la granularité du concept. Le concept peut avoir des sous concepts comme *January*, *February*, etc. L'ontologie *OWL-Time* définit aussi un concept *TimeZone* qui spécifie le fuseaux horaire de *DateTimeDescription* grâce à la propriété *timeZone*.

Le fichier OWL disponible¹⁷ ne modélise pas les intervalles non-convexes. Dans la documentation¹⁸ et dans [76] nous retrouvons des cas d'utilisation pour modéliser ce type d'intervalle que l'ontologie nomme *Temporal Aggregates*. L'exemple 3.1 montre la représentation en *OWL-Time* de l'expression *the first Sunday of every April*:

```

:tseq                                :tseq-everyApril
  a :TemporalSeq ;                    a :TemporalSeq ;
  :hasTemporalAggregateDescription   :hasTemporalAggregateDescription
    :firstSunEveryApril .            :everyApril .

:firstSunEveryApril                  :everyApril
  a :TemporalAggregateDescription;   a :TemporalAggregateDescription ;
  :hasContextTemporalSeq             :hasTemporalUnit
    :tseq-everyApril ;              :unitMonth ;
  :hasContextTemporalUnit            :hasithTemporalUnit 4 .
    :unitMonth ;
  :hasithTemporalUnit 7 ;
  :hasTemporalUnit :unitDay ;
  :hasPosition 1 .

```

Exemple 3.1 : Modélisation de *the first Sunday of every April* en *OWL-Time*

Pour représenter une agrégation, *OWL-Time* utilise des séquences temporelles, le concept *TemporalSeq*. Ainsi, pour modéliser l'expression précédente, l'ontologie définit deux *TemporalSeq*, une pour représenter *the first Sunday* et une autre pour *every April*. Ensuite les deux pointent sur des *TemporalAggregateDescription* qui précisent la description de l'agrégation. S'il y a des exceptions, elles sont définies

¹⁷<http://www.w3.org/2006/time>

¹⁸<http://www.w3.org/TR/owl-time/>

dans le *TemporalAggregateDescription* avec la propriété *hasGap* qui a comme valeur un entier positif. Enfin, les deux *TemporalSeq* sont reliés indirectement par la propriété *hasContextTemporalSeq*.

En résumé, *OWL-Time* est l'une des ontologies temporelles les plus complètes de LOV. D'ailleurs elle est réutilisée/étendue dans 47 jeux de données dans le LOV. Elle implémente aussi les 13 relations d'Allen. Cependant, elle ne modélise pas les *timestamp*, le temps relatif et le temps déictique.

Time Entry est une sous-ontologie de *OWL-Time* qui est beaucoup plus simple que *OWL-Time* et qui offre la plupart des concepts temporels de base et les relations dont les applications les plus simples auraient besoin, c'est-à-dire, un vocabulaire pour exprimer des faits sur les relations temporelles entre les instants, les intervalles et les événements, ainsi que des informations sur la durée, les dates et les heures.

Timeline Ontology étend les ontologies *OWL-Time* et *Time Entry* avec l'intégration de plusieurs calendriers pour les intervalles de temps et les instants. Elle a été conçue principalement pour la modélisation du temps pour les données multimédias.

The Time Interval Pattern¹⁹ et Intervals Ontology²⁰: *The Time Interval Pattern* est une spécification temporelle de la modélisation de *DOLCE-UltraLite*²¹. L'ontologie propose un seul concept, *TimeInterval* et deux attributs pour matérialiser le début et la fin de l'intervalle. Si ces deux propriétés ne sont pas spécifiées, l'intervalle de temps concerne toute la durée de l'intervalle. Par exemple l'intervalle *Samedi*, sans spécifier le début ni la fin signifie que l'intervalle commence le *Samedi* à *00H00 jusqu'à 23H59*. Cela pourrait être intéressant si on veut modéliser des événements dont on ne connaît pas le moment exact mais dont on connaît seulement le jour ou le mois ce qui peut être le cas pour des événements socioculturels.

Intervals Ontology est une ontologie qui définit différents intervalles de temps et importe l'ontologie *OWL-Time*. Dans sa modélisation, elle fait la distinction entre trois types d'intervalles : 1) *CalendarInterval* qui représente les intervalles d'une année civile, c'est-à-dire, que l'intervalle défini doit être entièrement inclus dans la même année. Ainsi elle propose différentes unités de temps (trimestre, semestre, mensuel, etc). 2) *BusinessInterval* qui représente des intervalles qui sont à cheval sur des années civiles et qui sont utilisés dans les entreprises. 3) *Interval* qui représente un intervalle convexe.

¹⁹<http://ontologydesignpatterns.org/wiki/Submissions:TimeInterval>

²⁰<http://reference.data.gov.uk/def/intervals>

²¹http://ontologydesignpatterns.org/wiki/Ontology:DOLCE+DnS_Ultralite

The Temporal Ordinal Reference Systems²² et Geologic Timescale model²³: *Temporal Ordinal Reference Systems* est une représentation en OWL du modèle Temporal Ordinal Reference Systems défini dans GeoSciML v3. GeoSciML (GeoScience Markup Language), composé de plusieurs modèles, est une norme de transfert de données basée sur XML pour l'échange de l'information géo-scientifique numérique qui représente et décrit des caractéristiques qu'on retrouve généralement sur les cartes géologiques. L'ontologie modélise trois concepts temporels principaux : 1) *TimeOrdinalEra* qui représente l'intervalle d'une ère (*era*) délimité par des 2) *TimeOrdinalEraBoundary* qui sont des points fixes dans le temps (instants). 3) *TimeOrdinalReferenceSystem* est un système de référence de temps composé d'un ensemble ordonné de périodes de temps (*TimeOrdinalEra*).

Geologic Timescale model est une représentation en RDF/OWL du modèle *Geologic Timescale* défini dans GeoSciML qui adapte le *Temporal Ordinal Reference Systems* en définissant un concept temporel, *GeologicTimescale*, qui permet de définir des échelles de temps grâce au *Temporal Ordinal Reference Systems*. Le vocabulaire défini permet la modélisation de structures de données complexes composées d'éléments abstraits qui représentent des intervalles de temps, des instants et leurs relations avec des représentations concrètes et spécifiques dans la géologie.

Vocabulary for temporal objects in Geographic Information²⁴ est une représentation en OWL des objets temporels et des systèmes de référence de la norme *ISO 19108:2002*. Cette norme définit les concepts pour décrire les caractéristiques temporelles de l'information géographique. Elle fournit une définition de base des caractéristiques temporelles des attributs, des opérations et des associations et définit les aspects temporels des métadonnées sur l'information géographique. L'ontologie définit différents types de temps. Le concept *Period* représente les intervalles et on peut utiliser différents types d'objet temporel comme *Calendar*, *ClockTime*. L'ontologie ne fait pas la distinction entre les intervalles infinis et fermés ni entre les intervalles convexes et non-convexes. Elle ne modélise pas non plus les 13 relations d'Allen.

Time Zone Ontology²⁵ est un vocabulaire qui décrit les fuseaux horaires et leur zone de couverture géographique.

The Time-indexed Value in Context²⁶ est un modèle léger (un concept et 4 propriétés) pour la description de scénarios qui impliquent des entités ayant une certaine valeur à un moment donné et dans un contexte particulier.

²²<http://resource.geosciml.org/ontology/timescale/thors>

²³<http://resource.geosciml.org/ontology/timescale/gts>

²⁴<http://def.seegrid.csiro.au/isotc211/iso19108/2002/temporal>

²⁵<http://www.w3.org/2006/timezone>

²⁶<http://www.essepuntato.it/2012/04/tvc>

Enumeration and Chronology of Periodicals Ontology²⁷ (ECPO) et **The Genealogisches Orts-Verzeichnis**²⁸ (GOV): ECPO est une ontologie bibliothécaire qui permet de définir des termes bibliographiques sur la chronologie des périodes d'emprunt d'un manuscrit. Elle permet de répondre à certaines questions comme *quelles sont les œuvres empruntées pour une période donnée ?*

GOV est une ontologie en OWL pour modéliser les informations administratives historiques comme la population d'une ville, les codes postaux, etc.

3.3.2.2 CNTRO

Clinical Narrative Temporal Relation Ontology (CNTRO) [68, 69] est une ontologie en OWL qui a été développée dans le but de permettre l'annotation sémantique temporelle des données cliniques provenant de bases de données ou de rapports cliniques.

L'ontologie définit quatre principaux concepts: *TimeInstant*, *TimeInterval*, *TimePhase* et *TimePeriod*. Le concept *TimeInstant* représente un instant dans le temps tandis que *TimeInterval* est un intervalle de temps délimité par deux *TimeInstant*. Pour modéliser les intervalles non-convexes, l'ontologie définit deux concepts : 1) *TimePhase*, sous concept de *TimeInterval*, qui représente chaque occurrence de l'intervalle à répéter et 2) *TimePeriod* qui spécifie la fréquence de répétitions de *TimePhase*. L'ontologie propose aussi la modélisation de la granularités, la représentation de temps imprécis et relatif. Elle modélise également les 13 relations d'Allen. Elle utilise d'autres classes comme *TemporalRelationStatement*, un sous concept de *rdf:Statement*, pour décrire la relation entre deux événements ou entre un événement et un *TimeInstant*.

L'exemple 3.2 modélise l'expression *every 8 hours for 10 days starting from today* en CNTRO:

```

:tphase                                :starting
  a                                     :TimePhase ;
  :hasTimePeriod                       :tperiod ;
  :startTime                           :starting ;
  :hasDuration                          :during .

                                     :during
:tperiod                                a           :Duration ;
  a                                     :TimePeriod ;
  :hasValue                             "every 8 hour" .
                                     :hasValue   "10" ;
                                     :hasUnit    "day" .

```

Exemple 3.2: Modélisation de *every 8 hours for 10 days starting from today* en CNTRO.

Dans la représentation précédente, la sémantique de la phrase est perdue puisque l'expression *every 8 hours* est modélisée comme une chaîne de caractères ce qui fait

²⁷<http://cklee.github.io/ecpo/ecpo.html>

²⁸<http://gov.genealogy.net/ontology.owl>

qu'elle ne supporte pas de raisonnement significatif. Par exemple, il est difficile de répondre à la question *quels sont les médicaments pris chaque Samedi ?* avec la représentation proposée. Un autre inconvénient de cette modélisation est que les valeurs de certains objets sont modélisées comme des chaînes de caractères, que se soient les heures, les jours, les mois ; cela diminue la capacité de raisonnement sur les concepts de l'ontologie.

3.3.2.3 TimeML

TimeML [70] est un langage d'annotation des informations temporelles dans des documents dont l'annotation est basée sur un système de balises standard: *tag* et *attributs*. Ainsi, TimeML répond à quatre objectifs principaux : 1) utiliser un *timestamp* (identification d'un événement et son ancrage dans le temps) comme date, 2) décrire l'ordonnancement des événements décrits dans les documents, 3) permettre le raisonnement temporel contextuel sur des expressions comme *la semaine dernière*, *deux semaines avant* et 4) permettre aussi le raisonnement sur la durée d'un événement. Pour cela, le langage spécifie les quatre structures (*tag*) suivantes :

1. *EVENT* correspond à toute situation qui se produit ou a eu lieu. Il peut aussi exister des états ou des circonstances dans lesquelles quelque chose devient vrai. Les *EVENT* sont généralement des verbes, bien que les valeurs nominales d'événements puissent également être annotées comme des *EVENT*. La balise *EVENT* est également utilisée pour annoter un sous-ensemble d'états dans un document. Ce sous-ensemble d'états comprend ceux qui sont soit transitoires soit explicitement marqués comme participant à une relation temporelle. Parmi ces attributs, l'attribut *CLASS* permet de décrire le type d'événement. Il peut avoir entre autres valeurs: 1) *REPORTING* pour décrire l'action d'une personne ou d'une organisation déclarant quelque chose, narrer un événement, informant sur un événement, etc. 2) *PERCEPTION* pour décrire la perception physique (voir, entendre, sentir, etc.) d'un événement. 3) *ASPECTUAL* décrit les facettes historiques des *EVENT* (*begin*, *finish*, *stop*, *continue*).
2. *TIMEX3* est utilisé pour marquer les expressions temporelles explicites, telles que les heures, les dates, la durée, etc. Parmi ces attributs, l'attribut *type* permet de spécifier quatre types d'expressions: (1) *DATE* pour exprimer une expression temporelle complète comme *le 11 Juin 1989*, *été 2002*; (2) *TIME* pour exprimer une expression temporelle imprécise comme *lundi*, *le mois prochain*, *l'année dernière*, *il y a deux jours*; (3) *DURATION* pour exprimer des durées comme *trois mois*, *deux ans*; (4) *SET* pour exprimer une expression temporelle répétitive comme *deux fois par mois*, *quotidien*. Les attributs *FREQ* et *QUANT* combinés avec la valeur *SET* de l'attribut *type* permet de spécifier la fréquence d'une expression répétitive. L'attribut *MOD* est utilisé comme modificateur temporel. L'attribut *VALUE* combiné avec la valeur *DURATION* de l'attribut *type* permet de spécifier une durée.

3. *SIGNAL* est utilisé pour annoter des sections de texte, il permet d'indiquer les relations temporelles entre les entités (*TIMEX3* et *EVENT*, *TIMEX3* et *TIMEX3* ou *EVENT* et *EVENT*).
4. *MAKEINSTANCE* est utilisé pour annoter des informations sur un *EVENT* donné. Parmi ces attributs, l'attribut *TENSE* qui permet de spécifier si *EVENT* est *PAST*, *PRESENT* ou *FUTURE*.

L'une des principales innovations introduites dans TimeML consiste en ses balises *LINK* qui codent les diverses relations existant entre les éléments temporels d'un document et établissent l'ordonnancement direct entre les événements. L'avantage des *LINK* est de distinguer les types et les instances de *EVENT* et de traiter de manière adéquate la subordination contextuelle impliquant la modalité et le discours rapporté. Il existe trois types de balises de liens :

1. *TLINK* : représente la relation temporelle entre les événements ou entre un événement et un temps. Elle établit aussi un lien entre les entités concernées.
2. *SLINK* : il s'agit d'un lien de subordination qui est utilisé dans des contextes impliquant la modalité, l'évidence, le conditionnel ou le factuel. Un *SLINK* est utilisé aussi dans les cas où une instance d'événement subordonne un autre type d'instance d'événement. Il est utilisé pour lier deux *EVENT* ou un *EVENT* et un *SIGNAL*
3. *ALINK* : est un lien aspectuel qui est un lien qui définit l'aspect temporel d'un *EVENT* c'est-à-dire *begin*, *finish*, *stop*, *continue*. Il indique une relation aspectuelle entre deux événements. Dans certains cas il est comme un croisement entre *TLINK* et *SLINK* car il indique à la fois une relation entre deux éléments temporels ainsi que la subordination aspectuelle.

Il faut aussi noter l'existence de la balise *CONFIDENCE* qui permet de donner le degré de confiance d'une balise ou de ses attributs. L'exemple 3.3 modélise en TimeML l'expression *John taught from September to December last year*.

John	<TIMEX3 tid="t5" type="DURATION"
<EVENT eid="e1"	value="P4M" beginPoint="t1"
class="OCCURRENCE">	endPoint="t2" temporalFunction="true"/>
taught	
</EVENT>	<TIMEX3 tid="t3" type="DATE"
	value="1995" temporalFunction="true"
<MAKEINSTANCE eiid="ei1"	anchorTimeID="t4">
eventID="e1" pos="VERB"	last year
tense="PAST" aspect="NONE"	</TIMEX3>
polarity="POS"/>	
<SIGNAL sid="s1"> from	<TIMEX3 tid="t4" type="DATE" value="1996-03-27"
</SIGNAL>	functionInDocument="CREATION_TIME">
	03-27-96
	</TIMEX3>
<TIMEX3 tid="t1"	
type="DATE" value="xxxx-09">	<TLINK timeID="t1" signalID="s1" relatedToTime="t5"
September	relType="BEGINS"/>
</TIMEX3>	
<SIGNAL sid="s2"> to </SIGNAL>	<TLINK timeID="t2" signalID="s2"
	relatedToTime="t5" relType="ENDS"/>
<TIMEX3 tid="t2"	<TLINK eventInstanceID="ei1"
type="DATE" value="xxxx-12">	relatedToTime="t5" relType="HOLDS"/>
December	
</TIMEX3>	

Exemple 3.3 : Modélisation de *John taught from September to December last year* en TimeML.

Ainsi, TimeML permet la modélisation d'intervalles de temps fermés ou infinis, la modélisation d'intervalles convexes ou non-convexes, la modélisation de certaines relations temporelles. L'inconvénient principal de ce langage est qu'il annote les événements et les expressions temporelles dans des segments textuels isolés. Ainsi dans l'exemple précédent, rien n'est dit entre l'intervalle *from September to December* et l'événement *John taught*. La conséquence principale d'isoler les événements et les expressions temporelles au sein des structures est de rendre ensuite l'exploitation difficile pour les systèmes de recherche d'information. De plus l'inférence ne permet pas de récupérer des informations qui n'ont pas été mentionnées explicitement dans le document (données non étiquetées). Certes le langage permet de représenter des intervalles non convexes mais il ne permet pas de représenter des expressions plus complexes comme celles du type *every 3rd Monday*. Il ne permet pas non plus la modélisation qualitative.

Dans cette sous section, nous avons présenté certaines ontologies qui proposent

une représentation d'expressions temporelles qui est la représentation de notions temporelles liées à la caractérisation d'une donnée (objet). Parmi les ontologies présentées, celles qui semblent être les plus complètes sont OWL-Time, TimeML, CIDOC-CRM et CNTR0 bien qu'il existe certaines limites dans ces ontologies. Il faut aussi noter les approches de The Time Interval Pattern et DOLCE qui ne modélisent que les intervalles ce qui semble intéressant pour les événements dont les instants sont inconnus.

3.4 Annotation temporelle de données

Dans cette sous section, nous ferons l'état de l'art des ontologies qui proposent une représentation temporelle liée à un triplet RDF, c'est-à-dire l'intervalle de temps ou la période durant lequel le triplet est vrai. Comme illustration, nous allons prendre l'exemple des triplets *address(John, Dakar)* qui a une validité dans l'intervalle de temps $(t1, t2)$ et *address(John, Nice)* qui a une validité dans l'intervalle $(t3, t4)$. La problématique ici est de pouvoir représenter ces données correctement et de pouvoir répondre à une requête par rapport à l'intervalle de temps choisi. De façon plus générale, la problématique est la représentation et l'interrogation des relations n-aires dans le Web sémantique. En effet, pour le langage du Web sémantique RDF, un énoncé (*statement*) est une relation binaire qui est utilisée pour relier deux individus (instances) ou un individu et une valeur. De ce fait avec la dimension temporelle, on ne manipule plus des relations binaires mais des relations ternaires.

3.4.1 Approches du W3C

Le W3C propose principalement trois approches pour la représentation de relations n-aires. Ces approches peuvent être utilisées pour ajouter une information temporelle à un triplet pour obtenir une relation ternaire. Ainsi les approches présentées ici sont les graphes nommés (named graphs), les relations n-aires (N-ary relations) [71] et la réification RDF.

3.4.1.1 Les graphes nommés

Les graphes nommés (*named graphs*) sont un concept du Web sémantique grâce auquel un graphe (ou un ensemble de triplets RDF) est identifié en utilisant un URI, permettant de contextualiser les triplets en ajoutant des métadonnées. Comme information additionnelle aux triplets, on peut annoter l'URI du graphe nommé avec des informations temporelles qui représentent la période de validité des triplets dans le graphe nommé. Cette approche a été utilisée par Tappolet et Berstein [78] qui ont proposé une syntaxe et un format de stockage pour exprimer des énoncés temporels en RDF. Ils ont aussi proposé un langage de requêtes temporelles, τ -SPARQL, qui peut être transformé en SPARQL standard. Ils ont aussi étendu l'ontologie OWL-Time en y ajoutant un nouveau format de date pour exprimer les représentations de temps non calendaire.

L'exemple 3.4 représente en graphes nommés les deux assertions suivantes: *address(John, Dakar):[t1, t2]* et *address(John, Nice):[t3, t4]*

```
@prefix : <http://example.org/>

:G1 { :John :address :Dakar .}

:G2 { :John :address :Nice . }

:G1 :hasInterval [ :hasBeginning t1 ;
                  :hasEnd t2 ].

:G2 :hasInterval [ :hasBeginning t3 ;
                  :hasEnd t4 ] .
```

Exemple 3.4 : Représentation en graphe nommé.

Le principe est de regrouper dans un même graphe nommé des triplets qui appartiennent à un même intervalle de temps et d'annoter temporellement l'URI du graphe. On considère ainsi, par convention, que l'annotation temporelle porte sur l'ensemble des triplets du graphe nommé annoté. Il faut noter qu'avec cette approche, on peut définir la plupart des notions temporelles présentées dans la section 3.2. Par exemple, en modélisant les 13 relations d'Allen à l'aide des graphes nommés cela permet de réduire le nombre d'inférences dans la base puisque les relations sont définies entre les graphes nommés plutôt qu'entre les triplets.

Hormis la prolifération des objets (création d'objet intermédiaire) et la redondance des données (duplication des données concernant John à cause des contraintes d'implications de SPARQL²⁹) le principal inconvénient des graphes nommés est lié au raisonnement temporel³⁰. De plus aucun raisonneur ne peut être utilisé sur des graphes nommés. La seule alternative est l'utilisation de requêtes CONSTRUCT en SPARQL ce qui est fastidieux pour faire du raisonnement. Il faut aussi noter que cette approche n'est optimale que si plusieurs triplets appartiennent à un même graphe (les triplets ont la même période de validité) sinon on serait obligé de créer pour chaque triplet un graphe nommé ce qui peut conduire à une diminution des performances lors de la recherche d'information.

3.4.1.2 Relations n-aires

Une relation n-aire est une relation qui lie un sujet à deux ou plusieurs termes. Cela est différent d'une relation binaire, qui lie un sujet à un seul terme. Dans [71], les auteurs proposent quatre façons de représenter des relations n-aires. Leur approche, basée sur RDF/OWL, utilise des relations binaires en créant une entité intermédiaire qui sert de sujet pour l'ensemble des relations. Cette entité à son tour

²⁹<https://www.w3.org/TR/rdf11-datasets/#relationship-with-sparql-entailment-regime>

³⁰<https://www.w3.org/TR/rdf11-datasets/#properties-of-this-dataset-semantics>

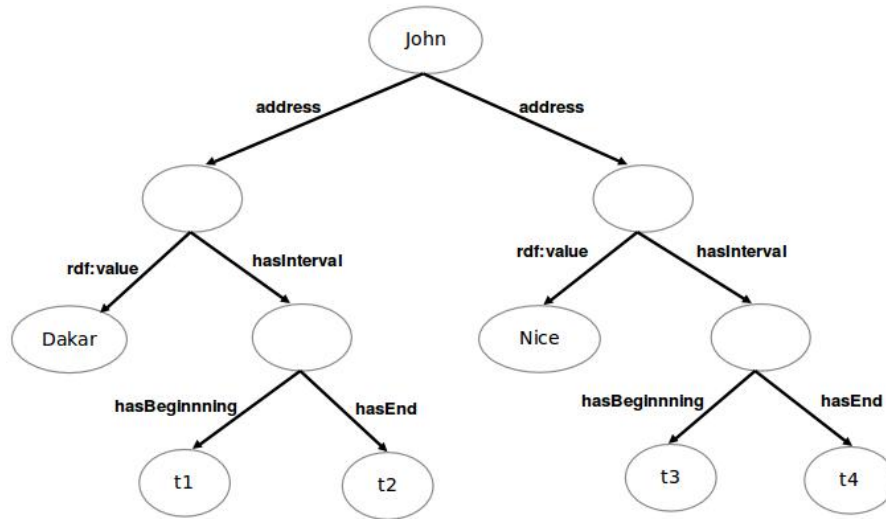


Figure 3.2: Représentation en relations n-aires.

fait l'objet d'une relation dans laquelle le sujet principal est l'entité. La figure 3.2 illustre la représentation à l'aide des relations n-aires des deux assertions suivantes: $address(John, Dakar):[t1, t2]$ et $address(John, Nice):[t3, t4]$.

Dans cette approche, pour représenter le triplet $address(John, Dakar):[t1, t2]$, il faut introduire un blank node qui est l'objet intermédiaire qui a pour valeur (*rdf:value*) l'objet de la relation binaire *address* (c'est-à-dire *Dakar*) et pour intervalle (*hasInterval*) la période de validité de la relation. Le tout est relié au sujet (*John*) au travers de la relation *address*.

Hormis la prolifération des données, le principal inconvénient est la perte de l'assertion entre le sujet et l'objet puisque la relation *address* relie maintenant le sujet et un blank node.

Notons que les relations n-aires offrent aussi la possibilité de modéliser des séquences c'est-à-dire des listes ordonnées.

3.4.1.3 La réification RDF

La réification permet de décrire un triplet RDF et de lui associer des métadonnées. Pour cela RDF définit les propriétés suivantes :

1. *rdf:subject*: identifie le sujet.
2. *rdf:predicate*: identifie la propriété.
3. *rdf:object*: identifie l'objet.
4. *rdf:type*: un triplet réifié est instance de *rdf:Statement*.

En plus de ces quatre propriétés, on peut ajouter autant d'information que l'on souhaite sur le triplet réifié pour contextualiser le *statement* comme le montre

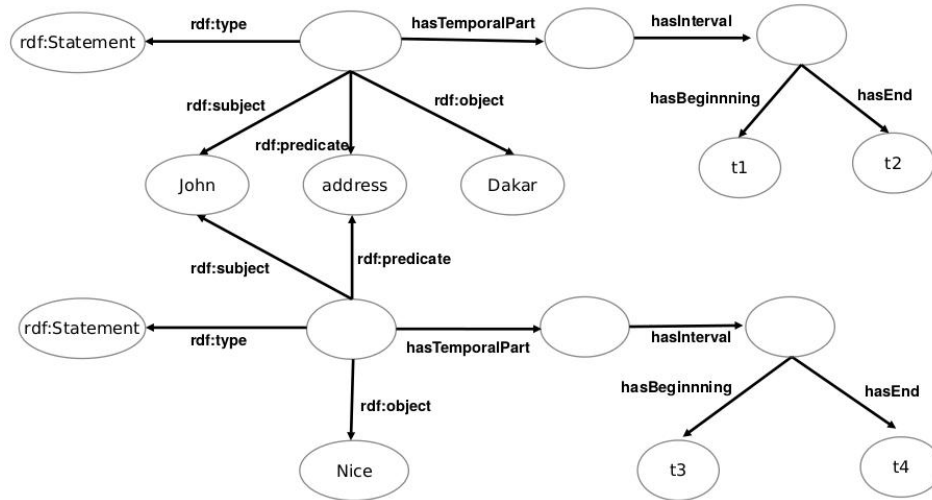


Figure 3.3: Représentation avec la réification RDF.

la figure 3.3 qui est une représentation de la réification RDF des deux assertions suivantes: $address(John, Dakar):[t1, t2]$ et $address(John, Nice):[t3, t4]$.

L'approche utilise des relations binaires en créant un blank node qui est de type $rdf:Statement$ et qui est relié au sujet de la relation originale³¹ par $rdf:subject$; le prédicat de la relation originale est relié au blank node par la propriété $rdf:predicate$; l'objet de la relation originale est relié au blank node par la propriété $rdf:object$. Ainsi pour contextualiser le triplet (la relation originale), on peut ajouter d'autres annotations au blank node. Dans le cas temporel, il suffit d'ajouter un nœud temporel (annoté par des propriétés temporelles) et relier ce dernier au blank node par la propriété $hasTemporalPart$.

La réification RDF présente les mêmes inconvénients que les relations n-aires; à savoir la duplication et la prolifération des données, la perte de l'assertion du triplet original, le problème pour représenter certaines propriétés algébrique comme l'*inverse*. L'avantage de la relation n-aire par rapport à la réification RDF est le nombre moins élevé de données répliquées et la possibilité de représenter des séquences ordonnées.

3.4.2 Approches spécifiques

Dans cette sous section nous avons regroupé les travaux dont l'objectif principal est de modéliser l'information temporelle dans un triplet. Ainsi, il existe l'approche Temporal RDF [79] qui est une amélioration de la réification RDF, l'approche 4D-Fluents [80] qui a été étendue dans SOWL [81, 82] et dans tOWL [83, 84]. Il existe aussi d'autres approches comme celles proposées par Rula et ses collègues [85], Scheuermann et ses collègues [86] et l'approche basée sur la logique de descriptions

³¹Dans l'exemple, une relation originale est le triplet $address(John, Dakar)$

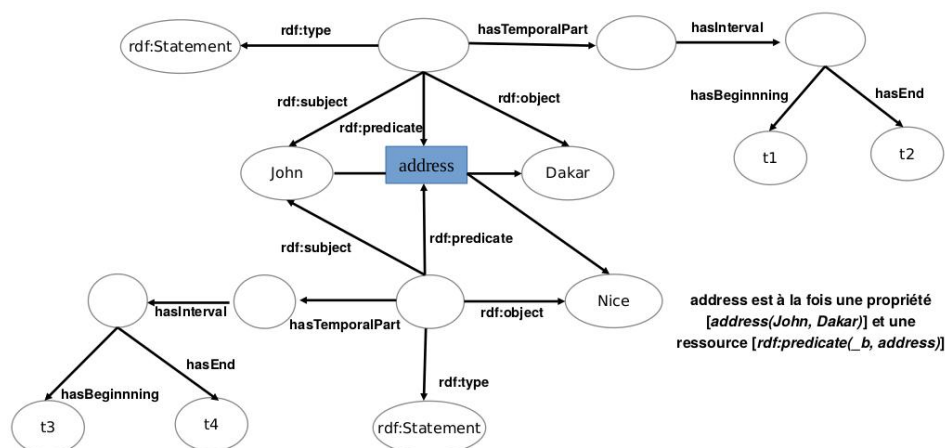


Figure 3.4: Représentation en Temporal RDF.

temporelles (Temporal Description Logics - TDL) [87, 88].

3.4.2.1 Temporal RDF

Temporal RDF [79] étend la réification *RDF Statement* d'un triplet où le quatrième élément est la période de validité du *RDF Statement*. Temporal RDF, encodé en utilisant la réification standard de RDF, permet l'utilisation de labels pour matérialiser la temporalité. Un label temporel est un élément temporel t dans un intervalle de temps qui représente la durée de validité d'un triplet (s, p, o) . Ceci entraîne une extension de la syntaxe et de la sémantique de RDF. La figure 3.4, ci-dessus, illustre la représentation en Temporal RDF des deux assertions suivantes: $address(John, Dakar):[t1, t2]$ et $address(John, Nice):[t3, t4]$.

Temporal RDF améliore la réification standard de RDF en gardant l'assertion du triplet original c'est-à-dire du triplet à réifier. Ainsi, Temporal RDF garde la sémantique (l'assertion) en utilisant la relation à la fois comme une relation dans la représentation originale et un concept dans la réification. De ce fait, le graphe peut être accédé à travers les deux vues selon qu'on s'intéresse à la temporalité ou à la représentation non temporelle. Si on s'intéresse à la temporalité, on gardera la vue globale du graphe en intégrant la réification RDF ce qui permet d'avoir à la fois les données et leur temporalité. Si on s'intéresse seulement aux données on pourra considérer la relation initiale c'est-à-dire omettre la réification RDF. Bien qu'elle introduise une redondance des données, cette approche présente de nombreux avantages comme la possibilité de représenter certaines propriétés algébriques comme l'inverse. Les auteurs ont aussi introduit des inférences (*entailments*) qui permettent de déduire certaines relations temporelles non explicitées.

3.4.2.2 4D-Fluents

La représentation en OWL des approches proposées par le W3C (la réification RDF, les relations n-aires et les graphes nommés) présente certains inconvénients. En effet la réification pose un problème de raisonnement en OWL puisque la relation du triplet à réifier est modélisée comme une ressource et la sémantique des propriétés de OWL n'est plus applicable (les propriétés d'une relation ne sont plus directement associées à la relation elle-même). Dans les approches basées sur les relations n-aires, la relation est scindée en deux, ce qui présente un inconvénient dans le raisonnement. De plus la représentation des propriétés inverses et symétriques accentue la redondance des données. En ce qui concerne les graphes nommés, ils ne font pas partie de la spécification de OWL³² et aucun raisonneur OWL ne les supporte.

C'est pourquoi Welty et Fikes [80] proposent une approche basée sur les endurements et les perdurants pour modéliser l'évolution temporelle des données en OWL. Les objets temporels sont représentés comme des objets à 4 dimensions ; la 4ème dimension étant le temps. Les propriétés qui ont une dimension temporelle sont appelées *fluents*. L'annotation temporelle est faite sur ces dernières en les rattachant aux parties concernées des objets. Ainsi les *fluents* sont les propriétés dont les valeurs sont susceptibles de changer dans le temps. La figure 3.5, illustre la représentation en 4D-Fluents des deux assertions suivantes: $address(John, Dakar):[t1, t2]$ et $address(John, Nice):[t3, t4]$. Dans cet exemple la propriété *address* est considérée comme un *fluent*.

Dans cette approche, les auteurs adoptent la théorie selon laquelle (i) tout objet a une partie temporelle et (ii) ce sont ces parties temporelles qui sont en interaction. Ainsi pour matérialiser la relation $address(John, Dakar):[t1, t2]$, ils considèrent que ce sont les parties temporelles de *Dakar* et de *John* qui sont en relation dans l'intervalle $[t1, t2]$. De ce fait, la relation temporelle est indiquée entre les parties temporelles. Puisque les changements sont modélisés par rapport aux propriétés, il faut choisir durant la modélisation quelles sont les propriétés qui peuvent conduire à des évolutions. Dans l'exemple précédent, si on voulait ajouter les propriétés *birthDate* et *buildDate* respectivement aux objets *John* et *Dakar* nous n'aurions pas besoin de créer une partie temporelle pour chaque objet puisque la valeur de chacune de ces deux propriétés est statique (elle ne change pas dans le temps). Ainsi, ces propriétés seraient rattachées directement aux objets et non à leurs parties temporelles.

L'approche 4D-Fluents souffre encore de redondance des données et de prolifération des objets. Contrairement aux autres approches du W3C en OWL, elle maintient l'expressivité de OWL et permet le raisonnement. De plus la partie temporelle de la modélisation peut être combinée avec les approches présentées dans la section 3.3 pour la modélisation d'expressions temporelles. Hormis la redondance de données, l'autre inconvénient est lié au chevauchement des parties temporelles d'un objet c'est-à-dire quand deux parties temporelles d'un objet sont en relation avec d'autres et que ces intervalles de temps se chevauchent. Le principal problème se

³²<http://www.w3.org/TR/owl2-syntax/>

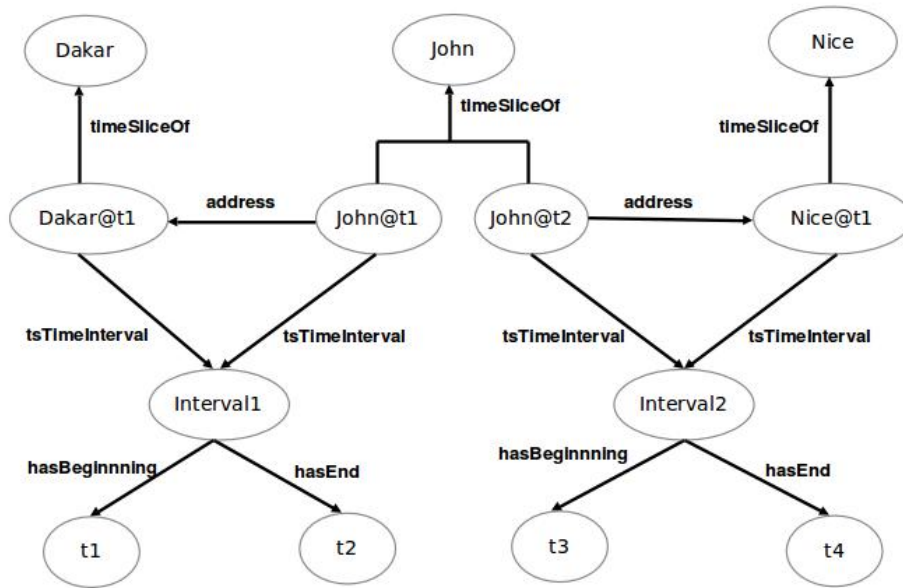


Figure 3.5: Représentation en 4D-Fluents.

pose pour raisonner et interroger la partie temporelle.

Ainsi, Baratis et ses collègues [81, 82] proposent une approche pour gérer ces situations en étendant 4D-Fluents avec les 13 relations d'Allen et en créant deux langages de requêtes, TOQL (Temporal Ontology Query Language) [81] et SOWL (Spatio-temporal OWL) [82], qui supportent des requêtes sur les informations quantitatives et qualitatives spatio-temporelles d'une ontologie en 4D-Fluents. De même, Frasinca et ses collègues proposent une extension du 4D-Fluents, tOWL, à trois niveaux [83, 84]. Dans le premier niveau, ils utilisent les domaines concrets (*concrete domains*) qui permettent la représentation de restrictions en utilisant des prédicats binaires. Le deuxième niveau est une représentation temporelle qui permet de modéliser des instants, des intervalles non convexes et les 13 relations d'Allen. Enfin le dernier niveau utilise l'approche 4D-Fluents en l'étendant avec des axiomes temporels complexes, comme les états de transition.

Hormis ces cinq méthodes, il existe d'autres approches pour l'annotation temporelle des données. Le *versioning* consiste à créer une nouvelle base de connaissances (BC) à chaque fois qu'un objet change de valeur. Cette technique présente plusieurs inconvénients : 1) un changement sur des attributs simples exige qu'une nouvelle version de la BC soit créée menant ainsi à une redondance des données, 2) la recherche d'informations nécessite des recherches exhaustives dans plusieurs versions de la BC et 3) les relations temporelles entre les versions ne sont pas clairement définies et les objets évolutifs ne sont pas spécifiés.

Rula et ses collègues proposent aussi une approche générique pour extraire des

informations temporelles du Web et leur durée de validité [85]. Scheuermann et ses collègues proposent également une approche empirique centrée sur les perspectives des utilisateurs, ce qui a permis de définir différents modèles temporels [86].

Une autre approche est celle des logiques de descriptions temporelles (Temporal Description Logics - TDL) [87, 88] qui étendent les logiques de description du Web sémantique avec des constructeurs supplémentaires. Les TDL offrent des capacités d'expressivité supplémentaires et restent décidables. Bien qu'elles constituent la seule famille d'approche qui ne crée pas de redondance de données, leur utilisation nécessite d'étendre la syntaxe et la sémantique des langages du Web sémantique comme RDF et OWL pour ajouter la dimension temporelle aux données.

3.5 Conclusion

Dans ce chapitre nous avons présenté des travaux sur la modélisation temporelle dans le Web sémantique. Le nombre élevé de travaux montre l'importance des notions temporelles dans les données manipulées au sein du Web sémantique. Dans ces travaux sur l'annotation temporelle, nous avons distingué deux domaines de modélisation. Le premier que nous avons appelé *modélisation d'expressions temporelles* qui consiste à modéliser des dates, des intervalles, des notions temporelles relatives, absolues, des relations entre les intervalles, etc. Le second, que nous avons appelé *annotation temporelle des données*, permet la représentation de notions temporelles de façon à annoter des connaissances pour garder l'évolution temporelle des connaissances (changement de valeur dans le temps).

Dans le domaine de *modélisation d'expressions temporelles*, nous retrouvons des ontologies de haut-niveau comme GFO [62], OpenCyc [63, 64], CIDOC CRM [65], SUMO [66] et DOLCE [67]. Dans ces ontologies sont modélisées différentes notions temporelles mais leur principal inconvénient est qu'elles ne modélisent ni les intervalles non convexes ni le temps déictique et encore moins la représentation qualitative.

Il existe également de nombreuses approches plus spécifiques au domaine temporel comme celles que nous pouvons retrouver dans le LOV et dont la plus complète et la plus réutilisée est OWL-Time [76, 77]. Cette approche modélise différentes notions mais présente comme inconvénient la non représentation du temps déictique pour les expressions humaines de temps. Il existe aussi l'ontologie CNTRO [68, 69] dans le domaine médical qui permet différentes représentations. Cependant, elle modélise la périodicité des intervalles non convexes dans des chaînes de caractères d'où la perte de la sémantique. La dernière ontologie présentée dans ce domaine est TimeML [70] qui est utilisée pour le TALN. Son principal inconvénient est qu'elle ne permet pas de représenter des intervalles non convexes complexes comme dans l'expressions *every 3rd Monday*. De plus, elle ne permet pas la modélisation qualitative.

Dans le domaine de l'*annotation temporelle des données*, il existe trois approches du W3C dont l'objectif principal est de représenter des relations n-aires. Ces approches peuvent être utilisées dans le domaine temporel où la problématique est la représentation des relations ternaires. Ainsi, la première approche s'appuie sur les graphes nommés (*named graphs*) qui permettent de regrouper des triplets qui ont la même période de validité dans un graphe (URI) et d'annoter temporellement ce graphe. Le principal inconvénient des graphes nommés se situe au niveau du raisonnement temporel. La deuxième approche s'appuie sur les relations n-aires (*N-ary relations*) qui sont décrites dans une note du W3C [71]. Hormis la redondance des données, le principal inconvénient est la perte de l'assertion des triplets avec l'ajout d'un nœud intermédiaire pour ajouter le temps. La dernière approche est la réification RDF qui comme les relations n-aires remplace le triplet original en créant un *Statement* qui peut être annoté temporellement.

Il existe aussi d'autres approches spécifiques à la modélisation temporelle comme Temporal RDF [79] qui propose d'étendre la réification RDF en gardant les triplets assertés. Cela permet d'avoir deux vues sur le graphe des données : une vue temporelle et une vue non temporelle. Cela facilite la recherche d'information. Cependant, l'utilisation de Temporal RDF nécessite une extension de la sémantique standard de RDF [89] à cause des inférences (*entailment*) introduites dans Temporal RDF. Il existe aussi l'approche 4D-Fluents [80], modélisée en OWL, qui considère que chaque objet a une partie temporelle et que ce sont ces parties temporelles qui interagissent. Le principal inconvénient de cette approche se situe au niveau du raisonnement temporel puisque les parties temporelles peuvent se chevaucher. C'est ainsi que Baratis et ses collègues proposent d'étendre l'approche 4D-Fluents en y intégrant les 13 relations d'Allen [81, 82]. Il existe aussi l'approche qui s'appuie sur les logiques de descriptions temporelles [87, 88] dont le seul inconvénient est la nécessité d'étendre la syntaxe et la sémantique des standards du Web sémantique.

Positionnement

Sommaire

4.1	Introduction	63
4.2	Capitalisation de connaissances	63
4.3	Ontologies Temporelles	67
4.4	Conclusion	69

4.1 Introduction

Ce dernier chapitre de l'état de l'art présente la synthèse des deux chapitres précédents. Dans ce chapitre, nous allons également présenter dans quel contexte les travaux abordés précédemment ont été utilisés pour cette thèse.

Ce chapitre est composé de deux grandes parties. En premier lieu, dans la section capitalisation de connaissances, nous montrons comment la CHAT¹ (théorie historico-culturelle de l'activité) peut être utilisée comme *méta-modèle* dans nos travaux. Ensuite, nous présentons comment nous comptons étendre l'approche de Vygotsky (socioculturelle), un modèle de la CHAT, à une communauté. Cette extension nous a amenés à positionner la notion de communauté, dans la deuxième partie de la section capitalisation de connaissances. La deuxième partie de ce chapitre concerne la modélisation temporelle qui est scindée en deux parties. En premier, nous présentons les approches les plus complètes pour la modélisation d'expression temporelles. À partir de là, nous comparons ces approches, ce qui permettra d'identifier les critères les plus importants dans la modélisation d'expression temporelles. En deuxième lieu, nous présentons les approches dans le domaine de l'annotation temporelle ce qui permettra également de mettre en évidence les critères les plus importants dans ce domaine. Enfin, nous terminons ce chapitre par une conclusion.

4.2 Capitalisation de connaissances

La CHAT (Théorie Historico-Culturelle de l'Activité) est un espace métaphorique représentant le lieu du développement cognitif, occupé par des pairs, des experts et d'autres agents susceptibles de contribuer au développement. Cette théorie considère

¹Cultural Historical Activity Theory

l'action humaine (activité) comme étant en relation avec des objets tels que des outils (matériel ou mental) qui transportent avec eux l'histoire et la culture de l'humanité. Dans cette théorie, le développement individuel ne peut pas être compris sans faire référence au contexte social dans lequel est intégré un tel développement. L'individu n'interagit pas directement avec les objets de l'environnement. Cette relation entre eux se fait au travers de moyens culturels, d'outils et de signes.

La CHAT n'est pas une méthodologie, encore moins une théorie prédictive ; c'est un cadre philosophique ou un outil de clarification pour étudier les pratiques humaines comme un processus de développement. Cette théorie considère que l'individu construit sa connaissance grâce à l'apprentissage de son environnement à travers les interactions sociales. Elle offre, au moins en principe, la possibilité de conceptualiser de façon scientifique des processus méta-cognitifs [56]. Ainsi, l'objectif principal de cette théorie est de surmonter le fossé entre, d'une part, l'être humain, et d'autre part, la culture et la société [24].

Ainsi la CHAT est une théorie intéressante pour la mise en place d'un cadre dont l'objectif principal est la transmission et l'apprentissage de connaissances dans un contexte collaboratif et collectif. L'absence de directives structurées et communes dans l'ingénierie ontologique (cf. chapitre I) fait qu'il n'existe pas un seul mode ou une seule méthodologie *correcte* pour développer des ontologies [11]. Ainsi, la CHAT pourrait servir de *méta-modèle* pour conceptualiser du domaine socioculturel.

Par conséquent, nous avons proposé une méthodologie pour identifier les caractéristiques représentant une communauté dans ses aspects sociaux (au sens large). La modélisation de ces caractéristiques forme l'ontologie socioculturelle. Ainsi, nous avons choisi d'étendre l'approche de Vygotsky à une communauté atomique et d'utiliser cette extension comme un *méta-modèle* pour la construction de l'ontologie socioculturelle. Nous entendons par méthodologie, les procédures de travail, les étapes, qui décrivent le pourquoi et le comment de la conceptualisation puis de l'artefact construit.

Les principaux points de différences entre notre approche et celle de Vygotsky peuvent être résumés comme suit (voir le tableau 4.1). Pour Vygotsky, la portée de l'activité repose sur l'individu bien qu'elle puisse être collective. En effet, au sein de la zone proximale de développement (ZPD) seuls les individus qui prennent part à l'activité peuvent bénéficier des résultats de l'apprentissage. Alors que dans notre approche, les résultats ne sont pas seulement bénéfiques aux individus qui co-construisent cette connaissance socioculturelle mais à toute personne ayant accès à la plateforme. Puisque l'Internet est ouvert, l'accès à la plateforme est universel pour peu que l'individu ait accès à l'Internet.

Le deuxième point de différence est l'exécution des tâches. En effet, pour Vygotsky bien que les tâches se fassent en collaboration avec les autres pairs dans la ZPD, l'objectif est que l'individu participe à ces tâches pour acquérir de la connaissance. Ainsi, dans cette interaction l'individu n'est pas un acteur passif. Par contre dans notre approche, l'action est menée de façon collective et in fine c'est le résultat de ces interactions qui est le produit de la co-construction. La principale différence est que chez Vygotsky la connaissance individuelle est mise en avant alors que dans

Aspects	Approche de Vygotsky	Approche au sein de la Plateforme
Portée	Individuelle (principalement)	Universelle
Exécution des tâches	Individuelle (principalement)	Collectivement
But	Apprentissage	Apprentissage & Co-construction
Domaine d'Application	Éducation (principalement)	Culture

Table 4.1: Comparaison de l'approche de Vygotsky (socioculturelle) et celle utilisée dans la plateforme.

notre approche c'est la connaissance collective qui est mise en avant. De plus nous comptons sur la complémentarité des pairs pour construire cette connaissance, ce qui n'est pas le cas de l'approche de Vygotsky.

Le troisième point de différence est le but ou l'objectif de l'activité. Bien que les deux approches partagent l'apprentissage comme objectif, notre approche propose un objectif en plus qui est la co-construction. En effet, puisque nous avons étendu l'approche de Vygotsky à une communauté, l'objectif de cette communauté est la co-construction de son patrimoine socioculturel². Et comme l'accès à une plateforme Web est universel, l'objectif des utilisateurs passifs (non membre de la communauté) est l'apprentissage bien qu'ils puissent aussi participer à cette co-construction.

Le dernier point de différence est le domaine d'application. L'approche Vygotsky est souvent utilisée dans l'éducation pour permettre à des pairs (apprenants) d'améliorer leur connaissance. Ce qui nécessite la plupart du temps l'existence d'un expert (enseignant) pour coordonner et dispenser le savoir. Par contre, dans notre approche, le domaine d'application est culturel ce qui ne nécessite pas un expert. En plus, la connaissance culturelle est un bien non exclusif (c'est-à-dire un bien auquel tout le monde peut accéder en même temps) donc nous avons besoin de la complémentarité des uns et des autres pour faciliter cette construction.

Dans la deuxième partie du chapitre 2 nous avons présenté différentes typologies au sein des communautés virtuelles. Les communautés utilisatrices de la plateforme s'apparentent aux communautés de transmission de savoir (chapitre 2) et particulièrement aux communautés de co-élaboration de connaissances (*Knowledge-building community* - KBC) [50, 51, 52]. Ainsi dans le cadre scolaire, la KBC est une stratégie pédagogique socio-constructiviste où le développement de la connaissance ne se fait pas simultanément mais de classe en classe. Ce type d'apprentissage cadre

²Le patrimoine socioculturel est l'ensemble des traditions ou les expressions vivantes héritées de nos ancêtres et transmises à nos descendants, comme les traditions orales, les arts du spectacle, les pratiques sociales, rituels et événements festifs, les connaissances et pratiques concernant la nature et l'univers ou les connaissances et le savoir-faire nécessaires à l'artisanat traditionnel.

Aspect	C3E	C4
Domaine	Éducation	Culturel
Impact de la Connaissance	Dans la Salle de Classe	Générationnel
Expert du Domaine	OUI	NON
Durée	Limitée	Infinie

Table 4.2: Comparaison de la Communauté de Co-élaboration de Connaissances Éducationnelles - C3E (Educational Knowledge-building Community - EKBC et de la Communauté de Co-élaboration de Connaissances Culturelles - C4 (Cultural Knowledge-building Community - CKBC)

avec l'objectif de la plateforme bien que le domaine d'application soit différent.

En effet, l'objectif général est de renforcer l'approche de développement socio-économique fondée sur les connaissances en mettant en œuvre un cadre durable permettant aux communautés d'identifier et de partager leurs ressources (en termes de connaissances, de savoirs, de potentiels, de projets, etc.) et leurs besoins (manque d'infrastructures, de moyens, etc.). Ce cadre fournit une vue d'ensemble des opportunités et des problèmes au niveau de tout le territoire, des informations venant de la base (des acteurs au quotidien dans les différentes communautés), plutôt que des éléments macro éloignés de la réalité du terrain. Ainsi, nous proposons d'étendre certains principes de la KBC dans le domaine culturel.

Les principaux points de différences entre la Communauté de Co-élaboration de Connaissances Éducationnelles (Educational Knowledge-building Community³ - EKBC) et notre approche, Communautés de Co-élaboration de Connaissances Culturelles (Cultural Knowledge-building Community - CKBC), peuvent être résumés comme suit (voir le tableau 4.2). Le premier point de différence se trouve dans le domaine d'application. Puisque la EKBC est plutôt utilisée dans le domaine de l'éducation, nous appliquons les CKBC dans le domaine culturel. Le mot culture fait référence aux langues, aux croyances, aux valeurs et normes, aux coutumes, à l'habillement, aux nourritures, aux rôles, aux connaissances et habilités et toutes autres choses que les personnes apprennent pour représenter la manière de vivre d'une société. Ainsi, tous les points de différence entre EKBC et CKBC résultent principalement du domaine d'application.

Le deuxième point de différence est l'impact de la connaissance. En effet, pour une EKBC la portée de la connaissance se limite à la classe, c'est-à-dire aux membres de la communauté, et ce sont les seuls à participer à cette élaboration. Tandis que pour la CKBC, l'impact de la connaissance passe de génération en génération. En effet, le patrimoine culturel est la *somme* des patrimoines des générations passées et de celle présente. Cet ensemble sera transmis aux générations futures. De plus,

³Puisque la KBC est utilisée dans le domaine de l'éducation, nous avons décidé de l'appeler Educational Knowledge-building Community (EKBC)

un acteur en dehors de la communauté a la possibilité d'apporter des informations par rapport à ce patrimoine.

Le troisième point de différence est le besoin d'un expert du domaine. Puisqu'une EKBC est dans le milieu scolaire, tout le travail de co-élaboration se fait grâce à un enseignant (expert) qui a un rôle essentiel dans cet environnement. Par contre au sein de la CKBC il n'y a pas d'expert mais plutôt de la complémentarité entre les membres. Ainsi, tous les membres ont le même degré d'importance au sein de la communauté bien qu'il puisse y avoir des profils d'experts qui se dégagent. Cependant, ces profils d'experts sont cachés dans le résultat final qui est le produit de la communauté.

Le dernier point de différence est la durée de la co-élaboration de connaissances. Puisque la EKBC est dans le milieu éducatif, la durée de la co-élaboration de connaissances est assujettie à la durée de la formation donc limitée dans le temps. Pour la CKBC, la durée est illimitée puisque le patrimoine culturel transcende les générations et que sa construction se fait d'une génération à l'autre.

4.3 Ontologies Temporelles

Dans la section 3, nous avons présenté plusieurs travaux sur la modélisation temporelle dans le Web sémantique. Ces travaux ont été regroupés en deux groupes que nous avons appelés : 1) *la modélisation d'expressions temporelles* qui permet de modéliser des intervalles, des dates, ... et 2) *l'annotation temporelle des données* qui permet d'annoter temporellement des triplets dans le but de conserver l'évolution des données (changement de valeur) dans le temps.

Dans le domaine de la modélisation d'expressions temporelles les ontologies les plus complètes sont OWL-Time [76, 77], CNTRO [68, 69] utilisée dans le domaine médical et TimeML [70] utilisée dans le TALN.

Le tableau 4.3 présente notre comparaison des trois ontologies les plus complètes dans le domaine de la modélisation d'expressions temporelles suivant sept critères. Ces critères portent entre autres sur la modélisation des temps humains qui sont les expressions temporelles que la personne utilise dans son discours. Au vu de ces critères, l'ontologie CNTRO est la plus complète. Cependant, elle présente certaines limites. En effet pour les intervalles non-convexes, CNTRO modélise la périodicité dans des chaînes de caractères ce qui diminue considérablement le raisonnement temporel sur les données et la perte de la sémantique. De plus pour le temps déictique, CNTRO utilise aussi des chaînes de caractères ce qui produit les mêmes effets qu'avec les intervalles non-convexes. Notons aussi que OWL-Time est la seule ontologie qui permet de définir des intervalles non-convexes plus complexes avec des exceptions.

Dans notre approche de la modélisation d'expressions temporelles, ces sept critères sont essentiels dans le domaine socioculturel et particulièrement pour la modélisation d'expressions humaines du temps qui sont liées à la culture [13].

	OWL-Time	CNTRO	TimeML
Intervalle non-convexe	✓	✓	✓
Intervalle Infini	✓	×	✓
Temps Déictique	×	✓	✓
Référence Temporelle	×	✓	×
Différentes Unités Temporelles	✓	✓	✓
Modélisation Qualitative	×	×	×
13 Relations d'Allen	✓	✓	partiellement

Table 4.3: Comparaison des différentes approches dans le domaine de la modélisation d'expressions temporelles (✓ signifie que l'ontologie intègre cette caractéristique sinon ×).

Ainsi, la satisfaction de ces sept critères permet de pouvoir exprimer des notions de temps fixe, régulier et linéaire (temps biologique et physique [13]) et des notions de temps subjectifs liés à la perception du temps (temps individuel [13]). La modélisation qualitative est importante dans le domaine culturel dans la mesure où elle dépend de la culture et de la position géographique. Cependant, elle n'est prise en considération par aucune des approches présentées dans le domaine de la modélisation d'expressions temporelles (cf. sous section 3.3).

Dans le domaine de l'annotation temporelle des données il existe six grandes approches (Tableau 4.4). Trois sont des initiatives du W3C dont l'objectif est la représentation de relations n-aires qui peuvent être utilisées pour la représentation temporelle qui nécessite une représentation ternaire. Il en existe aussi trois autres qui sont spécifiques au domaine temporel [79, 80, 88].

Le tableau 4.4 compare ces six approches suivant trois critères. Ces critères permettent d'évaluer les ontologies par rapport à:

1. La prolifération des objets (création d'objets intermédiaires pour ajouter la temporalité) et l'introduction de redondances (duplication des données).
2. La perte de l'assertion des triplets : en effet, une temporalité concerne un triplet ou un ensemble de triplets (*address(John, Dakar)*). La perte de l'assertion des triplets intervient quand la relation (sémantique) entre les deux objets (*John* et *Dakar*) disparaît.
3. L'extension des standards du W3C : ce sont des extensions syntaxiques et sémantiques.

Comme on le voit dans le tableau 4.4, pour toutes les approches exceptée TDL il existe une prolifération des objets et une redondance des données. Seules les approches basées sur les graphes nommés, Temporal RDF et TDL assertent les triplets initiaux. Pour les extensions des standards du W3C, seules les approches basées sur Temporal RDF et TDL nécessitent cela. En résumé, la meilleure approche

	Prolifération / Redondance des objets	Perte de l'assertion des triplets	Extension des standards du W3C
Réification RDF	✓	✓	×
Graphes nommés	✓	×	×
Relations n-aires	✓	✓	×
Temporal RDF	✓	×	✓
4D-Fluents	✓	✓	×
TDL	×	×	✓

Table 4.4: Comparaison des différentes approches dans le domaine de l'annotation temporelle des données (✓ signifie que l'ontologie intègre cette caractéristique sinon ×).

est celle basée sur les TDLs mais elle nécessite de définir de nouvelles sémantiques et l'ajout de nouveaux opérateurs. La deuxième meilleure approche est celle basée sur les graphes nommés mais elle n'est optimale que si plusieurs triplets partagent le même intervalle de temps. Il y a aussi le problème de chevauchement entre les graphes; par exemple quand une ressource est présente dans deux graphes dont les périodes de validité se recoupent : cela crée un problème pour le raisonnement temporel⁴. Les quatre autres approches (4D-Fluents, Réification RDF, Relations n-aires et Temporal RDF) présentent plus ou moins les mêmes inconvénients bien que Temporal RDF soit la seule à conserver l'assertion entre les triplets mais son principal inconvénient est l'extension de la sémantique de RDF.

Nous nous rendons compte que la prolifération et la redondance des données sont des maux nécessaires, cependant en fonction des approches nous pouvons diminuer cela. Ainsi, dans notre approche les deux critères que nous jugeons les plus importants : 1) utiliser les standards du W3C et 2) asserter les triplets.

4.4 Conclusion

Dans ce chapitre nous avons présenté les critères qui nous permettent d'évaluer les principales approches de la littérature. Dans le domaine de la capitalisation de connaissances, nous avons présenté l'approche de Vygotsky et celle de Leont'ev-Engeström. Le domaine d'application de la plateforme est plus approprié à l'approche de Vygotsky (socioculturelle) qui cadre beaucoup plus à un processus d'apprentissage collaboratif et de co-construction. Nous avons aussi montré : 1) les différences qui existent entre le domaine d'application de la plateforme et le cadre de Vygotsky et 2) comment le cadre de Vygotsky pourrait être étendu dans la plateforme.

Nous avons ensuite étudié la littérature relative aux différents types de communautés virtuelles afin de caractériser celles pour lesquelles nous avons travaillé.

⁴<https://www.w3.org/TR/rdf11-datasets/#relationship-with-sparql-entailment-regime>

Nos communautés utilisatrices sont une extension de la KBC. Puisque la KBC est appliquée dans le domaine de l'éducation, EKBC, nous avons proposé de l'étendre dans le domaine culturel appelées CKBC. Certains critères ont été mis en évidence dans le tableau 4.2 pour différencier les deux domaines d'applications.

Dans la deuxième partie de ce chapitre, nous avons également présenté les critères qui permettent d'évaluer les principales approches de la littérature dans le domaine de la modélisation temporelle dans le Web sémantique. Le travail se situe dans deux aspects de modélisations identifiés. Le premier est la modélisation d'expressions temporelles où le principal objectif est de modéliser le temps humain lié aux phénomènes socioculturels. Ainsi, nous avons proposé une ontologie pour modéliser le temps déictique, la modélisation qualitative et les intervalles non convexes et infinis.

Le second aspect de modélisation est l'annotation temporelle des données. Le tableau 4.4 présente les quatre principaux critères pour notre modélisation. Deux nous semblent particulièrement pertinents : 1) la conservation de l'assertion entre les triplets et 2) l'utilisation des standards du W3C.

Part II

Contributions

Capitalisation de Connaissances : Ontologie Socioculturelle et Communauté de Co-élaboration de Connaissances Culturelles

Sommaire

5.1	Introduction	73
5.2	Une ontologie socioculturelle pour le partage et la co-construction de connaissances	74
5.2.1	Extension du modèle de Vygotsky	75
5.2.2	Ontologie de haut-niveau	75
5.2.3	Alignement de schémas	78
5.2.4	Cas d'utilisation de USCO	80
5.3	Communauté de co-élaboration de connaissances culturelles	82
5.3.1	Introduire les CKBC aux membres	85
5.3.2	Technologie utilisée : MediaWiki	86
5.3.3	Animation des communautés utilisatrices	88
5.4	Synthèse	89

5.1 Introduction

La première partie de cette thèse est consacrée à un état de l'art sur, d'une part, la capitalisation de connaissances socioculturelles et, d'autre part, l'annotation temporelle de ces connaissances. Nos contributions en matière de capitalisation de connaissances se situent à deux niveaux. Le premier niveau est l'utilisation (extension) de la CHAT (Théorie Historico-Culturelle de l'Activité) comme méta-modèle pour la modélisation de l'ontologie socioculturelle. La CHAT est un cadre conceptuel issu de la tradition socio-culturelle de psychologues russes qui est compris comme l'interaction, la transformation et le développement entre les acteurs (sujets) et le

monde (objets). Le deuxième niveau résulte de l'extension de la CHAT à une communauté qui fait apparaître une communauté utilisatrice qu'il faudrait positionner parmi les différents types de communautés virtuelles existantes.

Deux approches dans l'application de la CHAT sont identifiées. Celle socioculturelle qui suit les travaux de Vygotsky [28, 23, 30, 20] et celle de la théorie de l'activité qui suit les travaux de Leont'ev-Engeström [25, 32, 33, 34, 24]. L'approche utilisée est celle socioculturelle qui fait référence au contexte social de l'individu pour promouvoir son développement psychique (connaissance). Cependant, cette approche est placée dans le contexte social d'une communauté qui est un ensemble de personnes ayant les mêmes aspirations.

L'état de l'art a aussi montré l'existence de différentes typologies de communautés virtuelles. Parmi ces typologies, nous nous intéressons à celle se focalisant sur la transmission de savoir et sur l'orientation sociale au sein des communautés. Il existe quatre grandes familles dans cette typologie dont celle rassemblant les communautés de co-élaboration de connaissances (Knowledge-building Community - KBC) [50, 52]. On retrouve ce type de communauté dans le domaine de l'éducation (communauté de co-élaboration de connaissances éducationnelles¹ [51, 52]). Dans cette thèse, nous préconisons la formation de KBC dans le domaine de la culture. Nous parlons donc de communauté de co-élaboration de connaissances culturelles (Cultural Knowledge-building Community - CKBC). Cette approche fournit aux communautés un cadre leur permettant de co-construire (co-élaborer) leurs connaissances socioculturelles.

Ce chapitre est composé de deux grandes parties. Dans la première partie de ce chapitre nous présentons les différentes phases de développement de l'ontologie socioculturelle de haut-niveau (Upper-level SocioCultural Ontology - USCO). Premièrement, nous présentons l'extension du modèle de Vygotsky à une communauté. Ensuite, nous présentons l'ontologie issue de cette extension du modèle de Vygotsky. Après cela, concernant la modélisation de l'ontologie de haut-niveau USCO, nous présentons l'alignement du schéma de USCO avec certains vocabulaires du LOD (Linked Open Data). En dernier lieu, nous présentons comment l'ontologie USCO peut être utilisée pour la construction de connaissances socioculturelles. Dans la deuxième partie de ce chapitre nous présentons l'adaptation de la notion de KBC au domaine culturel. Enfin, nous terminons ce chapitre par une synthèse.

5.2 Une ontologie socioculturelle pour le partage et la co-construction de connaissances

Dans cette sous section, nous présentons une méthodologie, en trois étapes, de construction d'une ontologie socioculturelle de haut-niveau (USCO²). La première étape

¹Educational Knowledge-building Community - EKBC

²USCO : Upper-level SocioCultural Ontology.

est l'extension du modèle socioculturel de la CHAT³ (Théorie Historico-Culturelle de l'Activité) à une communauté. La deuxième étape consiste en la définition des concepts et des propriétés de l'ontologie grâce à l'extension de la CHAT. La dernière étape utilise des vocabulaires du LOD pour diminuer la lourdeur du processus de construction de l'ontologie.

5.2.1 Extension du modèle de Vygotsky

Conformément à l'approche de Vygotsky [17] de la CHAT, nous considérons le processus de co-construction de connaissances comme étant l'interdépendance entre un individu ou groupe d'individus avec son environnement social. L'approche de Vygotsky est matérialisée par trois entités (figure 2.1, page 14): 1) un *Sujet* qui représente un individu engagé dans un processus, 2) un *Objet* qui peut-être le but de l'activité, les motivations pour participer à cette activité ou les matériels produits ou utilisés durant cette activité et 3) un *Artefact* qui représente la médiation entre *Sujet* et *Objet*.

Dans cette thèse, nous proposons de substituer l'individu (*Sujet*) à sa communauté et ainsi s'intéresser à l'action de la communauté au sein de son environnement social. Cette approche est moins portée sur les individus (qui sont classiquement les points centraux d'une communauté) que sur les connaissances que ceux-ci partagent. Ce changement de point de vue nous permet d'aborder une communauté comme une entité atomique. Ceci nous permet d'analyser l'impact des actions de la communauté sur son environnement.

Le mot russe traduit par *objet* a plusieurs significations [24, 26]. Cependant, nous considérons comme *objet* toute chose (matériel) qui se trouve dans l'environnement de la communauté et que cette communauté utilise dans ses activités. Nous considérons comme *Artefact* tout objet immatériel qui sert de médiation entre *communauté* et *objet*.

L'approche proposée adopte l'approche socioculturelle de la CHAT [20] qui considère la médiation comme unité de base de l'analyse donc de la co-construction de connaissances. En effet, c'est à travers les objets de la médiation que nos communautés vont partager et co-construire leur patrimoine culturel. Ces objets de médiation font le lien entre les communautés et les objets de leur environnement. Ils permettent également de savoir comment les communautés s'approprient les objets de leur environnement pour le développement de leur culture.

5.2.2 Ontologie de haut-niveau

Une ontologie de haut-niveau⁴ est une ontologie qui décrit les concepts (classes) et les propriétés (relations et attributs) généraux qui ont des significations partagées à travers le monde. L'objectif principal d'une ontologie de haut niveau est de fournir

³Cultural Historical Activity Theory

⁴Appelée parfois ontologie supérieure ou ontologie fondamentale.

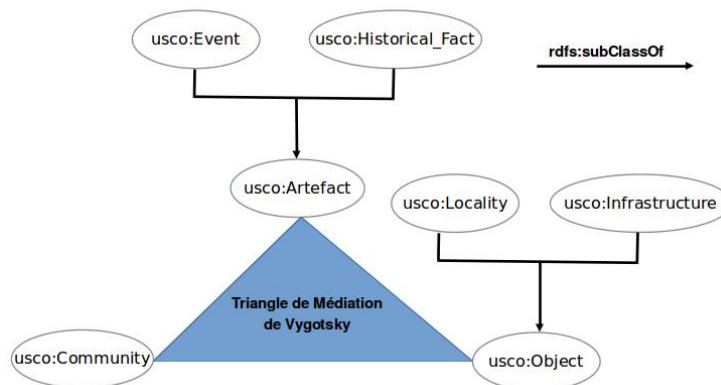


Figure 5.1: Ontologie Socioculturelle de haut niveau (USCO).

une très large interopérabilité sémantique entre différentes ontologies qui sont utilisées (développées) dans des domaines d'applications différentes.

L'extension du modèle de Vygotsky de la CHAT (socioculturelle) est utilisée comme point de départ pour construire l'ontologie de haut-niveau (Fig. 5.1):

- Le concept de communauté : une communauté est un groupe de personnes partageant une histoire, une culture, une ethnicité ou bien un intérêt et souhaitant échanger ou collaborer grâce au Web pour le partage de leurs connaissances concernant ce domaine.
- Le concept d'objet : regroupe toute entité qui se trouve dans l'environnement de la communauté. Particulièrement, il est composé de deux entités: 1) des localités qui sont des lieux d'habitation et 2) des infrastructures qui sont des bâtiments ou des espaces non bâtis dans l'environnement des communautés.
- Le concept d'artefact : nos outils de médiations sont principalement les événements organisés par les communautés dans leur environnement et les faits historiques qui se produisent dans l'environnement de la communauté.

L'ontologie socioculturelle de haut-niveau que nous proposons, *Upper-level Socio-Cultural Ontology* (USCO), permet d'avoir une connaissance sur les communautés mais également sur les activités que celles-ci organisent afin de pérenniser et de vulgariser leurs activités culturelles. USCO permet aussi de connaître les infrastructures et les faits historiques relatifs aux environnements des communautés.

Les concepts (classes) seuls ne fournissent pas assez d'information, il faut leur associer des attributs qui jouent un rôle essentiel dans le développement d'une ontologie. Les attributs décrivent les caractéristiques des classes et des instances. Comme le processus de définition de tous les attributs des concepts du domaine socioculturel est conséquent, nous avons proposé un moyen d'exploiter les connaissances du LOD pour définir ces attributs. Cela est présenté dans la sous section 5.2.3.

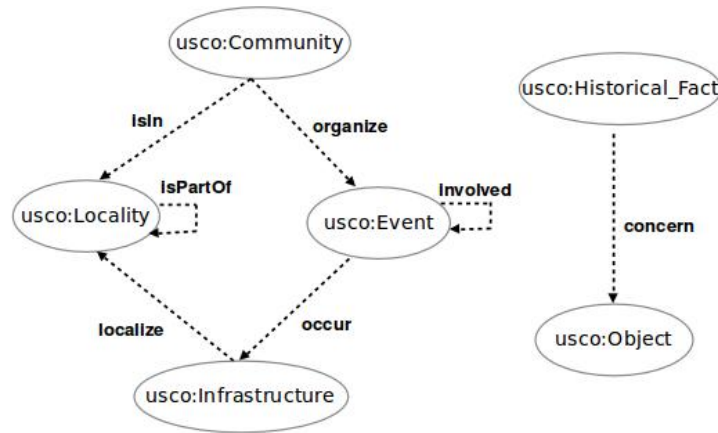


Figure 5.2: Relations entre les concepts.

Une autre chose importante dans le processus de développement d'une ontologie est de définir les relations entre les concepts. Un choix de modélisation qui doit être fait durant l'élaboration d'une ontologie est de décider si une connaissance doit être modélisée comme un attribut ou à l'aide d'une relation avec d'autres concepts. Un critère peut être de dire que c'est un attribut dès lors que les valeurs possibles sont d'un type dit primitif (entier, chaîne de caractères, etc), et que c'est une relation dès lors que les valeurs possibles sont d'un type dit complexe c'est-à-dire d'un autre concept de l'ontologie. Ainsi la figure 5.2 illustre les différentes relations qui existent entre les concepts de haut niveau de l'ontologie socioculturelle USCO.

Ces propriétés ont été définies en se basant sur le triangle de médiation de Vygotsky. En effet, selon ce triangle l'artefact joue le lien entre une communauté et l'objet. Ainsi, nous pouvons retrouver le triangle de médiation de Vygotsky à différents niveaux (*Community - Event - Infrastructure* et *Community - Historical_Fact - Locality*). Ces propriétés permettent de représenter différentes connaissances socioculturelles :

- Les propriétés *organize* et *occur* permettent de retrouver le triangle de médiation de Vygotsky. Ces propriétés permettent de modéliser les centres d'intérêts des communautés et comment elles utilisent les ressources de leur environnement. La propriété *involved* permet de savoir quelles sont les différentes activités associées à un événement si elles existent.
- Les propriétés *concern* et *isIn* permettent aussi de retrouver le triangle de médiation de Vygotsky. Ces propriétés permettent de modéliser les faits historiques relatifs à une communauté ou à son environnement.
- En combinant les propriétés *concern*, *isIn* et *localize* nous disposons de plus d'informations sur l'environnement d'une communauté. Une combinaison des

propriétés *concern* et *occur* permet de modéliser l'utilisation des ressources (objet) à la disposition des communautés dans leurs activités.

5.2.3 Alignement de schémas

L'alignement de schémas (*Schema Matching*) consiste à partir de structures de graphes (schéma) à établir une correspondance sémantique, en informatique, entre les nœuds de ces graphes. Un schéma est une structure formelle qui représente un artefact en ingénierie, comme un schéma SQL, un schéma XML, un schéma entité-relation ou la description d'une ontologie. Une correspondance sémantique consiste à trouver une équivalence sémantique entre un ou plusieurs éléments d'un schéma et un ou plusieurs éléments d'un autre.

Comme l'a montré Noy [90], il existe deux approches pour la découverte de correspondances entre des schémas ontologiques. La première approche consiste à utiliser une ontologie de haut-niveau largement acceptée par la communauté et à trouver des correspondances avec ce « schéma partagé ». La seconde approche repose sur des techniques heuristiques ou d'apprentissage automatique pour trouver les correspondances. L'approche utilisée dans cette thèse appartient à la première famille d'approches. En effet, nous considérons les vocabulaires du Linked Open Data (LOD) comme des vocabulaires de haut-niveau dans lesquels nous avons identifié des concepts similaires aux nôtres. Ces correspondances nous permettent d'exploiter les descriptions effectuées sur ces schémas et plus particulièrement les attributs définis dans ces schémas.

Le LOD est un mode de publication sur le Web des données structurées à l'aide de méta-données connexes permettant à ces dernières d'être connectées et enrichies. Un des principes importants du LOD est l'accessibilité des données sur le Web (données ouvertes).

L'alignement fait repose sur deux vocabulaires généraux (cf. figure 5.3). Le premier est Schema.org⁵ qui fournit une collection de schéma pour des données structurées sur les pages Web grâce à des balises. C'est une initiative de Yahoo, Bing et Google. Il est largement utilisé aujourd'hui par différentes organisations pour la publication de données sur le Web. Le deuxième schéma est le vocabulaire DBpedia. DBpedia est une base de connaissances extraite des *Infobox*⁶ de Wikipedia. Actuellement, la version anglaise de la base de connaissances de DBpedia décrit 4,58 millions de choses, dont 4,22 millions sont classées dans une ontologie cohérente⁷. L'ontologie couvre de nombreux domaines et représente un accord communautaire réel. La base de connaissances évolue automatiquement en fonction des changements de Wikipedia.

⁵schema.org

⁶Un Infobox est une table conçue pour être ajoutée dans le coin supérieur droit d'une page Wikipedia qui présente un résumé de la page et parfois améliore la navigation à d'autres articles connexes

⁷<http://wiki.dbpedia.org/about>, 04/01/2016.

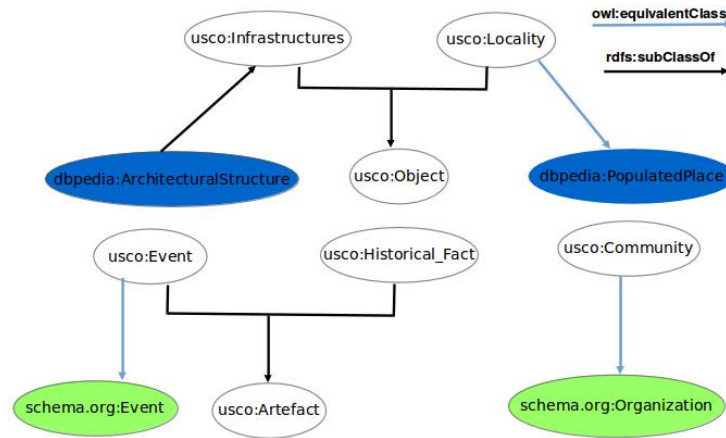


Figure 5.3: Alignement avec Schema.org et DBpedia.

Notre alignement est fait manuellement en utilisant deux opérateurs : 1) *owl:equivalentClass* qui est un axiome de OWL qui définit que deux concepts (classes) ont la même extension de classe c'est-à-dire que les deux extensions de classe contiennent exactement le même ensemble d'individus (instances) et 2) *rdfs:subClassOf* qui est un axiome de RDFS qui permet de définir qu'un concept C1 est le sous concept d'un autre concept C2, ainsi l'ensemble des individus de C1 est un sous-ensemble de l'ensemble des individus de C2.

Le premier alignement est fait avec le schéma de schema.org, grâce à l'opérateur *owl:equivalentClass*, entre les concepts *schema.org:Organization* et *usco:Community* et entre les concepts *schema.org:Event* et *usco:Event* (figure 5.3). Un *schema:Organization* est une entité, comprenant plusieurs personnes, comme une institution ou une association, qui a un but collectif. Le concept est utilisé dans plus d'un million de domaines d'application et compte plus d'une quinzaine d'attributs⁸. Un *schema.org:Event* permet de modéliser un événement (activité) qui se passe à un certain moment et dans un certain lieu, comme un concert, une conférence, un festival, etc. Le concept est utilisé dans cent mille à deux cent cinquante mille domaines d'applications et compte au moins une dizaine d'attributs⁹. Ainsi, l'alignement fait avec les deux concepts de Schema.org permet de bénéficier des attributs définis pour ces derniers. De plus, grâce à l'utilisation d'un raisonneur¹⁰, nous sommes en mesure de dire que toute entité qui est une instance de *schema.org:Organization* (respectivement de *schema:Event*) est aussi une instance de *usco:Community* (respectivement de *usco:Event*).

⁸<http://schema.org/Organization>

⁹<http://schema.org/Event>

¹⁰Un raisonneur est un logiciel qui permet de déduire des conséquences logiques d'un ensemble de faits et d'axiomes.

Le deuxième alignement est fait avec le schéma de DBpedia¹¹. Premièrement, grâce à l'opérateur *owl:equivalentClass*, nous avons aligné le concept *dbpedia:PopulatedPlace* avec le concept *usco:Locality*. Un *dbpedia:PopulatedPlace*¹² est un endroit ou une région qui peut être référencé grâce à ses coordonnées géographiques et dans lequel se trouve une population humaine permanente (ville, village, quartier, etc.). Le concept *dbpedia:PopulatedPlace* compte plus d'une cinquantaine d'attributs. Deuxièmement, grâce à l'opérateur *rdfs:subClassOf*, nous avons aligné le concept *usco:Infrastructure* à tous les sous concepts de *dbpedia:Place* excepté le concept *dbpedia:PopulatedPlace* qui est l'équivalent de *usco:Locality*. À titre illustratif dans la figure 5.3 nous avons représenté l'alignement avec le concept *dbpedia:ArchitecturalStructure* qui est un sous concept du concept *dbpedia:Place*.

L'alignement avec les schémas DBpedia et Schema.org a un double intérêt: 1) bénéficier des descriptions faites sur ces schémas (notamment les attributs) et 2) comme la création d'une base de connaissances (BC) est prévue (cf. section 7), cet alignement permet aussi de bénéficier du contenu de la base de connaissances de DBpedia qui est universelle et ainsi d'en extraire des données relatives au domaine contexte d'utilisation.

Le processus de construction de l'ontologie USCO est manuel. L'ontologie USCO fournit un cadre conceptuel pour des communautés afin de partager et de co-construire leur patrimoine socioculturel. La méthodologie proposée est fondée sur l'extension de la CHAT à une communauté. Pour alléger la construction de cette ontologie, un alignement est proposé avec deux vocabulaires du LOD, ce qui a permis de décrire les objets de l'ontologie USCO.

5.2.4 Cas d'utilisation de USCO

L'ontologie USCO est utilisée pour permettre à des communautés de partager et de co-construire leur patrimoine socioculturel. Cela se fait à travers les descriptions faites des individus instances d'événements socioculturels organisés par ces communautés, les descriptions faites des ressources à la disposition de ces communautés mais également les descriptions faites sur ces communautés elles-mêmes. Cette sous section montre comment l'ontologie USCO peut être utilisée pour le partage et la co-construction de connaissances socioculturelles.

Comme exemple illustratif d'utilisation de l'ontologie USCO, nous avons choisi l'individu *Mbappat* qui est un tournoi de luttes traditionnelles qui regroupe différentes équipes et au cours duquel d'autres activités peuvent être proposées.

La figure 5.4 représente les informations principales pour caractériser l'événement *Mbappat*. Cet événement intègre une activité comme *Louma* qui est un marché occasionnel. Les attributs ne sont pas représentés pour ne pas surcharger l'illustration. Comme exemple d'attributs, nous pourrions ajouter la durée des événements (*schema:duration*), les positions géographiques des infrastructures (*dbpedia:latm*),

¹¹<http://mappings.dbpedia.org/server/ontology/classes/>

¹²<http://mappings.dbpedia.org/server/ontology/classes/PopulatedPlace>

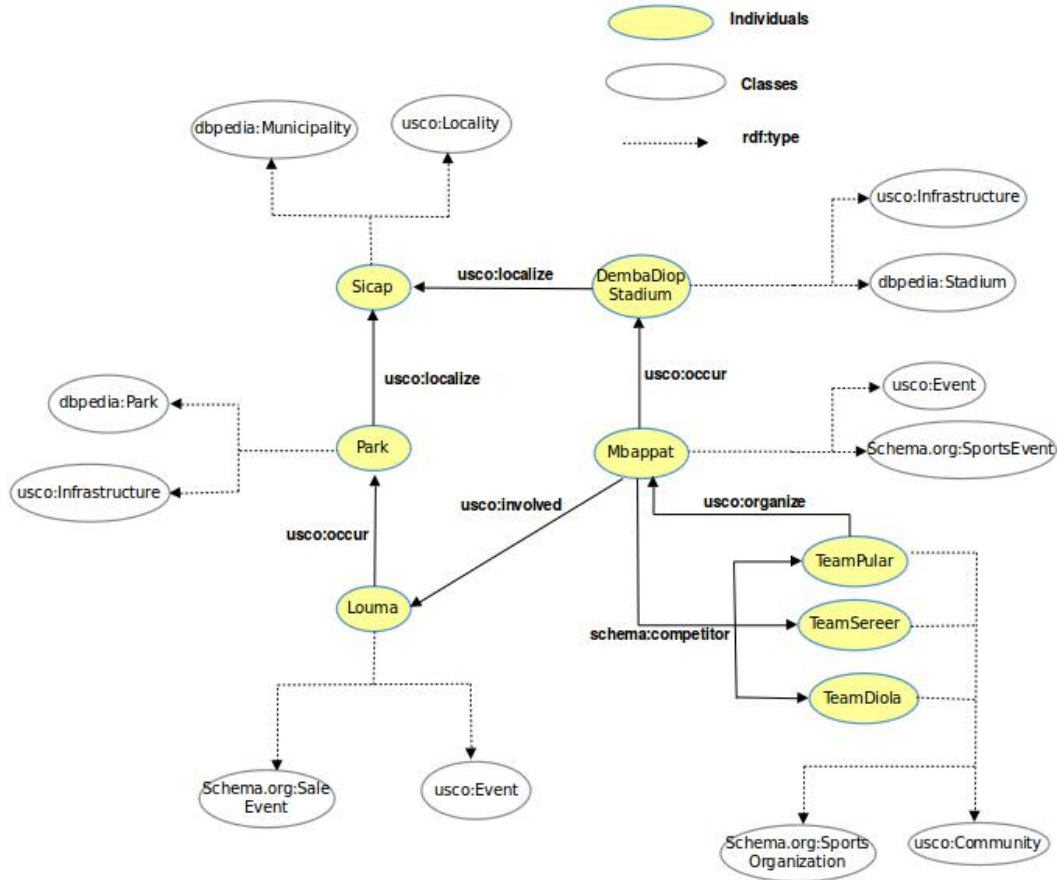


Figure 5.4: Cas d'utilisation de l'ontologie USCO.

etc. Cet exemple montre comment les connaissances sur les événements qui se produisent dans la municipalité de *Sicap* mais également sur les infrastructures qui s'y trouvent peuvent être modélisées à l'aide de l'ontologie USCO. L'analyse de ces connaissances permet par exemple de connaître l'utilisation que les communautés font des ressources à leur disposition. Elle peut également permettre de comprendre que le *Mbappat* regroupe des équipes souvent constituées en fonction de l'appartenance ethnique ou de l'appartenance à un même terroir (*TeamPular*, *TeamSereer* et *TeamDiola* sont des regroupements ethniques).

L'utilisation de l'ontologie USCO à une plus grande échelle et de façon systématique permet d'avoir une connaissance générale sur les communautés, les infrastructures, les activités, les localités et les faits historiques en relation avec ces derniers. Notons que le processus de construction de cette base de connaissances est collaboratif car chaque utilisateur a la possibilité d'apporter des informations sur les instances (individus) de l'ontologie USCO.

5.3 Communauté de co-élaboration de connaissances culturelles

Conformément à l'approche proposée par Scardamalia et Bereiter [50, 51, 52], nous considérons la plateforme de partage de connaissances socioculturelles comme un domaine de développement socio-constructiviste. Le socio-constructivisme est un processus d'apprentissage durant lequel les individus construisent leurs connaissances grâce aux interactions sociales au sein de leur environnement. Ainsi, selon la vision de Vygotsky [17], la plateforme constitue une Zone Proximale de Développement (ZPD) représentant tout ce que la personne n'est capable de connaître qu'avec l'aide et l'échange avec une personne ou avec un groupe de personnes. Ainsi les progrès au sein de cet espace communautaire génèrent continuellement de nouvelles avancées, avec des problèmes reformulés qui prennent un plus large éventail de connaissances en considération [51].

Ainsi, la co-élaboration de connaissances peut être définie comme la production et l'amélioration continue des idées de valeur d'une communauté [51]. Elle ne se résume pas à la somme des contributions individuelles mais plutôt à un échange (confrontation) d'idées entre les individus. La co-élaboration de connaissances, ainsi, va tout au long d'une vie d'une communauté et ne se limite pas à l'éducation. Appliquée à la culture, la co-élaboration de connaissances engage les individus dans un processus d'appropriation, de transmission, de production et de création de connaissances caractérisant une communauté.

Nous considérons une communauté comme un groupe de personnes partageant une histoire, une culture, une ethnicité ou bien un intérêt et souhaitant échanger ou collaborer grâce au Web pour le partage de leurs connaissances concernant ce domaine. Nos communautés sont caractérisées par trois axes : 1) un intérêt socioculturel¹³ commun, 2) un échange, une collaboration et un partage entre les membres et 3) l'utilisation d'Internet pour interagir. La culture et les coutumes sont au centre de tout ordre social au sein d'une communauté. Le mot culture fait référence aux langues, aux croyances, aux valeurs et normes, aux coutumes, à l'habillement, aux nourritures, aux rôles, aux connaissances et habilités et toutes autres choses que les personnes apprennent pour représenter la manière de vivre d'une société. C'est un système de règles qui sont à la base de ce que nous sommes et affectent la façon dont nous nous exprimons dans le cadre d'un groupe et en tant qu'individu. L'ontologie USCO permet de rassembler et d'annoter des connaissances concernant un ensemble de sous-cultures (*subculture*) qui sont une partie de la culture à laquelle appartient un groupe de personnes au sein d'une société. Le tout forme des cultures des gens ordinaires (*folks culture*) qui sont les cultures enracinées dans l'expérience, les coutumes et les croyances de tous les jours des personnes ordinaires dans une société.

Les activités au sein d'une communauté de co-élaboration de connaissances culturelles (Cultural Knowledge-Building Community - CKBC) sont un processus

¹³Aspect lié à la fois à la société et à la culture.

(i) de construction de connaissances à partir de sources (ressources) sociales et culturelles, et (ii) de leur l'intégration dans des connaissances pré-existantes (connaissances des générations précédentes). De ce fait, les membres d'une CKBC participent activement à trouver des idées, à l'identification de problèmes, à la recherche de solutions et de preuves, à débattre et à échanger avec leurs pairs. Ces activités ont un objectif clair pour la co-élaboration de connaissances socioculturelles. Elles permettent aux idées individuelles de se transformer en connaissances collectives.

Scardamalia définit 12 caractéristiques pour une co-élaboration de connaissances agissant comme un système pour faciliter le développement des EKBC¹⁴ [91]. Ces caractéristiques explicitent les processus cognitifs et sociaux dans un contexte collaboratif. Ainsi, chacune de ses caractéristiques est scindée en deux parties : 1) l'aspect socio-cognitif décrit l'objectif et/ou les principes visés par la caractéristique, 2) l'aspect technologique décrit le type de média appoprié pour supporter l'objectif de la caractéristique. Appliquée à la culture (CKBC), les 12 aspects socio-cognitifs des EKBC peuvent être regroupés en 4 caractéristiques :

1. Construction et complémentarité : la construction de connaissances est née d'un manque de connaissances sur un phénomène culturel, d'un vide à combler. La co-construction de connaissances est un engagement collectif des membres pour mieux comprendre leur environnement en analysant (traitant) des thèmes complexes. Par thème complexe, nous entendons un problème qui nécessite la collaboration de plusieurs individus pour en faire avancer la compréhension. Tous les participants ont la possibilité de s'engager dans le processus de co-construction et ont tous accès au savoir. Cette caractéristique regroupe les 4 caractéristiques des EKBC suivantes¹⁵ : démocratisation du savoir¹⁶ - diversité des idées¹⁷ - idées réelles, problèmes authentiques¹⁸ - intégration des idées débattues¹⁹.
2. Amélioration et évolution de la connaissance : la co-élaboration de connaissances culturelles est un long processus et demande une amélioration continue de la qualité et de la cohérence. Ceci permet un approfondissement des idées au regard de la thématique. Puisque la connaissance culturelle est générationnelle, il appartient à chaque génération de participer à la construction de son patrimoine culturel et de conserver les traces des générations antérieures. Ce processus permet également une évolution de la connaissance. Cette caractéristique regroupe les 3 caractéristiques des EKBC suivantes :

¹⁴Educational Knowledge Building Community.

¹⁵traduction extraite de <http://www.telelearning-pds.org/u/pv/princocons.htm>

¹⁶Democratizing Knowledge

¹⁷Idea Diversity

¹⁸Real Ideas, Authentic Problems

¹⁹Rise Above

démarche épistémologique²⁰ - discours transformatif²¹ - idées perfectibles²².

3. Communauté ouverte : puisque la connaissance culturelle est un bien non exclusif (c'est-à-dire un bien auquel tout le monde peut accéder en même temps), le fait que des individus hors de la communauté constituée sur la plateforme puissent participer est essentiel. En effet, tout individu doit être capable d'apporter sa contribution à la construction de la connaissance. De même, cette contribution doit être possible tout au long du processus de construction pour faciliter cette dernière. Cette caractéristique regroupe les 3 caractéristiques des EKBC suivantes : savoir communautaire, responsabilité collective²³ - ubiquité de l'élaboration de connaissances²⁴ - Avancement symétrique du savoir²⁵. Notons que l'aspect communauté ouverte des CKBC ajoute un principe en plus par rapport aux 3 aspects des EKBC. En effet, pour les trois aspects des EKBC cette ouverture signifie que les membres de la communauté ont des privilèges dans l'élaboration de la connaissance. Alors que pour les CKBC, cette ouverture signifie à la fois que les membres de la communauté ont des privilèges dans l'élaboration de la connaissance mais aussi que toute personne (qui a accès à la plateforme) peut apporter sa contribution dans l'élaboration de la connaissance au même niveau que les membres de la communauté. Ainsi, la participation des personnes externes à la communauté est la principale différence entre les CKBC et les EKBC par rapport à l'ouverture de la communauté.
4. Évaluation et remise en question : cet aspect est important pour la construction de la connaissance socioculturelle. En effet, les utilisateurs doivent avoir la possibilité de donner leurs références par rapport aux contenus qu'ils ajoutent ou qu'ils veulent supprimer. Ainsi, la communauté pourra discuter (évaluer) les sources (référencements) afin de permettre de ré-évaluer le contenu des échanges entre les utilisateurs. Cet aspect est aussi facilité par l'accessibilité des contenus à tout utilisateur (communauté ouverte). Cette caractéristique regroupe les 2 caractéristiques des EKBC suivantes : appel à des sources fiables²⁶ - évaluation simultanée, ancrée et transformative²⁷.

En résumé, la première différence principale entre une CKBC et une EKBC au niveau socio-cognitif se trouve au sein de l'environnement de travail (collaboration). En effet, pour une EKBC la collaboration se fait dans un environnement clos c'est-à-dire que toutes les contributions et les résultats sont tributaires seulement des membres de la communauté, alors que pour une CKBC, toutes les étapes

²⁰Epistemic Agency

²¹Knowledge Building Discourse

²²Impovable Ideas

²³Community Knowledge, Collective Responsibility

²⁴Pervasive Knowledge Building

²⁵Symmetric Knowledge Advance

²⁶Constructive Uses Of Authoritative Sources

²⁷Embedded, Concurrent and Transformative Assessment

du processus de co-élaboration et les résultats se font dans un environnement ouvert c'est-à-dire accessible à toute personne qui voudrait y participer. Les résultats d'une CKBC sont accessibles au delà de la communauté et même de la génération des membres de la CKBC. La deuxième différence est la présence d'un expert (enseignant) dans une EKBC. En effet, tout le travail au sein d'une EKBC s'effectue autour d'un enseignant qui coordonne l'action de tous les membres c'est-à-dire qu'il veille à ce que toutes les contributions restent dans les limites de la thématique définie. *A priori*, ce type de profil n'existe pas au sein d'une CKBC car les membres de la communauté ont les mêmes privilèges et c'est à la communauté de veiller à ce que toutes les contributions restent dans les limites de la thématique définie.

5.3.1 Introduire les CKBC aux membres

Créer une CKBC signifie changer la façon dont les gens partagent leurs connaissances, dont ils communiquent et dont ils fonctionnent. Cela signifie également que la structure et les activités d'une telle communauté doivent refléter la diversité des points de vues, encourager les résultats positifs de l'interaction entre les personnes et prendre en charge le partage du pouvoir entre les personnes.

Nous développons tous un certain type de culture. Ainsi, une CKBC doit permettre à des personnes qui partagent certains aspects culturels de pouvoir collaborer pour co-construire cette connaissance. Une fois la communauté créée, les échanges lui permettent d'être plus flexible par rapport aux différents point de vue et de devenir un système collaboratif proactif. Ceci permet à la communauté de se restructurer en permanence pour répondre à l'évolution de l'environnement socioculturel. L'existence d'une CKBC est assujettie aux capacités de ses membres à tous les niveaux qui doivent développer la compréhension culturelle et les compétences de communication appropriées. Ce développement permet l'émergence d'une collaboration et d'une interdépendance entre ses membres.

Pour assurer le développement d'une CKBC, toutes les personnes impliquées développent des modes d'interaction entre elles qui permettent la diversité de point de vue et de prendre en compte les connaissances antérieures. Ces modes d'interaction favorisent la compréhension entre les membres, ce qui est l'une des dynamiques de base dans le développement du travail en équipe. Ainsi pour un bon fonctionnement d'une CKBC, les composants suivants doivent marquer les relations au sein de la communauté:

- Objectif commun : les membres doivent s'assurer qu'ils partagent les mêmes objectifs et la même compréhension de ce qui doit être réalisé. Cela implique que la réalisation de l'objectif commun ne peut être atteint sans l'apport de chaque membre.

- Le partage du pouvoir : chaque membre a quelque chose de différent à apporter dans le processus de collaboration, que ce soit une critique ou un apport substantiel. Cet aspect est précieux dans la mesure où il permet à chaque

membre de participer à un même niveau de responsabilité au processus de co-construction.

- Le respect et la confiance mutuelle : ces deux fonctions doivent être tenues non seulement dans le début de la collaboration, mais tout au long du processus. Cet aspect permet une synergie au sein de la communauté qui arrive seulement si les membres se traitent mutuellement avec respect et communiquent efficacement les uns avec les autres.

Comme les activités des CKBC ont lieu en ligne autant qu'à l'extérieur de l'environnement virtuel, les stratégies de facilitation de la collaboration doivent englober à la fois les composants en ligne et non virtuels. La règle de base pour des CKBC est que la connaissance culturelle doit être intégrée dans toutes les facettes des communautés.

5.3.2 Technologie utilisée : MediaWiki

Un aspect important lors de la création d'une communauté virtuelle est le choix des technologies utilisées pour supporter l'interaction entre utilisateurs. Cependant cette technologie doit favoriser les principes susmentionnés pour une bonne co-élaboration de connaissances socioculturelles.

Nous devons proposer un environnement technologique (i.e. plateforme) qui permette à toute personne capable de participer à cette co-élaboration de le faire. C'est pourquoi nous avons construit notre plateforme à partir d'un Wiki. En effet, l'idée principale d'un Wiki est que n'importe quelle page peut être consultée et éditée par tous utilisateurs. Selon son inventeur « Un Wiki [...] est un site Web (ou toute autre collection de documents hypertextes) qui permet aux utilisateurs d'ajouter du contenu, comme sur un forum Internet, mais permet aussi que ce contenu soit édité par d'autres utilisateurs »²⁸. Ainsi, un Wiki fournit un moyen rapide d'éditer le contenu d'une page Web mais également de collaborer avec d'autres utilisateurs pour la rédaction de ce contenu.

Notre plateforme repose sur MediaWiki²⁹, un moteur de Wiki open-source, qui est utilisé par l'ensemble des projets de la Wikimedia Foundation³⁰. MediaWiki fournit des fonctionnalités qui facilitent une bonne co-élaboration de connaissances:

- Recherche thématique facile : MediaWiki offre une recherche dans le texte intégral ou dans les titres des pages du Wiki. Ainsi, les articles peuvent être consultés rapidement.

²⁸Ward Cunningham dans un de ses interviews en 2003,

<http://www.artima.com/intv/wiki.html>

²⁹<https://www.mediawiki.org/wiki/MediaWiki> .

³⁰Wikimedia Foundation, Inc. est une organisation à but non lucratif qui héberge les projets en ligne Wikipedia, Wiktionary, Wikiquote, Wikibooks, Wikimedia Commons, Wikisource, Wikispecies, Wikinews, Wikiversity, Wikivoyage et Wikidata.

- Facilité de contribution : La création et l'édition d'une page est facile. À partir du résultat de la recherche d'une thématique, MediaWiki propose de créer cette page de thématique si la recherche est infructueuse, MediaWiki propose par un lien de créer cette thématique, ce qui permet de créer une nouvelle page (thématique) dans le Wiki. Si la recherche est fructueuse, un lien permet d'éditer la page. De même pour chaque page, MediaWiki propose un espace, accessible à tous les utilisateurs, où ces derniers peuvent échanger durant la création de contenu. Ceci permet aux utilisateurs de mieux collaborer et de mieux « défendre » leur point de vue. MediaWiki permet également à un utilisateur de suivre l'évolution d'une page grâce à l'option *watchlist*.
- Traçabilité des contributions : MediaWiki propose une fonctionnalité appelée « historique ». Cette fonctionnalité permet de sauver toutes les versions précédentes ou toutes les modifications d'une page. Ainsi, il est possible de suivre exactement le processus d'édition d'un article, puisque tous les changements ont été documentés. L'historique permet de détailler les modifications entre deux versions. Ainsi les auteurs n'auront pas besoin de comparer les deux pages ligne par ligne. L'historique permet également à une version précédente d'un contenu d'être ouverte et enregistrée à nouveau, afin de restaurer le contenu original (*roll-back*). Ceci est utile contre les utilisateurs hostiles qui veulent détruire une page. L'historique indique aussi le nom ou l'adresse IP de l'auteur, ce qui permet notamment d'interpeller directement un contributeur.

MediaWiki (MW) offre des fonctionnalités qui permettent de supporter facilement les 4 caractéristiques définies précédemment pour une bonne co-élaboration des CKBC. En effet comme tout Wiki, MW donne la possibilité à tout utilisateur de modifier le contenu d'une page. Ainsi, les aspects de construction et de complémentarité et de communauté ouverte deviennent plus faciles car chaque utilisateur peut apporter sa couche dans la co-élaboration de connaissances. MW propose également un cadre (espace de discussion) où les utilisateurs peuvent échanger pour l'élaboration de la connaissance. Ce cadre facilite l'amélioration et l'évolution de la connaissance. En effet, les utilisateurs peuvent discuter des orientations à donner à la thématique ce qui permet d'améliorer le contenu de la co-élaboration. MW permet également aux utilisateurs de référencer leurs contributions en donnant des sources par rapport aux contenus qu'ils mettent. Cela permet d'évaluer le contenu de l'élaboration de connaissances en se basant sur les sources des contributeurs. Grâce à sa fonctionnalité de traçabilité, MW permet d'interpeller directement un contributeur par rapport à ses contributions ce qui facilite l'évaluation car on pourra lui demander de donner ses sources pour un tel contenu.

Ainsi, MW est un bon outil pour supporter une CKBC grâce aux différentes fonctionnalités qu'il propose. Notons également que cette co-élaboration de connaissances n'est pas exclusivement faite en ligne. Les membres peuvent aussi initier des rencontres face à face dont les résultats peuvent être partagés grâce à la plateforme.

5.3.3 Animation des communautés utilisatrices

Comme montré par l'étude de Ridings et Gefen [55], la recherche d'information est la première motivation pour rejoindre une communauté virtuelle. L'objectif principal est de créer un cadre qui permette à nos communautés utilisatrices de partager et de co-construire leurs connaissances socioculturelles et ainsi de permettre aux personnes intéressées par la thématique de la communauté de disposer d'informations sur ce domaine. La thématique doit avoir un impact socioculturel sur la vie des citoyens. Il existe un noyau (une personne ou un groupe de personnes) qui identifie un besoin ou une appropriation d'une ressource (connaître l'histoire) dans sa localité et qui décide d'utiliser la connaissance de la communauté (au sens de la population de la localité ou toute personne qui pourrait être intéressée) pour créer une connaissance collective sur ce domaine (thématique). Par exemple, il existe une la perte de connaissances des nouvelles générations du déroulement de la cérémonie du *Ndut*³¹ [92]. Ainsi, la plateforme permet à ce noyau de créer une page dans ce domaine et les utilisateurs pourront utiliser leurs connaissances pour créer ce corpus.

Au sein des communautés utilisatrices, le soutien social est lié au sentiment d'appartenance et d'affiliation à une communauté - appropriation de leur centre d'intérêt - mais également à la façon dont cette communauté répond au manque de connaissances socioculturelles de ses membres. Ainsi en ayant un espace communautaire avec un soutien social élevé, il est plus facile pour les membres de prendre part aux discours intellectuels d'où émergera une connaissance collective.

Ainsi, au sein de nos communauté utilisatrices, nous distinguons quatre types d'utilisateurs :

1. Noyau : le noyau représente le groupe de personnes (ou une personne) qui prend la conscience du manque de connaissances socioculturelles dans un domaine et qui prend l'initiative de créer un espace de travail au sein de la plateforme pour collaborer dans ce domaine. Le groupe doit aussi veiller à ce que les discussions restent dans le champ défini. Le mot noyau est préféré à celui de modérateur, qui représente des utilisateurs qui ont un certain droit sur la communauté. Cependant, le noyau doit jouer le rôle de modérateur concernant les directions des échanges qui doivent tourner autour de la thématique mais le noyau n'a pas d'autorité sur les utilisateurs, il a les mêmes droits que les autres membres de la communauté.
2. Membre actif : c'est un utilisateur actif au sein de la communauté, il contribue régulièrement aux interactions. Il doit être tenu au courant des évolutions au sein de la communauté.
3. Rôdeur : c'est un utilisateur qui intervient de temps à autre.

³¹le mot *Ndut* veut dire étymologiquement nid; le lieu où la mère oiseau initie ses oisillons au vol. De même, c'est dans le *ndut* que la communauté Sérèr apprend la vie au néophyte et il est souvent matérialisé par la case ou hutte où se passent les initiations. Le *Ndut* marque le passage de l'enfance à l'âge adulte chez les Sérèr.

4. Visiteur : un utilisateur non enregistré au sein de la plateforme.

Pour pouvoir participer au sein d'une communauté, il faut d'abord s'y inscrire et toutes les contributions doivent être liées à la thématique (problématique) définie par le noyau. Donc il revient à la communauté tout entière de veiller à ce que toutes les interactions soient dans cette thématique. Contrairement à un modérateur dans les réseaux sociaux, le noyau n'a pas de privilège particulier par rapport au processus de construction de connaissances car il dispose des mêmes droits d'écriture qu'un utilisateur rôdeur. Notons aussi que les membres peuvent interagir avec d'autres pour construire en collaboration la connaissance.

5.4 Synthèse

Dans ce chapitre, nous avons proposé une méthodologie pour la construction manuelle d'une ontologie socioculturelle de haut niveau. Nous avons utilisé l'approche socioculturelle de la théorie historico-culturelle de l'Activité (CHAT) comme méta-modèle. À travers un système de médiation entre l'individu et son milieu, cette théorie analyse l'activité humaine dans un processus d'apprentissage. Cette théorie, centrée sur l'individu, est étendue à la notion de communauté qui a permis de considérer une communauté comme une entité atomique et de s'intéresser cette fois-ci au partage de connaissances entre communautés. Nous avons identifié les trois concepts généraux pour le partage et la co-construction de la connaissance socioculturelle.

La deuxième étape du processus de construction de l'ontologie de haut-niveau USCO (*Upper-level SocioCultural Ontology*) est l'identification des objets de médiation possibles entre une communauté et son milieu dans un processus d'apprentissage et de co-construction d'un patrimoine culturel. Ces objets de médiation sont principalement les événements socioculturels organisés par ces communautés mais également les faits historiques relatifs à ces communautés ou à leur environnement (fig. 5.1).

La dernière étape de ce processus de construction est l'utilisation du Linked Open Data (LOD) pour enrichir le vocabulaire de l'ontologie USCO grâce à l'alignement de schémas. Le LOD expose via le Web des données structurées et connectées entre elles. Nous avons choisi deux vocabulaires (fig. 5.3) : 1) Schema.org qui est le résultat d'une activité communautaire de collaboration avec la mission de créer, de maintenir et de promouvoir des schémas de données structurées sur l'Internet et 2) DBpedia qui est un effort de la communauté universitaire pour extraire des informations structurées à partir de Wikipedia et pour rendre cette information disponible sur le Web. Ces deux alignements nous ont permis de bénéficier des descriptions faites sur ces deux schémas mais également de bénéficier, lors du développement d'une ontologie de domaine, de la base de connaissances de DBpedia qui est universelle.

Chapter 5. Capitalisation de Connaissances : Ontologie Socioculturelle 90 et Communauté de Co-élaboration de Connaissances Culturelles

L'extension de l'approche socioculturelle de la CHAT à une communauté fait apparaître un type de communauté virtuelle. Ces communautés virtuelles appartiennent à la famille des KBC [50, 52]. Une KBC est un lieu où les individus sont engagés dans un processus de création, en discutant et en testant les connaissances et où leur «principal travail» est la production collaborative et l'amélioration de la connaissance et des idées [50]. Cependant, cette théorie est appliquée dans le domaine de l'éducation [51, 52] que nous avons appelé communauté de co-élaboration de connaissances éducationnelles (Educational Knowledge-building Community - EKBC). Dans ce chapitre, nous avons présenté l'utilisation des KBC dans le domaine culturel appelé communauté de co-élaboration de connaissances culturelles (Cultural Knowledge-building Community - CKBC). Contrairement à une EKBC, la construction de connaissances au sein d'une CKBC est générationnelle et l'impact ne se limite pas à une entité mais passe de génération à génération. De plus, au sein d'une CKBC le rôle d'un expert n'est pas obligatoire contrairement dans une EKBC où il y a la présence d'un(e) enseignant(e). Le domaine culturel étant plus complexe à circonscrire qu'une activité d'enseignement, l'ouverture et la participation de toute personne capable de contribuer à la construction de la connaissance doit être possible et même nécessaire.

Ce chapitre décrit les deux premières contributions importantes de cette thèse. En premier, nous avons montré comment nous avons construit une ontologie socioculturelle de haut niveau à partir de l'extension d'une théorie socioculturelle de l'apprentissage à la notion de communauté. Dans le processus de construction que nous avons proposé nous avons utilisé le LOD pour décrire plus précisément les objets et artefacts d'une communauté socioculturelle. L'ontologie résultante a pour vocation de permettre à des communautés de vulgariser et de pérenniser leur culture à travers un partage des coutumes et l'histoire des différentes localités de leur environnement. La deuxième contribution décrite dans ce chapitre est l'adaptation de la notion KBC dans le domaine culturel (Cultural Knowledge-building Community - CKBC) en proposant 4 principes de co-élaboration de connaissances socioculturelles. Nous avons également choisi un environnement technologique qui permet de supporter le travail collaboratif des CKBC.

HuTO : Human Time Ontology

(une ontologie temporelle narrative)

Sommaire

6.1	Introduction	91
6.2	Modélisation d'expressions temporelles	93
6.2.1	Généralités	93
6.2.2	Date, temps calendaire et unités temporelles	94
6.2.3	Instant, intervalle et durée	97
6.2.4	HuTO et les relations d'Allen	100
6.3	Annotation temporelle des données	101
6.4	Règles temporelles	106
6.4.1	Normalisation de la représentation temporelle	106
6.4.2	Implications et inférences	110
6.5	Requêtes-types	112
6.5.1	Requête-type sur les ressources	112
6.5.2	Requêtes-types sur les éléments temporels	113
6.6	Transformations	116
6.6.1	Transformation RDF2Text	117
6.7	Synthèse	121
6.7.1	Comparaison de HuTO avec les approches de l'état de l'art	123

6.1 Introduction

Nous avons consacré la première partie de cette thèse à un état de l'art sur, d'une part, la capitalisation de connaissances et, d'autre part, les ontologies temporelles dans le Web sémantique. Dans ce domaine temporel, nous avons identifié deux axes de travail : *la modélisation d'expressions temporelles* et *l'annotation temporelle des données*. La modélisation d'expressions temporelles permet de modéliser une date, un intervalle, des notions temporelles répétitives (*chaque mercredi John joue*

au foot), relatives ou absolues, etc. Ainsi, dans cet état de l'art (section 3.3) nous avons présenté plusieurs approches dont la principale lacune est la faiblesse de la représentation du temps humain. Ce dernier, lié à la culture [13], permet d'exprimer principalement des notions de temps déictiques, qualitatives et d'utiliser une expression comme une référence temporelle.

L'annotation temporelle des données permet la représentation de notions temporelles de façon à annoter des connaissances (exprimées sous forme de triplets en RDF) et cela en conservant l'évolution des données (changement de valeur) dans le temps. Dans ce domaine aussi, nous avons présenté plusieurs approches (section 3.4) dont les principaux inconvénients sont la perte de la sémantique des triplets (perte de l'assertion entre les triplets) et l'extension des standards du W3C.

Dans ce chapitre, nous proposons une ontologie temporelle narrative, Human Time Ontology¹ (HuTO). La modélisation de HuTO a été faite suivant les deux axes de travail identifiés précédemment. Dans le domaine de la modélisation d'expressions temporelles, l'objectif principal de HuTO est de pouvoir représenter le temps humain ce qui est important dans la représentation de données socioculturelles. Dans le domaine de l'annotation temporelle des données, l'objectif principal de HuTO est de conserver la sémantique des triplets tout en utilisant les standards du W3C.

Ainsi, HuTO est une ontologie formalisée en RDFS permettant l'annotation temporelle de ressources en RDF à l'aide d'expressions temporelles du langage courant. Cette ontologie permet également de définir des concepts comme des références temporelles liées au contexte et de capturer les changements temporels associés aux ressources annotées. Elle rend possible l'interrogation temporelle de la base de connaissances à l'aide de requêtes SPARQL. Plus précisément, HuTO permet de:

- Modéliser des expressions temporelles:
 1. **Explicites** : elles sont immédiatement ancrées ; par exemple : *30 Août 2014*.
 2. **Déictiques** : elles forment une relation spécifique avec le temps du discours ; par exemple : *aujourd'hui, demain*.
 3. **De durée** : elles indiquent un intervalle de temps ; par exemple : *2 heures, 20 minutes*.
 4. **Cycliques** : elles permettent de modéliser des dates répétitives ; par exemple : *chaque lundi, tous les deux mois*.
 5. **Qualitatives** : elles permettent de définir des notions temporelles relatives à une culture ou à une zone géographique qui peuvent être utilisées comme des références temporelles ; par exemple : *été 2014*.
 6. **Mixtes** : elles combinent les expressions pré-citées ; par exemple *deux mois l'année dernière*.

¹<http://ns.inria.fr/huto/>

- Normaliser les expressions temporelles afin de pouvoir leur appliquer des raisonnements et de pouvoir les interroger.

Ce chapitre continue comme suit. Dans la section 6.2, après avoir défini nos deux modèles, nous présentons les concepts de HuTO qui servent à modéliser les expressions temporelles. Dans la section 6.3, nous présentons notre approche pour l'annotation temporelle des données. La section 6.4 présente les règles temporelles de HuTO qui permettent de normaliser les intervalles définis précédemment et de raisonner sur les données. La section 6.5 présente les requêtes types qui permettent d'interroger une base de connaissances en utilisant HuTO. La section 6.6 présente nos transformations pour rendre les résultats des requêtes plus lisibles et plus compréhensibles à l'être humain. Enfin, nous terminons ce chapitre par une synthèse où une comparaison est faite entre HuTO et les approches présentées dans le chapitre 3.

6.2 Modélisation d'expressions temporelles

6.2.1 Généralités

Avant que nous puissions caractériser les expressions temporelles, nous avons besoin de préciser à quelle logique temporelle nous nous référons. En effet, Hayes a identifié six logiques temporelles [93]. La première et la plus fondamentale est la dimension temporelle (*time-dimension*) qui se réfère aux dimensions physiques de base. Cette dernière est un bloc contigu de l'espace-temps sur lequel le raisonnement peut être mis en œuvre. La seconde est le *time-plenum* qui est un continuum temporel c'est-à-dire un grand «espace» temporel au sein duquel tous les événements sont situés. La troisième logique considère les intervalles de temps (*time-intervals*) comme des parties du *time-plenum*. La quatrième logique définit les instants (*timepoint*) comme des intervalles infinitésimaux qui font partie d'un intervalle de temps (*time-intervals*). La cinquième logique définit les durées (*durations*) comme des quantités de temps et en particulier des quantités de temps des intervalles (*time-intervals*). La dernière logique identifiée par Hayes, *time-position*, permet de préciser une position dans un système de coordonnées temporelles comme les dates calendaires. Ces dates constituent généralement la réponse appropriée à une question concernant le déroulement d'une activité par exemple [93].

Dans cette thèse, nous utilisons deux logiques temporelles : les intervalles de temps (*time-intervals* [93]) et les positions dans un système de coordonnées temporelles (*time-position* [93]). La plupart des approches présentées dans l'état de l'art utilisent à la fois les intervalles et les instants. Cependant, nous avons fait le choix d'utiliser les intervalles et non les instants. En effet, comme dans les approches DOLCE [67] et Time Interval Pattern², nous considérons que les intervalles peuvent être considérés comme liés à des instants. Les instants peuvent être vus comme des intervalles aussi courts que possible. Ainsi, un instant peut être défini comme étant

²<http://ontologydesignpatterns.org/wiki/Submissions:TimeInterval>

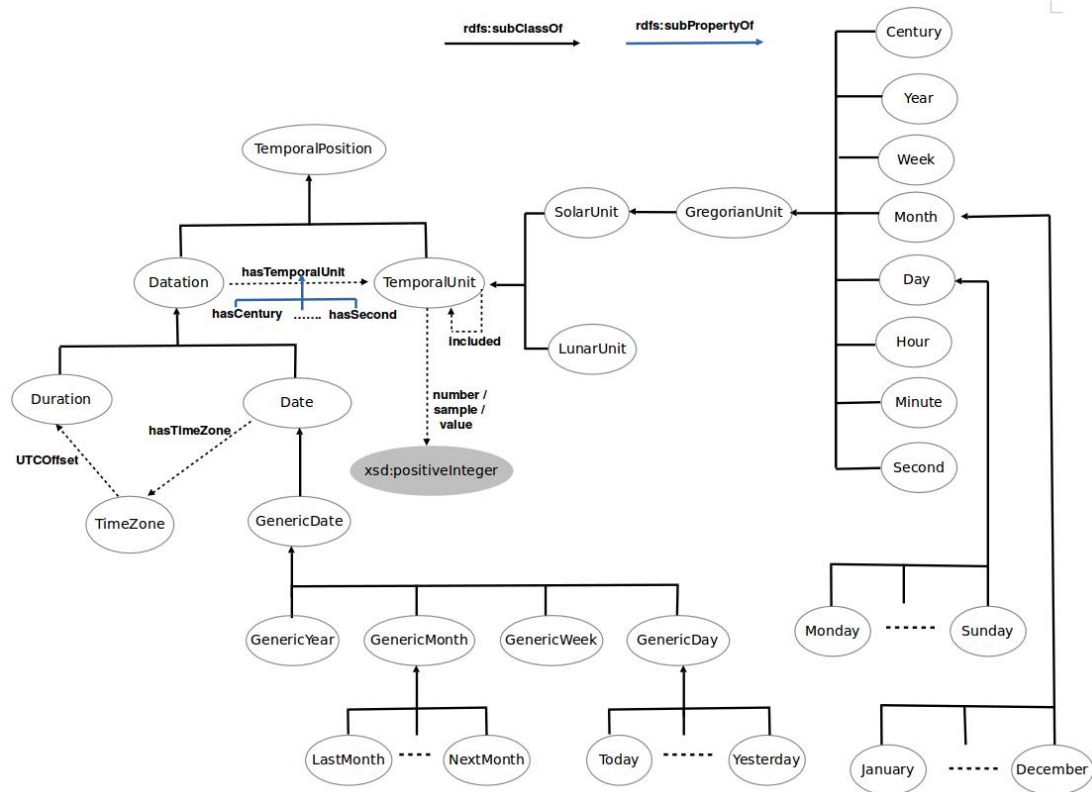


Figure 6.1: Principaux sous concepts et propriétés du concept *TemporalPosition*.

un moment pendant lequel rien ne peut arriver, qui ne possède aucune structure interne. Cela signifie qu'il n'y a pas de réel besoin pour modéliser les instants, étant donné que même les secondes, qui sont les unités de base du système international de temps, ont des structures internes. Dans notre modélisation, nous distinguons deux types d'intervalles : 1) les intervalles convexes qui sont des intervalles continus et 2) les intervalles non convexes qui sont des intervalles répétitifs.

Notons que dans notre modélisation, nous ne considérons que les intervalles et les *time-position*. Dans ce cas, un intervalle peut être représenté par une date, par des dates (date de début et date de fin), par une durée ou par une combinaison d'une date et d'une durée. Par conséquent, pour représenter un intervalle afin de mieux répondre à des questions temporelles, nous allons aussi modéliser les durées et les dates calendaires qui sont des éléments caractérisant les intervalles.

6.2.2 Date, temps calendaire et unités temporelles

Dans HuTO, le concept RDFS principal pour modéliser une position dans un système de coordonnées temporelles est *TemporalPosition* dont dérivent les concepts *Datation* et *TemporalUnit* (Fig. 6.1). Ces trois concepts sont des concepts

abstrait³. Le concept *Datation* a comme sous concepts *Date* et *Duration*. Le concept *Date* permet de modéliser des dates comme celles du type *xsd:dateTime*⁴ (exemples 6.1a et 6.1b). Il représente une date et une heure spécifiques. Un fuseau horaire peut être ajouté grâce au concept *TimeZone* en donnant la différence avec le temps universel (*Coordinated Universal Time*) grâce à la propriété *positive* (un *boolean*) qui est à *true* si la différence est positive par rapport au temps universel et à *false* si la valeur est négative par rapport au le temps universel. Si aucune valeur de fuseau horaire n'est donnée, il est supposé que l'heure est en temps universel (UTC). Le concept *Duration* permet de définir des durées comme celles du type *xsd:duration*⁵ (exemple 6.1d). Il représente une durée exprimée en nombre d'années, mois, jours, heures, minutes et secondes (exemple 6.1c).

HuTO permet de modéliser des dates déictiques qui sont des dates qui forment une relation spécifique avec le temps du discours. Dans notre modélisation, les temps déictiques sont des sous concepts du concept *Date* (Fig. 6.1). En effet, on ne peut pas savoir à quoi réfère une date déictique sans connaître la position temporelle du discours. Le choix de modélisation fait, comme dans l'exemple 6.1d, est de modéliser le temps déictique en y associant les propriétés de la date effective associée au temps déictique.

Notre ontologie permet de prendre en considération différentes unités temporelles grâce au concept *TemporalUnit*. Actuellement, nous ne considérons que les unités solaires (*SolarUnit*) et les unités lunaires (*LunarUnit*). Cependant, HuTO a été conçue de façon à permettre de modéliser différentes unités calendaires. Ainsi, il est possible de définir d'autres unités temporelles comme celles des calendriers Maya ou Aztèque qui sont des spécialisations du concept abstrait *TemporalUnit*. Comme nous nous intéressons plus particulièrement à la modélisation des événements socio-culturels sénégalais, dans la figure 6.1, nous avons détaillé les unités de temps du calendrier grégorien pour lequel les granularités définies vont de *Century* à *Second*. En effet plusieurs calendriers sont des calendriers luni-solaires c'est-à-dire une combinaison du cycle annuel du soleil (*SolarUnit*) et sur le cycle régulier des phases de la Lune (*LunarUnit*). Nous adoptons cette approche car plusieurs événements socioculturels sénégalais utilisent les unités solaires et lunaires.

Pour hiérarchiser les unités temporelles, nous avons défini la propriété *included* (Fig. 6.1) qui est utilisée pour ordonner les unités temporelles. Par exemple, pour les unités grégoriennes, nous avons *included(Year, Century)* pour signifier que l'unité *Year* est incluse dans l'unité *Century*. Notons également que la propriété *hasTemporalUnit* (Fig. 6.1) est spécialisée par les propriétés *hasCentury*, *hasYear*, etc. qui permettent de spécifier respectivement le siècle, l'année, etc. d'une date ou d'une durée.

³Un concept abstrait n'a pas d'instances directes.

⁴http://www.datypic.com/sc/xsd/t-xsd_dateTime.html

⁵http://www.datypic.com/sc/xsd/t-xsd_duration.html

Dans l'exemple 6.1a comme avec *xsd:dateTime*, pour spécifier une date on donne la valeur des différentes unités de la date. Nous proposons deux types de notations : 1) la notation calendaire (exemple 6.1a) et la notation numérique (exemple 6.1b). La propriété *value* est utilisée pour spécifier la valeur en jour ou en année de la date. Comme spécifié précédemment, une date déictique est liée directement à la date réelle grâce à ces propriétés. Comme pour les dates, une durée est représentée en spécifiant les valeurs de ses unités (exemple 6.1c). La propriété *number* est utilisée pour spécifier la magnitude de la durée.

a. Date(Tuesday, February 17, 2015 at 10 pm⁶)

```
[a :Date;  
  :hasHour [a :Hour;  
            :value 22];  
  :hasDay [a :Tuesday;  
          :value 17];  
  :hasMonth [a :February];  
  :hasYear [a :Year;  
           :value 2015]].
```

b. Date(04 15) "April 15"

```
[a :Date;  
  :hasDay [a :Day;  
          :value 15];  
  :hasMonth [a :Month;  
            :value 4]].
```

c. 2 hours 30 minutes

```
[a :Duration;  
  :hasHour [a :Hour;  
            :number 2];  
  :hasMinute [a :Minute;  
             :number 30]].
```

d. Date(Today)- Friday April 25,
2014

```
[a :Today;  
  :hasDay [a :Friday;  
          :value 25];  
  :hasMonth [a :April];  
  :hasYear [a :Year;  
           :value 2014]].
```

Exemple 6.1: Modélisation de date et de durée.

La modélisation de la date que nous avons proposée permet de représenter des dates du même type que *xsd:dateTime* et *xsd:duration* qui permettent de représenter des

⁶Nous utilisons le format 24 heures pour représenter les heures.

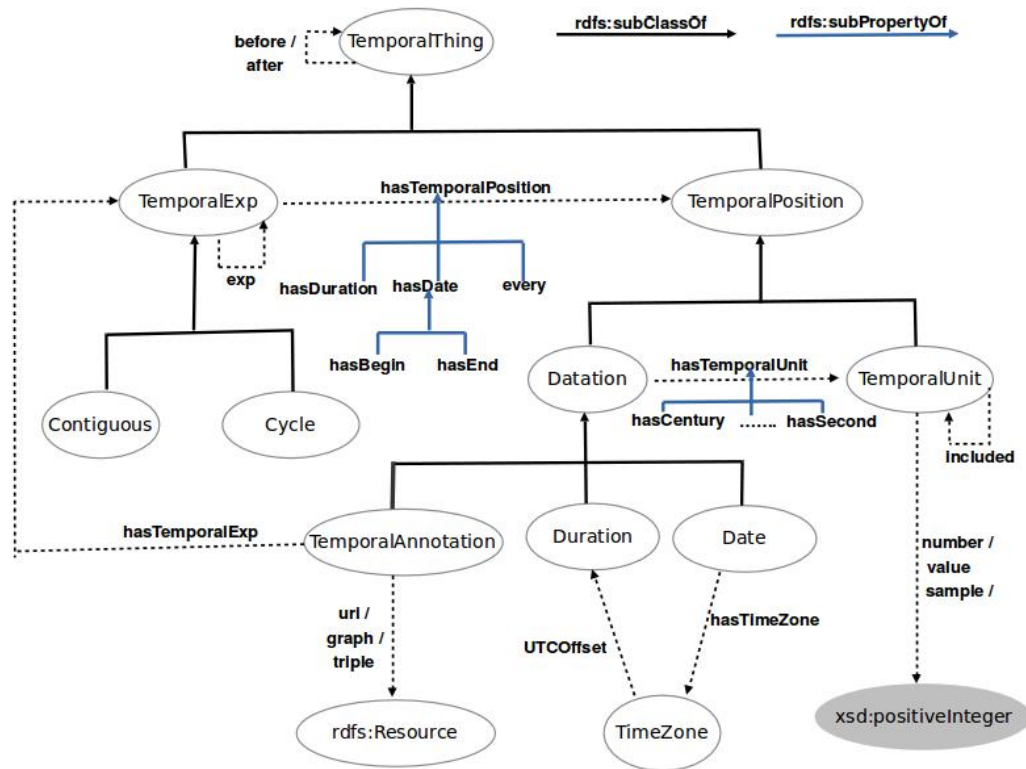


Figure 6.2: Les intervalles et leurs propriétés dans HuTO.

dates grégoriennes. De plus, cette modélisation permet de représenter des dates déictiques et prend en charge d'autres unités temporelles comme les unités lunaires.

6.2.3 Instant, intervalle et durée

La deuxième partie du modèle de HuTO permet de décrire des intervalles. En effet, HuTO permet de représenter un instant et une durée dans le sens de Hayes [93] comme un intervalle (*Contiguous* et *Cycle* dans la figure 6.2). Pour un intervalle, si la date de fin ou la durée n'est pas spécifiée alors l'intervalle considéré est celui de l'unité tout entière. Par exemple, la date *Vendredi 29 Janvier 2016* est considérée comme un intervalle de *24 heures*. Une durée représente la magnitude d'un intervalle. Ainsi une durée est modélisée par un intervalle avec sa magnitude.

Dans notre modélisation, nous considérons deux types d'intervalles (Fig. 6.2) : 1) les intervalles convexes, c'est-à-dire continus ou contigus, modélisés grâce au concept *Contiguous* et 2) les intervalles non convexes, c'est-à-dire discontinus ou cycliques, modélisés grâce au concept *Cycle*.

Ainsi, le concept *Contiguous* est utilisé pour représenter :

- un instant dans le sens de Hayes [93] avec la propriété *hasDate* (exemple 6.2a).
- une durée dans le sens de Hayes [93] avec la propriété *hasDuration* (exemple 6.2d).
- un intervalle dans le sens de Hayes [93] avec les propriétés :
 1. *hasBegin* et *hasEnd* (exemple 6.2c).
 2. *hasBegin* et *hasDuration* (exemple 6.2b).
 3. *hasEnd* et *hasDuration*.

Notons que dans la modélisation proposée, nous faisons la distinction entre les intervalles infinis et les intervalles fermés :

- Un intervalle est fermé si une des trois conditions est réalisée :
 1. la propriété *hasDate* est utilisée ;
 2. les propriétés *hasBegin* et *hasEnd* sont utilisées toutes les deux ;
 3. Une des propriétés *hasBegin* ou *hasEnd* est utilisée avec la propriété *hasDuration* ;
- Sinon l'intervalle est infini.

a. Interval of Tuesday, February 17, 2015

```
[a :Contiguous;  
 :hasDate  
  [a :Date;  
   :hasDay [a :Tuesday;  
           :value 17];  
   :hasMonth [a :February];  
   :hasYear [a :Year;  
            :value 2015]]].
```

b. From Tuesday 17 February 2015 to two months

```
[a :Contiguous;  
 :hasBegin  
  [a :Date;  
   :hasDay [a :Tuesday;  
           :value 17];  
   :hasMonth [a :February];  
   :hasYear [a :Year;  
            :value 2015]].  
:hasDuration  
 [a :Duration;  
  :hasMonth [a :Month;  
            :number 2]]].
```

c. From Tuesday 17 to Friday 20 February 2015

```
[a :Contiguous;
  :hasBegin
    [a :Date;
      :hasDay [a :Tuesday;
        :value 17];
      :hasMonth [a :February];
      :hasYear [a :Year;
        :value 2015]].
  :hasEnd
    [a :Date;
      :hasDay [a :Friday;
        :value 20];
      :hasMonth [a :February];
      :hasYear [a :Year;
        :value 2015]]].
```

d. Two days

```
[a :Contiguous;
  :hasDuration
    [a :Duration;
      :hasDay [a :Day;
        :number 2]]].
```

Exemple 6.2: Modélisation d'intervalles convexes.

Le concept *Cycle* est utilisé pour modéliser les intervalles non convexes (répétitifs). Les intervalles non convexes sont caractérisés par deux éléments : 1) la fréquence de répétition et 2) l'occurrence de l'intervalle convexe à répéter. Cette fréquence est un sous concept du concept *TemporalUnit* qui représente l'unité de temps avec laquelle la répétition est exprimée. L'intervalle convexe est relié au concept *Cycle* par la relation *exp*. De façon générale, la propriété *exp* (Fig. 6.2) est utilisée pour connecter un intervalle non convexe (*Cycle*) à un intervalle convexe (*Contiguous*) (exemple 6.3a) ou à un autre intervalle non convexe. Elle est également utilisée pour connecter un intervalle convexe à un intervalle non convexe (exemple 6.3b).

L'exemple 6.3a représente un intervalle non convexe. La fréquence de cet intervalle répétitif est annuelle, aussi l'unité de la propriété *every* est *Year*. L'intervalle répétitif est un *Sunday*, le premier (*number 1*). La propriété *number* est utilisée pour modéliser une magnitude comme dans une durée. L'exemple 6.3b représente un intervalle non convexe délimité dans le temps. Par conséquent, l'intervalle non convexe est contenu dans un intervalle convexe. Ainsi, l'intervalle convexe a deux parties : la première pour modéliser la durée et la seconde (propriété *exp*) pour modéliser l'intervalle non convexe. La propriété *sample* est utilisée pour modéliser

un échantillon d'une date comme dans l'expression *every 8 hours*.

a. The first Sunday of every April.

```
[a :Cycle;
  :every [a :Year];
  :exp [a :Contiguous;
    :hasDate
      [a :Date;
        :hasDay
          [a :Sunday;
            :number 1];
        :hasMonth
          [a :April]]]]].
```

b. Every 8H for 10 days starting from today⁷

```
[a :Contiguous;
  :hasBegin [a :Today];
  :hasDuration
    [a :Duration;
      :hasDay [a :Day;
        :number 10]];
  :exp [a :Cycle;
    :every [a :Hour;
      :sample 8]]].
```

Exemple 6.3: Modélisation d'intervalles non convexes.

Notons que la propriété *every* est utilisée entre les concepts *Cycle* et *TemporalPosition* (Fig. 6.2). De même, les propriétés *hasDate* et *hasDuration* sont utilisées entre les concepts *Contiguous* et *TemporalPosition* (Fig. 6.2).

6.2.4 HuTO et les relations d'Allen

Allen a défini treize relations fondamentales entre les intervalles (Fig. 3.1) et fournit une table de composition qui peut être utilisée comme une base de raisonnement sur des entités temporelles [57, 58]. Dans la définition d'Allen, un intervalle est convexe, fermé et ordonné.

Dans la version actuelle de HuTO, nous n'avons modélisé que les propriétés *before* et *after* (Fig. 6.2). Ces deux propriétés nous permettent de modéliser la représentation du temps en datation relative, c'est-à-dire que la dimension temporelle d'une ressource est exprimée par rapport à la dimension temporelle d'une autre ressource (exemple 6.4b). Notons que cette utilisation nous permet d'avoir deux informations implicites (cf. section 6.4.2): 1) la date de la ressource référencée et 2) les deux relations d'Allen entre les ressources. HuTO permet de spécifier les relations *before*

⁷Le contexte du *Today* a été omis volontairement pour ne pas surcharger l'exemple.

et *after* entre deux intervalles, entre deux ressources et entre une ressource et un intervalle.

6.3 Annotation temporelle des données

L'annotation temporelle des données consiste à lier une donnée (une ressource, un triplet ou un graphe nommé) à sa dimension temporelle. Grâce à notre modélisation deux dimensions peuvent co-exister : 1) une temporelle et 2) une non temporelle. La dimension temporelle est spécifiée à l'aide des concepts de HuTO et la dimension non temporelle, celle du domaine de connaissances, est spécifiée au travers des triplets décrivant des aspects autres que ceux temporels. Ainsi, la recherche d'information est facilitée en considérant ou non la partie temporelle.

De façon générale, l'annotation temporelle peut être associée à une ressource, un triplet ou un graphe nommé. Pour annoter ces objets, une ressource *TemporalAnnotation* est créée et est associée à quatre propriétés (Fig. 6.2). La première, la propriété *hasTemporalExp*, permet de connecter l'expression temporelle qui anote la ressource. Les trois autres propriétés sont associées à la ressource à annoter temporellement comme suit :

1. Si c'est une ressource, la propriété *uri* est utilisée pour la lier avec l'URI de la ressource (exemple 6.4d).
2. Si c'est un triplet, la propriété *triple* est utilisée avec une réification RDF (Exemple 6.4a).
3. Si c'est un graphe nommé, la propriété *graph* est utilisée pour la lier avec l'URI du graphe nommé (exemple 6.4c).

L'exemple 6.4a représente l'annotation d'un triplet réifié. De ce fait, un objet *TemporalAnnotation* est créé qui utilise la propriété *hasTemporalExp* pour modéliser l'intervalle d'occurrence et la propriété *triple* pour la réification RDF. L'exemple 6.4b est un exemple d'une des utilisations de la propriété *after* au sens d'Allen. La propriété est définie entre deux objets *TemporalAnnotation* et la valeur de la propriété *uri* de chaque objet *TemporalAnnotation* est liée à une ressource. L'exemple 6.4c représente l'annotation d'un ensemble de triplets (graphe nommé). Ainsi, un objet *TemporalAnnotation* est créé et annoté temporellement (la propriété *hasTemporalExp*). La propriété *graph* est utilisée pour faire le lien avec l'URI du graphe. Ce dernier réunit en son sein des triplets dont l'union partage le même intervalle de temps. L'exemple 6.4d représente l'annotation d'une ressource. Comme pour les autres annotations, un objet *TemporalAnnotation* est créé qui est temporellement annoté grâce à la propriété *hasTemporalExp* et la propriété *uri* est utilisée pour faire le lien avec la ressource à annoter temporellement. L'exemple 6.4f représente l'annotation d'une expression complexe contenant un intervalle non convexe qui contient un intervalle convexe qui contient lui même un intervalle non convexe.

a. Senghor was the President of Senegal from September 1960 to December 1980

```
[a :TemporalAnnotation;  
  :hasTemporalExp  
    [a :Contiguous;  
      :hasBegin  
        [a :Date;  
          :hasMonth [a :September];  
          :hasYear  
            [a :Year;  
              :value 1960]]];  
      :hasEnd  
        [a :Date;  
          :hasMonth [a :December];  
          :hasYear  
            [a :Year;  
              :value 1980]]];  
  :triple  
    [ rdf:subject    <Senghor>;  
      rdf:predicate  <presidentOf>;  
      rdf:object     <Senegal>]].
```

b. The Battle of Dekheulé took place after the Battle of Mékhé

```
[a :TemporalAnnotation;  
  :uri <BattleOfDekheule>]  
:after  
  [a :TemporalAnnotation;  
    :uri <BattleOfMekhe>].
```

c. In 2011, the City of Dakar have 1,056,009 inhabitants, it is the most populated and its mayor is Mr Sall.

```
[a :TemporalAnnotation;  
  :hasTemporalExp  
    [a :Contiguous;  
      :hasDate [a :Date;  
        :hasYear  
          [a :Year;  
            :value 2011]]];  
  :graph <http://example.org/g/>].
```

```
<http://example.org/g/>{  
  <Dakar> <population> 1056009;  
    <rank> 1;  
    <mayor> <Sall>}.
```

d. The first Saturday of each December, the Fanal of Ndar⁸ is organized

```
[a :TemporalAnnotation;
  :hasTemporalExp
    [a :Cycle;
      :every [a :Year];
      :exp [a :Contiguous;
        :hasDate
          [a :Date;
            :hasDay
              [a :Saturday ;
                :number 1];
            :hasMonth
              [a :December]]]];
  :uri <FanalOfNdar>] .
```

f. Every year, during summer, every Saturday, go to the beach

```
[a :TemporalAnnotation ;
  :hasTemporalExp [a :Cycle ;
    :every [a :Year] ;
    :exp [a :Contiguous ;
      :hasDate [a :Summer];
      :exp [a :Cycle ;
        :every [a :Saturday]]]];
  :uri <GoToBeach>] .
```

Exemple 6.4: Annotation temporelle de données.

Notons que HuTO permet de modéliser les notions temporelles qualitatives (Fig. 6.3). En effet, HuTO permet à une ressource d'être utilisée comme une référence temporelle (une date) à l'aide du concept *TemporalAnnotation*. Dans ce cas, la ressource est considérée comme un objet temporel et peut être utilisée comme un ancrage dans le temps. Ainsi, au lieu de se référer à une date, on utilise la ressource annotée comme référence temporelle comme dans l'exemple 6.4b.

HuTO permet également de modéliser plusieurs notions temporelles par rapport à une culture ou à une position géographique. Par exemple, la notion *PartOfDay* est partagée à travers le monde, mais la durée dépend de la position géographique. Ainsi, la notion de nuit (*Night*) a la même signification à Dakar (Sénégal) et à Sydney (Australie), mais les intervalles ne sont pas les mêmes. À la date du 29 Juillet 2015, la notion de nuit est de 21 heures à 5 heures 30 à Dakar⁹ alors qu'à Sydney¹⁰ elle est de 18 heures 30 à 5 heures 30. Aussi pour modéliser cette différence, les concepts *DakarNight* et *SydneyNight* peuvent être créés en tant que sous concepts du concept

⁸Fanal est un carnaval et Ndar (Saint-Louis) est une ville au nord du Sénégal

⁹<http://www.timeanddate.com/astronomy/senegal/dakar>

¹⁰<http://www.timeanddate.com/astronomy/australia/sydney>

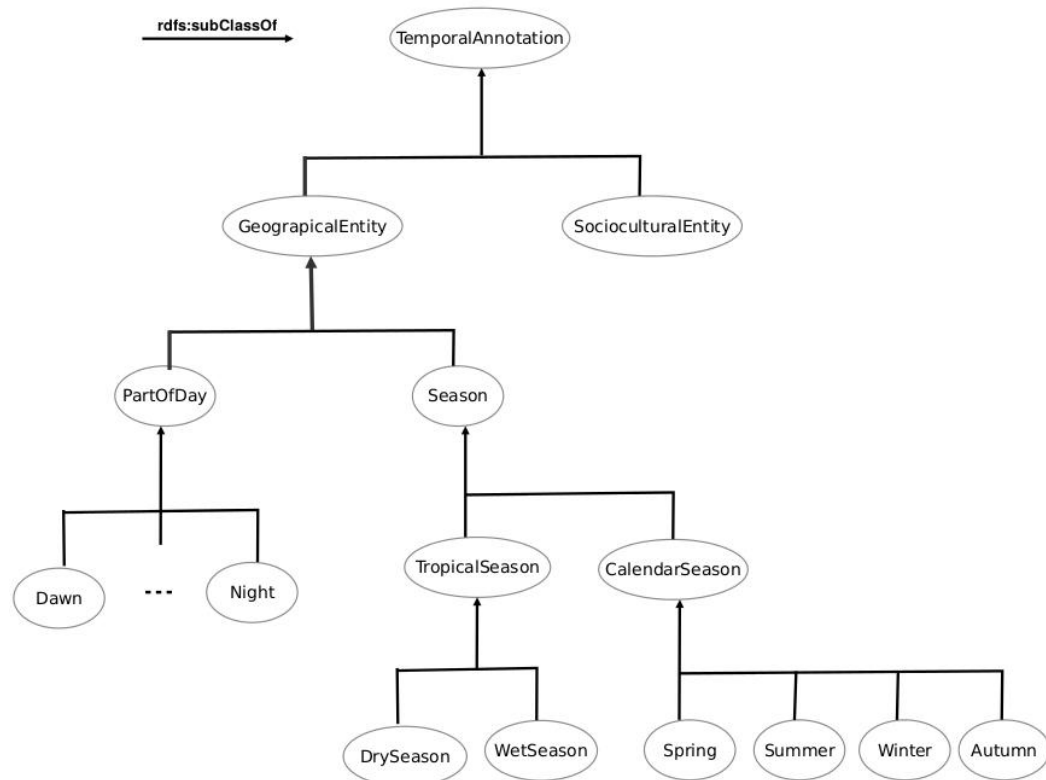


Figure 6.3: Modélisation temporelle qualitative dans HuTO.

Night et chacun d'eux aura son propre intervalle. Ainsi, pour modéliser des notions temporelles contextuelles, la modélisation qualitative de HuTO peut être utilisée en spécifiant le bon intervalle par rapport à la position géographique. Après cela, la notion peut être utilisée comme dans l'exemple 6.4b comme un ancrage dans le temps. La même démarche est effectuée pour la notion de saison (*Season*). Les pays Africains comme le Sénégal ont une saison tropicale et la durée et la période de cette saison sont relatives à la position géographique du pays.

HuTO permet aussi de définir des notions temporelles spécifiques à la culture d'une région comme des sous concepts du concept *SocioculturalEntity* qui est un sous concept de *TemporalAnnotation* (fig. 6.3). Par exemple au Sénégal, l'événement Gamou¹¹ (Mawlid¹²) est une date de référence importante où de nombreux événements religieux sont organisés. Pour faire référence à la date de ces événements, l'expression **durant le Gamou** est utilisée. Ainsi avec la modélisation HuTO, le concept **Gamou** peut être un sous concept de *SocioculturalEntity* et on peut l'utiliser comme un ancrage dans le temps pour dater ces événements (exemple

¹¹Période de 12 jours pour célébrer la naissance du prophète de l'Islam.

¹²*Gamou* est le terme utilisé au Sénégal et *Mawlid an-nabawiy* est la transcription arabe pour *naissance du prophète*.

6.5).

a. Religious conference is organized during the Gamou

```
[a :TemporalAnnotation;
 :hasTemporalExp
  [a :Contiguous;
   :hasDate
    [a :TemporalAnnotation;
     :uri <Gamou>]];
 :uri <ReligiousConference>].
```

b. Gamou starts on December 24, 2015 for 12 days

```
[a :TemporalAnnotation;
 :hasTemporalExp
  [a :Contiguous;
   :hasBegin
    [a :Date;
     :hasDay
      [a :Day;
       :value 24]
     :hasMonth [a :December];
     :hasYear
      [a :Year;
       :value 2015]];
   :hasDuration
    [a :Duration;
     :hasDay
      [a :Day;
       :number 12]]];
 :uri <Gamou>]].
```

Exemple 6.5: Modélisation qualitative dans HuTO.

L'utilisation de HuTO présente certains avantages comparée aux approches présentées dans le chapitre 3. Dans la modélisation des expressions temporelles, HuTO permet de représenter des énoncés complexes comme dans l'exemple 6.4f. HuTO intègre aussi la modélisation des intervalles fermés et infinis et permet aussi d'utiliser une ressource comme référence temporelle (ancrage dans le temps). Ces aspects ne sont pas considérés par les autres approches présentées. HuTO permet également de modéliser le temps déictique ce que ne font pas les autres approches excepté CN-TRO qui le modélise dans une chaîne de caractère. Un autre avantage majeur de HuTO est la modélisation de notions temporelles qualitatives qui permet de définir des notions temporelles spécifiques à une culture (exemple 6.5) ou à une position géographique.

Pour l'annotation temporelle des données, HuTO propose une représentation qui permet d'annoter une ressource, un triplet ou l'ensemble des triplets d'un graphe nommé. Ceci nous permet de séparer la partie temporelle des données de celles du domaine contrairement aux autres approches, excepté Temporal RDF, où la sémantique des triplets est perdue par l'introduction d'un *blank-node* (n-ary relations, la réification RDF) ou par le déplacement des relations sur des *timeSlice* (4D-Fluents). Hormis la modélisation des expressions temporelles que HuTO intègre, la principale différence entre HuTO et Temporal RDF est que ce dernier nécessite une extension légère du vocabulaire de RDF [79].

6.4 Règles temporelles

HuTO fournit un modèle conceptuel en RDFS pour modéliser des expressions temporelles et pour annoter des ressources en RDF. Cependant, de nombreuses relations temporelles sont exprimées implicitement dans les occurrences d'événements (datation relative). Les réponses à de nombreuses questions axées sur le temps ne sont pas nécessairement représentées explicitement mais peuvent être déduites. Pour cela, nous avons proposé un ensemble de règles permettant de normaliser la représentation des données temporelles et également des règles d'inférence.

6.4.1 Normalisation de la représentation temporelle

Puisque HuTO est une ontologie en RDFS, nous avons proposé des règles, exprimées sous forme de requêtes CONSTRUCT en SPARQL ayant pour objectif de déduire et d'explicitier le maximum d'information temporelle afin de permettre le raisonnement sur les données.

Les informations temporelles peuvent être exprimées de différentes façons. Par exemple, une date (le mois) peut être représentée soit en utilisant la représentation calendaire (Exemple 6.1a), soit à l'aide de chiffres (exemple 6.1b). Aussi, nous avons créé des règles pour normaliser ces deux types d'écritures. De ce fait, quel que soit le mode d'écriture utilisé, toutes les représentations possibles seront ajoutées dans le graphe des données (exemple 6.6).

```
a. PREFIX huto: <http://ns.inria.fr/huto/>
   CONSTRUCT{
     ?x huto:number ?m ;
       huto:numberOfDay ?d ;
       huto:even ?e}
   WHERE{
     ?x rdf:type ?o
     ?o rdfs:subClassOf huto:Month ;
       huto:number ?m ;
       huto:even ?e .
     OPTIONAL{?x rdf:type/huto:numberOfDay ?d}
```

```

}

b. PREFIX huto: <http://ns.inria.fr/huto/>
   CONSTRUCT { ?x a ?c }
   WHERE { ?x huto:number ?m ;
           rdf:type huto:Month .
           ?c huto:number ?m .
           ?c rdfs:subClassOf huto:Month
         }

```

Exemple 6.6: Exemple de règles de normalisation des mois.

L'exemple 6.6a montre pour un mois grégorien donné comment ajouter le nombre de jours du mois, la parité du mois et le numéro du mois. La règle de l'exemple 6.6b permet d'ajouter le mois grégorien à un mois exprimé en numérique.

Nous avons également proposé deux règles pour déterminer les années bissextiles. Ces règles nous permettent, entre autres, de connaître le nombre de jours dans l'année ce qui est utile pour répondre à certaines requêtes (exemple 6.7).

```

a. PREFIX huto: <http://ns.inria.fr/huto/>
   CONSTRUCT { ?x huto:leapYear true }
   WHERE {
     ?x a huto:Year ;
     huto:value ?y .
     BIND((?y / 4) as ?y1)
     BIND(ROUND(?y1) as ?y2)
     FILTER(?y1 = ?y2)
     BIND((?y / 100) as ?y3)
     BIND(ROUND(?y3) as ?y4)
     FILTER(?y3 != ?y4)
   }

b. CONSTRUCT { ?x huto:leapYear true }
   WHERE {
     ?x a huto:Year ;
     huto:value ?y .
     BIND((?y / 400) as ?y1)
     BIND(ROUND(?y1) as ?y2)
     FILTER(?y1 = ?y2)
   }

```

Exemple 6.7: Règles pour la détermination d'une année bissextile.

Les deux règles de l'exemple 6.7 permettent de déterminer si une année est bissextile. En effet, l'exemple 6.7a vérifie la règle

$\text{Year mod } 4 = 0 \text{ and } \text{Year mod } 100 \neq 0$ et l'exemple 6.7b vérifie la règle $\text{Year mod } 400 = 0$. Si l'une des règles est vérifiée alors l'année est bissextile.

La modélisation de HuTO permet à une ressource d'être utilisée comme une référence temporelle (exemple 6.5a). Ainsi pour pouvoir retrouver ces ressources lors de la recherche d'information temporelle, nous avons défini une règle qui ajoute explicitement l'intervalle d'occurrence de la ressource utilisée comme référence temporelle (exemple 6.8).

```
PREFIX huto: <http://ns.inria.fr/huto/>
CONSTRUCT { ?h ?m ?j }
WHERE {
    ?h ?o ?l .
    ?l a huto:TemporalAnnotation ;
        (huto:uri|huto:triple|huto:graph) ?y .
    ?x a huto:TemporalAnnotation ;
        (huto:uri|huto:triple|huto:graph) ?y ;
        huto:hasTemporalExp ?k .
    ?k ?m ?j .
    FILTER EXISTS{ ?m rdfs:subPropertyOf+ huto:hasTemporalPosition }
    FILTER(?x != ?l)
}
```

Exemple 6.8: Règle de normalisation d'une ressource utilisée comme référence temporelle.

Ainsi, la règle de l'exemple 6.8 appliquée à l'exemple 6.5 permet d'ajouter la date d'occurrence du *Gamou* comme date absolue de *ReligiousConference*. En effet, les trois premières lignes de la clause WHERE montrent que la ressource *?h* utilise la ressource *?y* comme référence temporelle. De ce fait, les quatre lignes qui suivent permettent d'extraire la date d'occurrence de la ressource *?y* qui est utilisée comme date d'occurrence de la ressource *?h* (clause CONSTRUCT).

Nous avons également normalisé les intervalles définis par leur durée en ajoutant explicitement la date de fin (*hasEnd*) ou la date de début (*hasBegin*) de l'intervalle. L'exemple 6.9 montre une règle pour expliciter la date de fin d'un intervalle défini par sa durée. Notons qu'il existe sept règles pour normaliser les intervalles définis par leur durée (car la règle dépend du type de la durée qui peut être en siècle, en année, en mois, en semaine, en jour, en minute ou en seconde).

Dans l'exemple 6.9 est présentée la règle qui ajoute la date de fin (*hasEnd*) à un intervalle exprimé par une durée et une date de début (*hasBegin*). Dans cette règle, le calcul de la date de fin se fait sur le mois et l'année car la durée est exprimée en mois. Ainsi, nous récupérons toutes les informations liées à la date de début (*?p ?j ?h*) que nous ajoutons à la date de fin, c'est-à-dire l'heure, le jour, la semaine s'ils existent dans la date de début. Les BIND de la clause WHERE permettent de calculer le mois et l'année de la date de fin.

Cette règle, appliquée à l'exemple 6.2b, ajoute la propriété *hasEnd* à l'intervalle avec les propriétés *hasDay*, *Month* et *hasYear*.

```

PREFIX huto: <http://ns.inria.fr/huto/>
CONSTRUCT{?x huto:hasEnd
  [ a huto:Date ;
    ?j ?h ;
    huto:hasMonth [a huto:Month ;
                   huto:number ?m3] ;
    huto:hasYear [a huto:Year ;
                  huto:value ?y1]]}
WHERE { ?x huto:hasBegin ?p ;
        huto:hasDuration/huto:hasMonth/huto:number ?m1 .
        ?p huto:hasMonth/huto:number ?m ;
        huto:hasYear/huto:value ?y ;
        ?j ?h .

        BIND(?m + ?m1 as ?m2)
        BIND(IF(?m2 > 12, ?m2 - 12, ?m2) as ?m3)
        BIND(IF(?m2 != ?m3, ?y+1, ?y) as ?y1)

        FILTER( ?j != huto:hasMonth && ?j != huto:hasYear)
        FILTER NOT EXISTS{ ?x huto:hasEnd ?l }
}

```

Exemple 6.9: Règle de normalisation d'une durée exprimée en mois¹³.

Pour vérifier la cohérence entre les concepts *Contiguous* et *Cycle*, nous avons également créé une requête de vérification et de correction. En général, si le concept de *Cycle* contient le concept *Contiguous* alors l'unité de la fréquence (*every*) doit être supérieure ou égale à la plus grande unité de l'intervalle *Contiguous*. Ainsi, dans l'exemple 6.4f, on ne peut pas inverser les positions des intervalles *Cycle* (le premier) et *Contiguous* parce que nous avons la relation *included(Summer, Year)*. De même, si le concept *Contiguous* contient un concept *Cycle* alors la plus petite unité de temps de l'intervalle *Contiguous* doit être supérieure à l'unité de la fréquence de l'intervalle *Cycle*. Ainsi, dans l'exemple 6.3b, on ne peut pas inverser les positions des intervalles *Contiguous* et *Cycle* parce que nous avons la relation *included(Hour, Day)* et *Day* est la plus petite unité de l'intervalle *Contiguous* et *Hour* est l'unité de la fréquence de l'intervalle *Cycle*.

Nous avons défini deux règles qui vérifient la cohérence des unités entre les intervalles *Contiguous* et *Cycle*. Le cas échéant, elles font les rectifications nécessaires. Il existe une règle lorsqu'un intervalle convexe (*Contiguous*) contient un intervalle

¹³Notons qu'une durée exprimée en mois n'excède pas 11 mois. Le cas échéant, elle est exprimée en année. D'où la valeur de ?m2 est toujours inférieur à 24.

non convexe (*Cycle*) (exemple 6.10) et une autre règle lorsqu'un intervalle non convexe (*Cycle*) contient un intervalle convexe (*Contiguous*). Notons que les deux règles vérifient et corrigent le cas échéant. Ces requêtes retournent l'ensemble des triplets qui ont été mis à jour. Notons aussi que tant que ces règles retournent des résultats il faut les ré-exécuter à nouveau.

La règle de l'exemple 6.10 extrait le type de l'intervalle *Contiguous* (?y) et celui de l'intervalle *Cycle* (?z). Ensuite, elle compare si leur ordre d'inclusion est inversé (*included*). Si tel est le cas, la requête interchange leur position grâce aux clause DELETE et INSERT.

```
PREFIX huto: <http://ns.inria.fr/huto/>
DELETE { ?p ?o ?k
         ?m ?f ?j }
INSERT { ?p ?f ?j
         ?m ?o ?k }
WHERE {
  ?x a huto:Contiguous ;
      huto:hasDate ?p ;
      huto:exp/huto:every ?m .

  ?p ?o ?k .
  ?k rdf:type ?y .

  ?m rdf:type ?z ;
      ?f ?j .

  ?y huto:included ?t .
  ?z rdfs:subClassOf* ?t .

  FILTER (?y != ?t)
}
```

Exemple 6.10: Règle de correction d'un intervalle *Contiguous* qui contient un intervalle *Cycle*.

6.4.2 Implications et inférences

Dans la modélisation de HuTO, nous avons choisi d'utiliser le langage RDFS plutôt que OWL. Le choix s'est porté sur RDFS parce que la sémantique de RDFS est plus légère que celle de OWL et suffisante pour permettre de répondre à la plupart des requêtes temporelles. En effet comme le montre Ter Hort [94], la sémantique de RDFS est définie en utilisant des conditions IF (si) alors que celle de OWL est définie en utilisant des conditions IF-and-If only (si et seulement si). Une

condition *IF-and-If only* est plus puissante, dans la mesure où elle conduit à plus d'implications mais également plus contraignante. De plus, comme notre objectif est d'annoter des ressources RDF et que RDFS intègre la plupart des inférences qui nous intéressent (*RDFS entailments*), nous avons choisi d'utiliser RDFS comme langage pour la modélisation de HuTO.

Cependant, comme RDFS n'implémente pas certaines inférences de base comme la transitivité ou la réflexivité, nous avons dû créer quelques règles d'inférence à cet effet. Ainsi, nous avons défini des règles d'inférence pour la transitivité des propriétés *before/after*. De même, si une relation (*after* ou *before*) est exprimée entre deux événements (respectivement intervalles), il est nécessaire de propager cette relation entre les intervalles (respectivement ressources) concernés. Pour cela, nous avons proposé des règles de propagation. Au total, 26 règles d'implications et d'inférences ont été définies pour les propriétés *before* et *after*. Ces règles sont définies pour la transitivité, l'inverse et la propagation.

```

PREFIX huto: <http://ns.inria.fr/huto/>
CONSTRUCT{ ?x huto:before ?y }
WHERE{ ?s huto:before ?o .
      ?x a huto:TemporalAnnotation ;
      huto:hasTemporalExp ?s .
      ?y a huto:TemporalAnnotation ;
      huto:hasTemporalExp ?o .

      FILTER NOT EXISTS{ ?x huto:before ?y }
}

```

Exemple 6.11: Règle de propagation de la propriété *before* entre des ressources.

La règle de l'exemple 6.11 propage la relation *before* exprimée entre deux intervalles sur les ressources qui ont ces intervalles comme occurrence.

Pour vérifier l'ordre d'emboîtement des concepts *Cycle* et *Contiguous*, nous avons défini la propriété *included* (Fig. 6.2) qui permet de hiérarchiser les unités temporelles. Ainsi, nous avons explicité dans HuTO sept relations entre les unités temporelles (*included(Year, Century)*, *included(Month, Year)*, ...). Nous avons également défini deux règles de propagation: une règle pour la transitivité et une autre pour la transitivité par la subordination : si *included(d2, d1)* et *rdfs:subClassOf(d3, d2)* alors *included(d3, d1)* (exemple 6.12).

```

PREFIX huto: <http://ns.inria.fr/huto/>
CONSTRUCT { ?y huto:included ?x }
WHERE { ?y rdfs:subClassOf+/huto:included ?x }

```

Exemple 6.12: Règle de propagation par la subordination de la propriété *included*.

Notons que nous avons défini 52 règles¹⁴ de normalisation et d'inférence et que toutes ces règles sont exprimées sous forme de requêtes CONSTRUCT en SPARQL interprétées comme des règles. Nous distinguons deux types de règles. Le premier permet d'ajouter des informations dans le graphe RDF, conséquence de la modélisation du temps humain où certaines informations sont implicites. Le deuxième type de règle permet de réaliser des inférences, conséquence de l'utilisation du langage RDFS qui a certaines limites d'inférences.

Toutes ces règles sont interprétées comme des règles en chaînage avant à saturation et la base est donc enrichie.

6.5 Requêtes-types

Une requête-type est une requête paramétrée (avec des variables) qui doit être instanciée avec des valeurs spécifiques avant son utilisation. L'instanciation se fait avec la clause `VALUES` de SPARQL. Les requêtes-types permettent de répondre à des questions récurrentes sur les annotations temporelles. Ainsi, elles permettent une utilisation plus facile de l'ontologie HuTO dans la recherche d'information temporelles.

L'ontologie HuTO (ontologie et règles) a été testée avec Corese [95], un moteur sémantique qui permet le traitement de ressources en RDF/S, de requêtes en SPARQL et dispose d'un langage de règles adapté à RDF. Notre jeu de données compte 1256 triplets. L'exécution des implications de RDFS (*RDFS entailments*) l'a étendu à 1999 triplets et l'application de nos règles à 2705 triplets. Ce jeu de données porte sur l'annotation temporelle de connaissances socioculturelles au Sénégal. Il a été utilisé pour tester les requêtes-types définies dans cette section.

Nous distinguons deux requêtes-types :

1. Une requête-type sur les ressources : elle détermine la période d'occurrence d'une ressource donnée.
2. Les requêtes-types sur les éléments temporels : elles déterminent les ressources relatives à une période d'occurrence donnée ou relatives à une ressource temporellement annotée.

6.5.1 Requête-type sur les ressources

Une requête temporelle sur les ressources d'une base de connaissances permet de déterminer la période d'occurrence des ressources. La requête-type prend en compte les trois représentations qui peuvent être utilisées pour l'annotation temporelle d'une

¹⁴<http://ns.inria.fr/huto/rules>

donnée (une ressource, un triplet ou un graphe nommé). Ainsi, quelle que soit la représentation utilisée, cette requête-type permet de retrouver la temporalité de la ressource.

La requête SPARQL de l'exemple de 6.13 est un exemple de requête écrite à partir de la requête-type sur les ressources¹⁵ dans laquelle la ressource recherchée a été précisée. Elle permet de déterminer la temporalité de la ressource *data:Gamou*.

```
PREFIX huto: <http://ns.inria.fr/huto/>
PREFIX data: <http://example.org/data/>
DESCRIBE ?x
WHERE{
  { ?x huto:uri ?resource }
  UNION
  { ?x huto:triple/(rdf:subject|rdf:object) ?resource }
  UNION
  { ?x huto:graph ?g .
    graph ?g{ { ?resource ?p ?o }
              UNION
              { ?s ?p ?resource }
            }
  }
  FILTER NOT EXISTS { ?j ?k ?x }
}
VALUES ?resource { data:Gamou }
```

Exemple 6.13: Requête déterminant la temporalité de la ressource *data:Gamou*.

6.5.2 Requêtes-types sur les éléments temporels

Les requêtes-types sur les éléments temporels permettent de déterminer les ressources relatives à une période donnée ou à une ressource temporellement annotée. Nous avons identifié six requêtes-types : deux sont relatives à une ressource et quatre relatives à un intervalle (période).

Les deux requêtes-types relatives à une ressource permettent de déterminer les ressources dont la période d'occurrence est avant ou après une ressource temporellement annotée.

La requête SPARQL de l'exemple 6.14 ci-dessous est un exemple de requête écrite à partir de la requête-type pour chercher les événements (ressources) ayant eu lieu avant une ressource donnée. Dans cet exemple, la ressource référence a été précisée (clause **VALUES**) : *data:Gamou*.

Ainsi, cette requête renvoie toutes les ressources dont la date d'occurrence (date de début) vient après celle de la ressource *data:Gamou*. La requête ne concerne que

¹⁵En effet, dans les requêtes-types il n'y a pas de clause **VALUES**. La clause est utilisée pour instancier des variables (dans cet exemple la variable **?resource**).


```

33.           FILTER(?d1 > ?d2)
34.         })
35.       )
36.     })
37.   )
39. }
40. VALUES ?resource {data:Gamou }

```

Exemple 6.14: Requête déterminant la temporalité des ressources dont l'occurrence est avant celle de la ressource *data:Gamou*.

La requête-type qui détermine les ressources dont la période d'occurrence est après une ressource temporellement annotée a la même forme que la requête de l'exemple 6.14 à la différence qu'au lieu d'utiliser l'opérateur supérieur (>) pour la comparaison, nous utilisons l'opérateur inférieur (<).

Les quatre requêtes-types relatives à un intervalle (période) donné permettent de rechercher :

1. Les ressources répétitives par rapport à une fréquence donnée. Cette fréquence peut être hebdomadaire, mensuelle (exemple 6.15), annuelle, etc.
2. Les ressources qui sont vraies à une date donnée. Cette requête renvoie toutes les ressources dont la durée d'occurrence contient la date donnée. Par exemple les événements qui ont eu lieu le 30 Mars 2015.
3. Les ressources qui sont vraies pendant un intervalle donné. Cette requête renvoie toutes les ressources dont la durée d'occurrence se trouve dans l'intervalle demandé. Par exemple les événements qui ont eu lieu entre le 30 Mars et le 5 Avril 2016.
4. Les ressources qui sont vraies sur une période donnée. Cette requête renvoie toutes les ressources dont la durée d'occurrence se trouve dans la période demandée. Par exemple les événements qui ont eu lieu à partir du 30 Mars 2016 et pour les 45 jours qui suivent le 30 Mars 2016.

La requête SPARQL présentée dans l'exemple 6.15 est un exemple de requête écrite à partir de la requête-type pour rechercher des ressources répétitives en fonction de leur périodicité. Dans cet exemple, la fréquence a été précisée par `Month`.

Toutes les requêtes-types sont données en annexe.

Notons que pour toutes ces requêtes-types, les résultats sont des ressources temporellement annotées à l'aide d'intervalles convexes, non convexes ou d'autres ressources utilisées comme référence temporelle.

```
PREFIX huto: <http://ns.inria.fr/huto/> .
DESCRIBE ?s
WHERE {
    BIND ( CONCAT ('huto:', ?period) as ?frequency)

    ?s huto:hasTemporalExp ?o .
    ?o rdf:type huto:Cycle ;
        huto:every ?q .
    ?q rdf:type ?frequency .

    FILTER NOT EXISTS { ?q huto:sample ?t }
}
VALUES ?period {Month}
```

Exemple 6.15: Requête déterminant les ressources qui se produisent mensuellement.

6.6 Transformations

Dans cette section, le langage de transformation STTL (SPARQL Template Transformation Language) [96, 97] est utilisé pour une présentation plus adaptée des résultats des requêtes aux utilisateurs. STTL permet de spécifier des transformations de graphes RDF dans des formats tels que HTML et même en langue naturelle.

STTL repose sur deux extensions de SPARQL: (i) une requête **Template** qui permet d'exprimer des règles de transformation et (ii) des fonctions d'extension pour appeler de manière récursive d'autres **Template** (`st:apply-templates`¹⁶). Une requête **Template** est composée d'une clause **Template** et d'une clause **WHERE** standard [96]. La clause **WHERE** donne les conditions d'application de la transformation en précisant les nœuds sur lesquels la transformation doit s'appliquer. La clause **Template** permet de définir la façon de présenter les nœuds RDF sélectionnés.

Un moteur STTL a été intégré dans Corese [95] et nous avons utilisé ce dernier pour proposer des transformations sur les résultats des requêtes de la section 6.5. Nous avons utilisé l'extension `st:apply-templates` de STTL pour ces transformations. Étant donné un graphe RDF, cette extension essaie successivement d'appliquer les **Templates** sur les nœuds jusqu'à ce que l'un d'eux réussisse [96]. Ainsi, un **Template** qui réussit est un **Template** qui vérifie les conditions de la clause **WHERE** et qui retourne un résultat.

Une transformation est un ensemble cohérent de **Templates**. Nos transformations sont faites sur le jeu de données décrit précédemment qui est un jeu de données constitué de ressources socioculturelles concernant le Sénégal.

¹⁶Le préfixe `st:` correspond au namespace <http://ns.inria.fr/sparql-template/>

```

Graph XML Validate
</result>
</result>
<binding name='z'><bnode>_b97</bnode></binding>
</result>
</result>
<binding name='z'><bnode>_b103</bnode></binding>
</result>
</result>
<binding name='z'><bnode>_b138</bnode></binding>
</result>
</result>
<binding name='z'><bnode>_b232</bnode></binding>
</result>
</result>
<binding name='z'><bnode>_b270</bnode></binding>
</result>
</result>
<binding name='z'><bnode>_b283</bnode></binding>
</result>
</result>
<binding name='z'><bnode>_b295</bnode></binding>
</result>
</result>
<binding name='z'><bnode>_b301</bnode></binding>
</result>
</result>
<binding name='z'><bnode>_b310</bnode></binding>
</result>
</result>
<binding name='z'><bnode>_b318</bnode></binding>
</result>
</result>
<binding name='z'><bnode>_b324</bnode></binding>
</result>
</result>
<binding name='z'><bnode>_b332</bnode></binding>
</result>
</result>
<binding name='z'><bnode>_b372</bnode></binding>
</result>
</result>
</result>
</result>
<rdf:Description rdf:nodeID='_b324'>
<chuto:hasTemporalExp rdf:nodeID='_b325'/>
<chuto:uri rdf:resource='http://example.org/data/WorldWarOne'/>
<rdf:type rdf:resource='http://ns.inria.fr/huto/TemporalAnnotation'/>
</rdf:Description>
</result>
</result>
<rdf:Description rdf:nodeID='_b138'>
<chuto:hasTemporalExp rdf:nodeID='_b139'/>
<chuto:uri rdf:resource='http://example.org/data/Simb16'/>
<rdf:type rdf:resource='http://ns.inria.fr/huto/TemporalAnnotation'/>
</rdf:Description>
</result>
</result>
<rdf:Description rdf:nodeID='_b103'>
<chuto:hasTemporalExp rdf:nodeID='_b104'/>
<chuto:uri rdf:resource='http://example.org/data/Simb102'/>
<rdf:type rdf:resource='http://ns.inria.fr/huto/TemporalAnnotation'/>
</rdf:Description>
</result>
</result>
<rdf:Description rdf:nodeID='_b270'>
<chuto:hasTemporalExp rdf:nodeID='_b271'/>
<chuto:uri rdf:resource='http://example.org/data/DoingSport'/>
<rdf:type rdf:resource='http://ns.inria.fr/huto/TemporalAnnotation'/>
</rdf:Description>
</result>
</result>
<rdf:Description rdf:nodeID='_b332'>
<chuto:hasTemporalExp rdf:nodeID='_b333'/>
<chuto:uri rdf:resource='http://example.org/data/02014'/>
<rdf:type rdf:resource='http://ns.inria.fr/huto/TemporalAnnotation'/>
</rdf:Description>
</result>
</result>
<rdf:Description rdf:nodeID='_b301'>
<chuto:hasTemporalExp rdf:nodeID='_b303'/>
<chuto:triple rdf:nodeID='_b302'/>
<rdf:type rdf:resource='http://ns.inria.fr/huto/TemporalAnnotation'/>
</rdf:Description>
</result>
</result>
<rdf:Description rdf:nodeID='_b283'>
<chuto:hasTemporalExp rdf:nodeID='_b284'/>

```

Figure 6.4: Résultats de clauses SELECT (à gauche) et DESCRIBE (à droite).

6.6.1 Transformation RDF2Text

La transformation RDF2Text que nous avons proposée engendre du texte qui précise l'URI de la ressource et sa date d'occurrence. En effet dans la modélisation de l'ontologie HuTO, les ressources sont emboîtées dans des *blank nodes* et l'utilisation des clauses standard de SPARQL comme SELECT et DESCRIBE donne des résultats peu lisibles. Une clause SELECT renvoie des *blank nodes* (à gauche de la figure 6.4) et une clause DESCRIBE met les informations temporelles dans des blank nodes (à droite de la figure 6.4).

La transformation RDF2Text est constituée de 22 **Templates**¹⁷. Chaque **Template** est spécifique à un type de nœud. Le langage offre la possibilité de choisir un nœud de départ dans le graphe pour démarrer la transformation grâce au **Template st:start**. Dans notre cas, nous avons identifié que les nœuds de départ sont ceux de type *TemporalAnnotation* (exemple 6.17). Ainsi à partir d'un nœud de type *TemporalAnnotation*, nous définissons deux types de transformations : (i) une transformation pour les ressources (au sens large) et (ii) une transformation pour les données temporelles.

¹⁷<http://ns.inria.fr/huto/rdf2text/>

```
PREFIX huto: <http://ns.inria.fr/huto/> .
TEMPLATE st:start {
  st:apply-templates(?resource)
  "Date: "
  st:apply-templates(?exp)
}
WHERE {
  ?in a huto:TemporalAnnotation ;
      (huto:triple|huto:uri|huto:graph) ?resource ;
      huto:hasTemporalExp ?exp .

  FILTER NOT EXISTS { ?y ?p ?in }
}
```

Exemple 6.17: Template d'entrée des transformations.

Comme HuTO permet d'annoter temporellement des ressources, des triplets et des graphes nommés (cf. section 6.3), nous avons défini trois **Templates** pour les données annotées : 1) un **Template** pour une ressource, 2) un **Template** pour une réification RDF et 3) un **Template** pour un graphe nommé. Ainsi, si le résultat à afficher est :

1. Une ressource : l'URI de la ressource est affiché.
2. Une réification RDF : le triplet asserté est affiché (exemple 6.18).
3. Un graphe nommé : les triplets qui constituent le graphe sont affichés.

L'exemple 6.18 représente le **Template** d'affichage d'une triplet.

```
PREFIX huto: <http://ns.inria.fr/huto/> .
TEMPLATE { ?s " " ?p " " ?o }
WHERE { ?in rdf:subject      ?s ;
         rdf:predicate      ?p ;
         rdf:object         ?o .
}
```

Exemple 6.18: Template d'une réification RDF.

Pour afficher les intervalles résultats des requêtes, nous avons défini seize **Templates**. Parmi ces **Templates**, cinq sont utilisés pour les intervalles non convexes, chacun pour un cycle particulier.

L'exemple 6.19 montre le **Template** de traduction d'un cycle mensuel.

```

PREFIX huto: <http://ns.inria.fr/huto/> .
TEMPLATE {
  "each month at " st:apply-templates(?exp)
}
WHERE {
  ?in a huto:Cycle ;
  huto:every/rdf:type huto:Month ;
  huto:exp ?exp .
}

```

Exemple 6.19: Template pour les cycles mensuels.

Dix autres **Templates** ont été écrits pour traduire les intervalles convexes. En effet dans la modélisation de HuTO, un intervalle non convexe est composé d'une fréquence et d'un intervalle convexe qui se répète par rapport à sa fréquence. Chaque **Template** permet de traduire un format d'intervalle convexe. En effet avec la modélisation de HuTO, un intervalle convexe (`huto:Contiguous`) peut être spécifié avec la propriété `huto:hasDate`, avec les propriétés `huto:hasBegin` et `huto:hasEnd` et avec la propriété `huto:hasDuration` combinée avec l'une des propriétés `huto:hasBegin` ou `huto:hasEnd`.

L'exemple 6.20 montre le **Template** pour traduire l'intervalle répétitif délimité par une date de début et une date de fin dans le traitement d'un intervalle convexe (dans la variable `?e` se trouve l'intervalle répétitif).

```

PREFIX huto: <http://ns.inria.fr/huto/> .
TEMPLATE {"From: "st:apply-templates(?begin)
  " To: "st:apply-templates(?end) ", "
  st:apply-templates(?e) }
WHERE { ?in a huto:Contiguous ;
  huto:hasBegin ?begin ;
  huto:hasEnd ?end ;
  huto:exp ?e .}

```

Exemple 6.20: Template pour les intervalles exprimés à l'aide de leurs dates de début et de fin.

Nous avons également proposé un **Template** spécifique à la traduction d'une date grégorienne (exemple 6.21) qui est utilisé par tous les **Templates** pour afficher la date d'un intervalle convexe.

```

PREFIX huto: <http://ns.inria.fr/huto/> .
TEMPLATE { str(?day)
            str(?month)
            str(?week)
            str(?year)
            str(?century)
            str(?hour)
            str(?minutes)
}
WHERE {
    ?in rdf:type huto:Date .
    OPTIONAL{?in huto:hasMinute/huto:value ?min}
    OPTIONAL{?in huto:hasHour/huto:value ?h}
    OPTIONAL{?in huto:hasDay/huto:value ?d1}
    OPTIONAL{?in huto:hasDay/rdfs:label ?d2 .
              FILTER langMatches(lang(?d2), "EN" )}
    OPTIONAL{?in huto:hasDay/huto:number ?n}
    OPTIONAL{?in huto:hasMonth/rdfs:label ?m
              FILTER langMatches(lang(?m), "EN")}
    OPTIONAL{?in huto:hasWeek/huto:value ?w}
    OPTIONAL{?in huto:hasYear/huto:value ?y}
    OPTIONAL{?in huto:hasCentury/huto:value ?c}

    BIND(IF(?d1= 1, "st ", IF(?d1= 2, "nd ", IF(?d1= 3, "rd ", "th "
    ))) as ?nd1)

    BIND(IF(?n = 1, "st ", IF(?n = 2, "nd ", IF(?n = 3, "rd ", "th "
    ))) as ?nn)

    BIND(IF(BOUND(?min), CONCAT(":",?min,"min "), "" ) as ?minutes)

    BIND(IF(BOUND(?h), CONCAT(" at ",?h,"H "), "" ) as ?hour)

    BIND(IF(BOUND(?d1), CONCAT(?d1, ?nd1), IF(BOUND(?d2),
    IF(BOUND(?n), CONCAT("the ", ?n, ?nn, ?d2), ?d2),""))
    as ?day)

    BIND(IF(BOUND(?w), CONCAT(" of the week: ", ?w), "" ) as ?week)
    BIND(IF(BOUND(?m), CONCAT(" ", ?m), "" ) as ?month)
    BIND(IF(BOUND(?y), CONCAT(" ", ?y), "" ) as ?year)
    BIND(IF(BOUND(?c), CONCAT("Century ", ?c), "" ) as ?century)
}

```

Exemple 6.21: Template d'une date grégorienne.

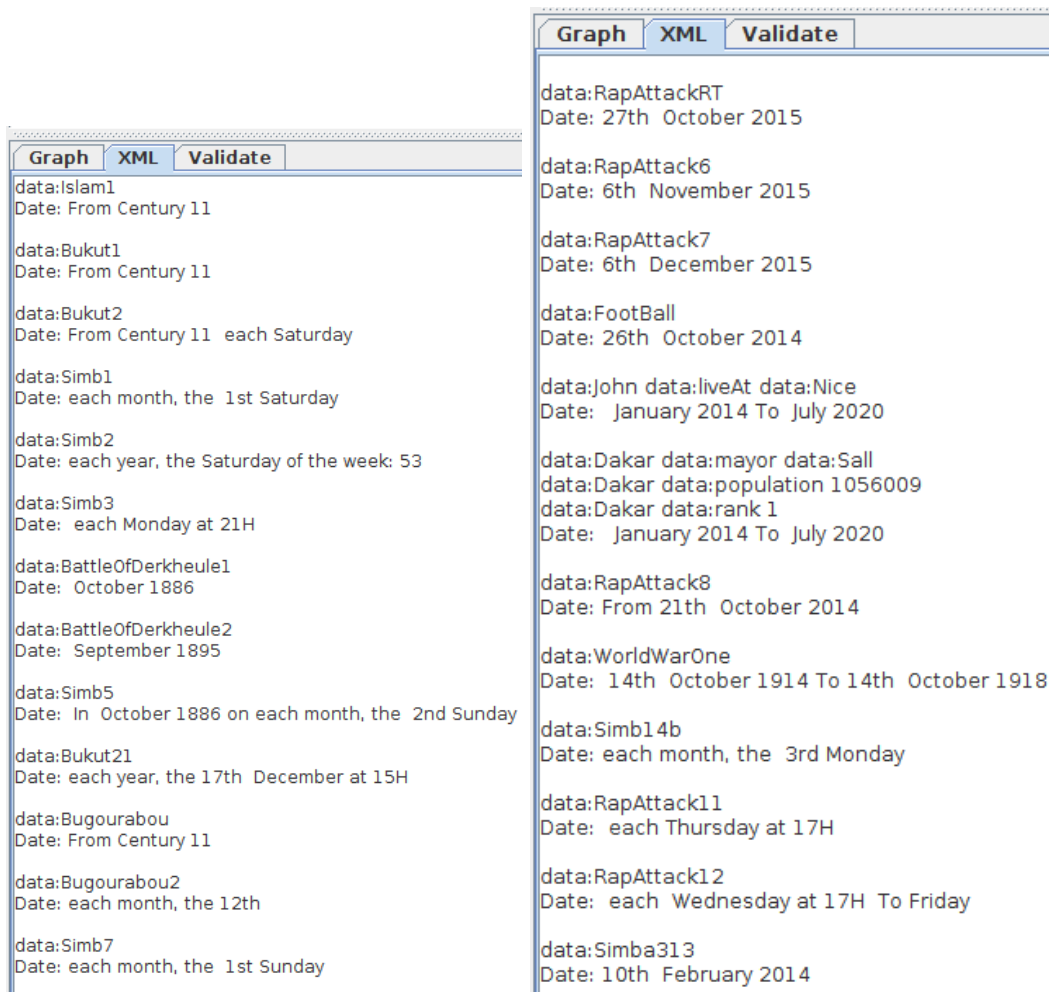


Figure 6.5: Exemple de résultats de transformations sur une requête temporelle.

La figure 6.5 montre un exemple de résultat de l'application des transformations sur une requête temporelle. Ainsi, la transformation `RDF2Text`, qui permet d'obtenir un résultat plus lisible et compréhensible pour une personne, prend en compte tous les types d'objets de la modélisation de HuTO. En effet, chaque type de nœud a un `Template` spécifique et le résultat produit est spécifique à ce nœud.

L'utilisation du langage STTL a permis de définir une transformation `RDF2Text` qui permet un rendu plus lisible et compréhensible des requêtes SPARQL sur des données annotées grâce à HuTO.

6.7 Synthèse

Dans ce chapitre, nous avons présenté l'ontologie temporelle narrative, HuTO (Human Time Ontology). HuTO est une ontologie en RDFS pour annoter temporelle-

ment des données en RDF à l'aide d'expressions du langage courant. Ce travail repose sur deux domaines de recherche dans la modélisation temporelle en Web sémantique : 1) la modélisation d'expressions temporelles et 2) l'annotation temporelle des données.

La modélisation d'expressions temporelles permet de modéliser une date, un intervalle, des notions temporelles répétitives, relatives ou absolues. Dans ce domaine, HuTO permet la modélisation d'énoncés de temps complexe (exemple 6.4f). Dans notre approche nous considérons toute entité temporelle comme un intervalle pouvant être défini à l'aide de différentes unités (granularités) calendaires. Une correspondance existe d'une part entre le type *xsd:duration* et le concept *Date* de HuTO et d'autre part entre le type *xsd:dateTime* et le concept *Duration* de HuTO. Une distinction est faite entre les intervalles fermés et infinis de même entre les intervalles convexes et non convexes. Notre ontologie intègre aussi les relations temporelles *before* et *after* telles que définies par Allen [57, 58]. Ces relations sont définies soit entre deux intervalles soit entre deux ressources soit entre un intervalle et une ressource. HuTO permet aussi d'utiliser une ressource comme une référence temporelle pour dater une autre ressource. Une des contributions majeures de HuTO est la modélisation des intervalles non convexes de façon à permettre l'écriture de requêtes SPARQL qui permettent de considérer tout type d'intervalle. Différentes unités temporelles sont définies dans HuTO qui peuvent être étendues à d'autres unités. HuTO permet de modéliser plusieurs expressions temporelles déictiques comme *Hier*, *Aujourd'hui*. HuTO permet aussi une représentation temporelle qualitative qui est une modélisation de notions temporelles relatives aux positions géographiques ou à la culture.

En ce qui concerne l'annotation temporelle de données, HuTO propose une approche qui associe une dimension temporelle distincte des données qui rend la recherche d'information plus facile en permettant de considérer ou non l'aspect temporel. Cette approche permet de garder l'évolution temporelle des données que ce soit une ressource, un triplet ou un ensemble de triplets (graphe nommé).

Hormis ces contributions, l'ontologie HuTO intègre également des règles¹⁸ pour la normalisation et le raisonnement. Ces règles permettent d'ajouter des informations dans le graphe RDF soit pour expliciter certaines informations (règles de normalisation) soit pour ajouter des informations implicites (règles de raisonnement).

Nous avons aussi proposé des requêtes-types pour interroger des données annotées grâce à HuTO. Ainsi, nous avons identifié sept requêtes-types : une requête-type qui permet de déterminer la période d'occurrence d'une ressource et six requêtes-types pour déterminer les ressources par rapport à une période donnée ou relative à une ressource annotée. Les résultats de ces requêtes sont des ressources annotées à l'aide d'intervalles convexes, non convexes et des ressources annotées par des références temporelles.

Pour avoir des réponses plus lisibles par un humain, nous avons défini une trans-

¹⁸<http://ns.inria.fr/huto/rules/>

	OWL-Time	CNTRO	TimeML	HuTO
Intervalle non convexe	✓	✓	✓	✓
Intervalle Infini	✓	×	✓	✓
Temps Déictique	×	✓	✓	✓
Référence Temporelle	×	✓	×	✓
Différentes Unités Temporelles	✓	✓	✓	✓
Modélisation Qualitative	×	×	×	✓
13 Relations d'Allen	✓	✓	partiel- lement	partiel- lement

Table 6.1: Comparaison de HuTO avec les trois approches les plus significatives dans le domaine de la modélisation d'expressions temporelles (✓ signifie que l'ontologie intègre cette caractéristique sinon ×).

formation basée sur le langage STTL [96, 97]. Cette transformation RDF2Text¹⁹ permet d'obtenir un affichage en langue naturelle plus compréhensible pour les humains.

6.7.1 Comparaison de HuTO avec les approches de l'état de l'art

Dans le chapitre 4, nous avons proposé des critères de comparaisons des ontologies temporelles. Ces critères prennent en consideration la modélisation d'expressions temporelles et l'annotation temporelle des données.

Le tableau 6.1 présente la comparaison des trois ontologies les plus complètes dans le domaine de la modélisation d'expressions temporelles avec HuTO suivant sept critères. Parmi ces quatre approches, seule OWL-Time permet de modéliser des exceptions dans les intervalles non convexes. Cependant, OWL-Time ne permet pas de modéliser des expressions comme *Every month except July*. OWL-Time ne prend pas en compte la modélisation du temps déictique et CNTRO le modélise dans des chaînes de caractères ce qui limite le raisonnement. Une référence temporelle permet à une ressource d'être utilisée comme une date. Seules les approches HuTO et CNTRO modélisent cet aspect. Toutes les approches du tableau 4.3 modélisent les unités temporelles grégoriennes. L'un des points forts de l'ontologie HuTO par rapport aux autres approches est qu'elle ne se limite pas aux unités grégoriennes mais intègre par exemple les unités lunaires. Notre approche peut être étendue à d'autres unités comme des unités Mayas et Aztèques. Seule l'approche HuTO permet une modélisation qualitative, c'est-à-dire la modélisation de notions temporelles relatives à une culture ou à une zone géographique de telle sorte qu'elles puissent être utilisées comme des références temporelles. En ce qui concerne les relations d'Allen, TimeML et HuTO n'intègrent que deux relations alors que OWL-Time et CNTRO intègrent toutes les relations et leurs caractéristiques

¹⁹<http://ns.inria.fr/huto/rdf2text/>

	Prolifération / Redondance des objets	Perte de l'assertion des triplets	Extension des standards du W3C
Réification RDF	✓	✓	×
Graphes nommés	✓	×	×
Relations n-aires	✓	✓	×
Temporal RDF	✓	×	✓
4D-Fluents	✓	✓	×
TDL	×	×	✓
HuTO	✓	×	×

Table 6.2: Comparaison de HuTO avec les approches majeures dans le domaine de l'annotation temporelle des données (✓ signifie que l'ontologie intègre cette caractéristique sinon ×).

logiques. Contrairement à TimeML, HuTO intègre les caractéristiques logiques des relations *before* et *after*.

Dans le domaine de la modélisation d'expressions temporelles, HuTO est donc plus complète que les autres approches surtout en ce qui concerne l'expression du temps humain qui est importante dans le domaine culturel.

Le tableau 6.2 présente la comparaison suivant trois critères de six approches dans le domaine de la modélisation d'expressions temporelles avec HuTO . Dans HuTO, nous combinons les approches de la réification RDF et des graphes nommés. En effet, si un triplet doit être annoté, c'est l'approche par réification RDF qui est choisie. Si c'est un ensemble de triplets, l'approche par les graphes nommés est choisie. Cela a comme effet de diminuer la prolifération et la redondance des objets comparé aux approches de la réification de RDF, Temporal RDF et aux relations n-aires si plusieurs triplets partagent la même période de temps. En effet, dans ce cas il est plus avantageux d'utiliser l'approche des graphes nommés où un seul nœud temporel est utilisé pour l'ensemble des triplets. En ce qui concerne l'approche 4D-Fluent, la modélisation temporelle nécessite deux objets supplémentaires (les parties temporelles) et six triplets alors que pour les autres approches (HuTO, réification RDF, Temporal RDF et Relations n-aires) seulement un objet supplémentaire est nécessaire, quatre triplets pour les approches HuTO, réification RDF et Temporal RDF et trois triplets pour l'approche à l'aide des relations n-aires. En ce qui concerne la perte de la sémantique du triplet, HuTO adopte l'approche Temporal RDF ce qui permet d'éviter cet écueil. L'utilisation de HuTO ne nécessite pas d'étendre les standards du W3C comme le font Temporal RDF et les approches TDL. L'extension STTL n'est utilisée que pour l'affichage des requêtes SPARQL. Elle ne fait pas partie de la spécification de HuTO [98].

En résumé, les avantages majeurs de l'ontologie HuTO par rapport à toutes les approches présentées dans le chapitre 3 sont :

1. La combinaison de la modélisation d'expressions temporelles et de l'annotation temporelle de données.
2. La modélisation d'intervalles convexes et non convexes et aussi la modélisation d'intervalles fermés et infinis.
3. La modélisation de différentes unités temporelles (grégorienne, lunaire) qui peuvent être étendues à d'autres unités (Maya, Aztèque, etc).
4. La modélisation d'expressions temporelles déictiques.
5. La modélisation de notions temporelles qualitatives qui peuvent être utilisées comme des références temporelles.

Dans ce chapitre, nous avons décrit la troisième contribution. Elle s'inscrit dans le domaine de la modélisation temporelle pour annoter des données du Web sémantique.

Implémentation et Validation

Sommaire

7.1	Introduction	127
7.2	Architecture de la plateforme	128
7.2.1	Couche IHM	129
7.2.2	Couche persistance	132
7.2.3	Couche sémantique	134
7.3	Ontologie socioculturelle du Sénégal	135
7.4	Concepts temporels de la plateforme	137
7.5	Prototype	141
7.5.1	Outils installés	142
7.5.2	Implémentation	143
7.5.3	Maquette	149
7.6	Synthèse	154

7.1 Introduction

Dans ce chapitre, nous présentons le prototype de la plateforme WestAfricape-
dia, que nous avons développé, pour permettre aux communautés sénégalaises de
partager et de co-construire leur patrimoine culturel. Cette plateforme intègre les
ontologies présentées dans les chapitres précédents de ce mémoire. Ainsi, la plate-
forme repose essentiellement sur l'ontologie USCO qui permet à des communautés
de pouvoir partager et co-construire leurs connaissances socioculturelles. Puisque
l'ontologie USCO est alignée avec les vocabulaires de Schema.org et DBpedia, ces
derniers sont également intégrés et utilisés dans la plateforme. Pour permettre
l'annotation temporelle des ressources, l'ontologie HuTO est utilisée.

Les fonctionnalités de la plateforme doivent également permettre aux membres
des communautés la co-élaboration de connaissances et permettre d'échanger et
confronter leur point de vue dans le processus de co-construction de connaissances.

Ce chapitre continue comme suit. Dans la deuxième partie, l'architecture de la
plateforme est présentée avec ses trois couches. La partie trois présente l'ontologie
USCO dans le contexte sénégalais qui est utilisée dans la plateforme. La quatrième

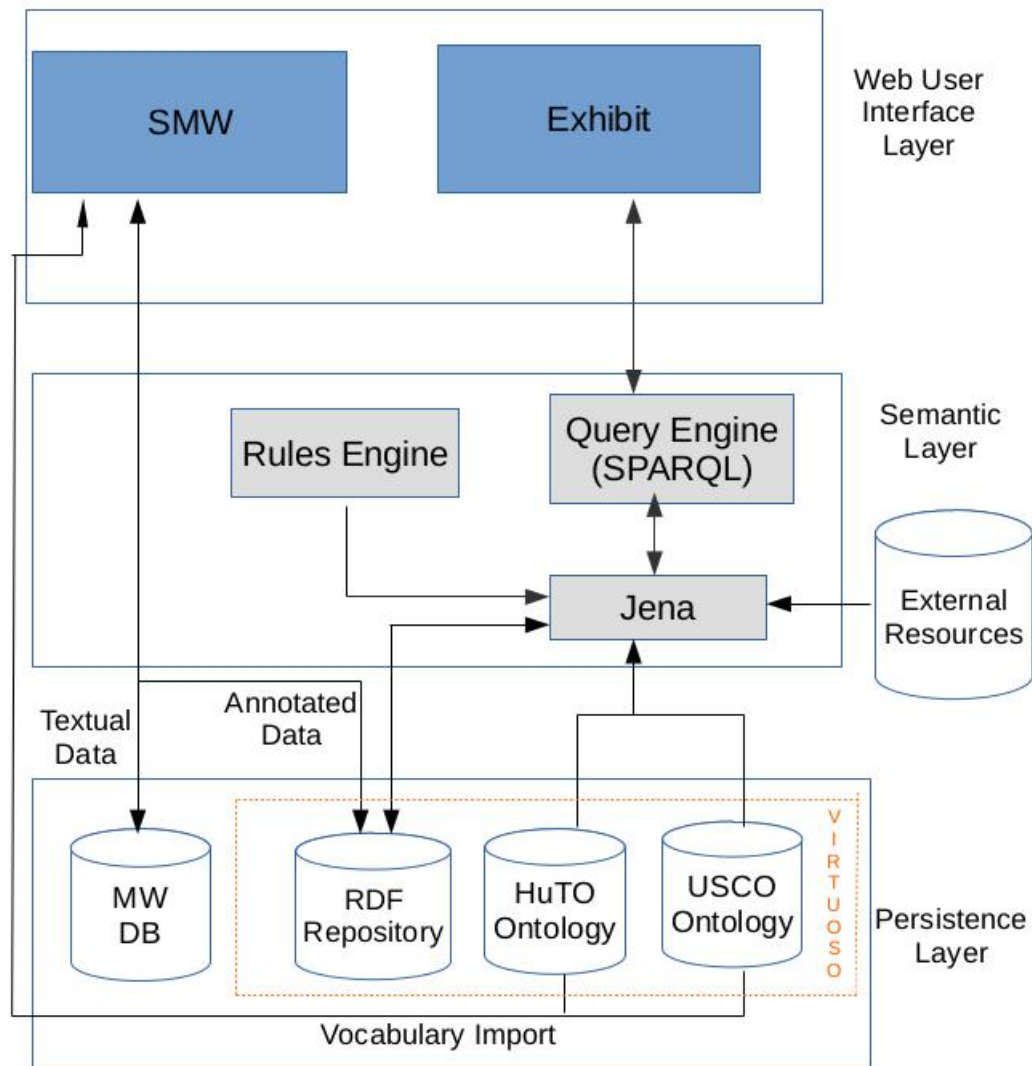


Figure 7.1: Architecture de la plateforme WestAfricapedia.

partie présente les notions temporelles qualitatives de HuTO dans le contexte sénégalais. La cinquième partie présente le prototype de la plateforme en présentant les outils utilisés, l'implémentation faite et la maquette de la plateforme. Enfin, nous terminons ce chapitre par une synthèse.

7.2 Architecture de la plateforme

Le prototype développé repose sur trois entités : 1) Semantic Mediawiki (SMW) qui est le socle de la plateforme, 2) Virtuoso qui représente notre triple store (base de données RDF) et Exhibit qui est notre outil de visualisations des données. La figure 7.1 représente les différentes couches de l'architecture de la plateforme WestAfrica-

pedia. Cette architecture est composée de trois couches. Une Interface Homme-Machine (IHM) supportée par SMW et Exhibit. Cette couche permet aux utilisateurs de participer à la co-construction de connaissances. Elle permet aussi de présenter les données de façon conviviale (Exhibit). La couche sémantique permet de disposer d'une interface de requêtes et d'un moteur de règles. La couche de persistance des données composée d'une base de données relationnelle et d'un triple store (base de données sémantiques).

7.2.1 Couche IHM

La couche IHM est composée d'outils mis en œuvre afin qu'un humain puisse contrôler et communiquer avec la plateforme. Le premier outil est SMW, une extension de MediaWiki, un Wiki sémantique. Le deuxième outil est Exhibit qui est un outil du projet SIMILE.

7.2.1.1 Semantic Mediawiki (SMW)

Dans la section 5.3.2, nous avons présenté MediaWiki qui est la technologie utilisée pour permettre aux communautés de co-élaborer des connaissances culturelles, de co-construire et de partager leurs connaissances socioculturelles. Cependant, comme nous comptons avoir une couche sémantique pour permettre à ces communautés d'annoter les données, nous utilisons SMW qui permet d'ajouter cette couche sémantique à MediaWiki. Ainsi, l'utilisation de l'extension SMW permet de bénéficier de toutes les caractéristiques de MediaWiki. De plus, elle permet d'ajouter une couche sémantique pour permettre aux utilisateurs d'annoter les données.

Semantic MediaWiki (SMW) est une extension de MediaWiki, le moteur Wiki sur lequel repose Wikipedia. Contrairement aux wikis traditionnels qui ne peuvent contenir que du texte ne pouvant être ni compris, ni évalué par les agents logiciels, SMW ajoute des annotations sémantiques dans les pages du Wiki en utilisant une extension du langage de script de MediaWiki. Ainsi, SMW permet à un MediaWiki de fonctionner comme une base de connaissances collaborative. La première version de SMW a été diffusée en 2005. Aujourd'hui SMW compte plus d'une vingtaine d'extensions qui permettent d'étendre la capacité de modifier, d'afficher et de naviguer au travers des données stockées par SMW : le terme "Semantic MediaWiki" est quelquefois utilisé pour désigner la famille complète des extensions.

SMW et Web sémantique

L'unité de base dans un site MediaWiki (MW) est une **page Wiki**. Le nom complet d'une page Wiki est défini sous la forme `NAMESPACE:NOM`, où des espaces de noms sont prédéfinis comme `Category` et `User`. MW dispose d'un langage de script pour décrire le contenu des pages du Wiki. Ce langage a été étendu par SMW avec les trois ensembles d'annotations sémantiques suivantes :

1. **Classes et Propriétés** : SMW réutilise l'espace de noms (`namespace`) `Category` pour définir des classes. Par exemple, une page Wiki avec le nom `Category:Person` est destinée à représenter la classe de toutes les personnes. De même, un nouvel espace de noms appelé `Property` est introduit pour définir les propriétés des concepts. En fonction du type de la propriété, il s'agit d'une relation (relation entre deux pages: `Type:Page`) ou d'un attribut (relation entre une page et un type prédéfini¹).
2. **Axiomes** : SMW permet de déclarer des relations de sous classes et de sous propriétés. Par exemple, dans la page `Category:Student`, l'annotation `[[Category:Person]]` indique que `Student` est une sous classe de `Person`. SMW permet de définir des relations d'égalité entre deux pages du Wiki (`owl:SameIndividual`) ou entre deux classes (`owl:EquivalentClass`), par l'intermédiaire du mécanisme de redirection de MW. La redirection de MW permet de dire que deux ou plusieurs pages parlent de la même chose (thématique). SMW permet également de définir des inverses de propriétés.
3. **Instances** : SMW permet à une page d'être déclarée comme une instance d'une classe ou comme sujet/objet d'un triplet RDF. Par exemple, si la page Wiki `Dakar` contient les mentions `[[Category:City]]` et `[[population::1056009]]`, cela signifie que la page `Dakar` est instanciée comme objet de la classe `City` et sa population est de 1056009 habitants.

Ainsi, la collection de données sémantiques dans SMW est faite en permettant aux utilisateurs d'ajouter des annotations sur les pages des articles à travers un système de balises spéciales. Chaque article correspond exactement à un élément ontologique c'est-à-dire à un élément de l'un des types de RDF: `owl:Thing`, `rdfs:Class`, `owl:ObjectProperty`, `owl:DatatypeProperty` et `owl:AnnotationProperty`. En particulier, un article est une instance de `owl:NamedIndividual`, les catégories deviennent des classes et les propriétés deviennent, en fonction de leur type dans le Wiki, des relations entre les articles ou des attributs pour les articles. Ainsi, la plupart des annotations dans SMW correspondent au langage OWL DL. En RDF, la modélisation des données est tripartite: **sujet - prédicat - objet** mais dans SMW la modélisation est simplifiée pour correspondre au modèle de page du Wiki (annotation faite sur une page) **page - propriété - valeur**. En effet dans l'extraction du Wiki en fichier RDF, les pages qui sont dans l'espace de noms `Category` deviennent des concepts, les pages qui sont dans l'espace de noms `Property` deviennent des propriétés et les pages qui ne sont pas dans l'espace de noms par défaut deviennent des instances (individus). Le tableau 7.1 représente une correspondance entre la syntaxe de SMW et OWL. Les trois premières lignes sont des annotations faites dans la page `Dakar` et la dernière ligne est une annotation faites dans la page `Category:City`

¹Les types prédéfinis sont disponibles sur https://www.semantic-mediawiki.org/wiki/Help:Property_declaration#List_of_datatypes

SMW	Syntaxe	OWL
Relation	[[Capital::Senegal]]	owl:ObjectProperty
Attribut	[[population::1056009]]	owl:DatatypeProperty
Catégorie (dans un article)	[[Category:City]]	owl:NamedIndividual
Catégorie (dans l'espace de noms <code>Category</code>)	[[Category:Locality]]	rdfs:subClassOf

Table 7.1: Correspondance entre SMW et OWL.

(espace de noms `Category`).

SMW permet aux utilisateurs d'importer et de réutiliser des vocabulaires existants en associant les éléments des vocabulaires avec des termes Wiki. De même, SMW permet aussi de générer dans un fichier au format RDF/OWL toutes les annotations faites par les utilisateurs.

Par défaut, SMW stocke toutes les données dans une base de données MySQL qui est celle utilisée par MediaWiki. Cela garantit une configuration simple, mais ne constitue pas une solution idéale pour les formats de données RDF. Un modèle plus adéquat est d'utiliser un format de données qui organise les informations sous forme de graphe (**triple store**) plutôt que dans une base de données avec des tables fixes. Il est possible d'utiliser de tels systèmes, en plus de la base de données MySQL pour la gestion des données annotées de SMW (voir section 7.2.2).

7.2.1.2 Exhibit

Exhibit est un outil développé durant le projet SIMILE (Semantic Interoperability of Metadata and Information in unLike Environments) qui était une collaboration entre MIT Computer Science, Artificial Intelligence Lab (CSAIL) et le World Wide Web Consortium (W3C). L'objectif du projet SIMILE a été de développer des outils pour augmenter l'interopérabilité des collections numériques disparates et de développer des outils open source solides qui permettent aux utilisateurs d'accéder, de gérer, de visualiser et de réutiliser des schémas/vocabulaires/ontologies et des métadonnées.

Exhibit fait partie de ces outils et il permet de créer des pages Web avec un support de tri, de filtrage et de visualisations riches en écrivant seulement du code HTML et éventuellement un peu de CSS et du code JavaScript. Aucune base de données n'est créée par l'utilisateur pour utiliser Exhibit. Cependant lors de son chargement, Exhibit crée une base de données légère pour charger les données. L'interface d'utilisateur d'Exhibit est composée de trois types de composants principaux:

1. **Collection** : une collection contient un sous-ensemble des éléments actuellement dans la base de données de Exhibit. Une collection peut être définie

de manière à contenir des éléments de certains types particuliers. Chaque collection a un ID, qui peut être spécifié en utilisant l'attribut `id` HTML.

2. **Widget** : il existe trois types de widgets:
 - **Facet** qui est un élément dont le but est de filtrer un ensemble de collections de données.
 - **View** qui est un composant qui rend l'ensemble des objets filtrés dans une collection à laquelle il est attaché. Comme la collection est filtrée par les facettes qui lui sont attachées, les vues attachées à cette collection se mettent à jour pour afficher l'ensemble des collections filtrées.
 - **Lens** : une lentille présente la disposition des items dans la page HTML.
3. **Coder** : Un codeur associe un élément d'information à une caractéristique visuelle. Pour connecter une vue ou une facette à un codeur, il faut spécifier à la vue ou à la facette quel codeur utiliser et quelles sont les données de la base à utiliser.

Ainsi, Exhibit propose deux types de visualisations : 1) une visualisation sur une carte Google Map, dans ce cas les données doivent être géolocalisées et 2) une visualisation sur une page HTML simple. Un autre avantage de l'utilisation de Exhibit est le fait qu'il utilise la technologie JavaScript. Ainsi, tous les calculs sont effectués localement sur le navigateur Web de l'utilisateur.

La couche IHM permet aux utilisateurs de participer à la co-construction de connaissances socioculturelles grâce à SMW qui peut aussi être utilisé pour la recherche d'information. Cependant, Exhibit permet une meilleure visualisation sur les données annotées (RDF) en proposant différents filtres et permet même une visualisation sur une carte Google Map pour les ressources géolocalisées.

7.2.2 Couche persistance

Typiquement, SMW stocke ces données dans une base de données relationnelle. Cependant pour des données RDF, la meilleure option est d'utiliser un format de données qui organise l'information sous la forme d'un graphe (**triple store**) plutôt que dans des tables de base de données relationnelle. En effet, l'utilisation d'un triple store a plusieurs avantages parmi lesquels nous pouvons citer :

- Une meilleure performance pour la recherche d'information : les bases de données RDF sont configurées pour répondre à des requêtes SPARQL qui est une recommandation du W3C. SPARQL offre une meilleure performance dans la recherche d'information que l'utilisation de requêtes SQL dans une base de données relationnelle. En effet, les requêtes SPARQL permettent de bénéficier de toute la sémantique des données RDF.

	Blazegraph	Jena TDB	Sesame	Virtuoso	4store
Modèle Supporté	Graphique et RDF stores	RDF stores	RDF stores	Graphique et RDF stores	RDF stores
Nombre de triplets stockable	12,7B	1,7B	70M	15,4B	15B
Classement en RDF store	15	3	4	2	8

Table 7.2: Comparaison entre les triplestores utilisés par SMW (M signifie million et B signifie milliard. À titre indicatif, DBpedia compte moins de 440M de triplets).

- Une interface additionnelle : avec un SPARQL Endpoint (interface de requêtes sur le Web) nous avons la possibilité de faire des requêtes SPARQL sur les données sans recourir à l'interface de SMW. Ceci permet à d'autres applications d'avoir accès aux données.
- Une fonctionnalité de raisonnement : les langages du Web sémantique comme RDFS et OWL offrent des caractéristiques de modélisation puissantes comme la possibilité de définir des relations de subsumptions entre les concepts ou les caractéristiques de certaines propriétés comme par exemple la transitivité. Ainsi, une base de données RDF peut inférer sur ces caractéristiques pour obtenir des réponses.
- la réutilisation d'ontologies et l'intégration de données : il est possible d'ajouter des données externes dans la base de données RDF et d'utiliser SMW pour les mettre à jour. Ainsi, la base de données RDF peut être vue comme une plateforme d'intégration de données et de réutilisation d'ontologies.
- La séparation physique des ressources : la séparation des données utilisées par MediaWiki de celles de SMW fournit un moyen de répartir les tâches entre plusieurs serveurs. En particulier, les requêtes complexes, qui consomment une grande quantité de puissance de calcul peuvent ainsi ne pas affecter le fonctionnement de base du Wiki.

Ainsi, SMW peut fonctionner avec cinq triple stores différents. Le tableau 7.2 fait une comparaison² entre ces triple stores. Une base de données de graphe représente les données dans des structures de graphes composés de nœuds et d'arêtes. Ceci permet un traitement facile des données par un calcul des propriétés spécifiques du graphe, telles que le nombre d'étapes (nœuds) nécessaires pour aller d'un nœud à un autre nœud. Un triple store implémente un modèle de graphe qui interprète le prédicat comme le label d'un lien entre sujet et objet.

² Comparaison extraite de <https://www.w3.org/Wiki/LargeTripleStores> et de <http://db-engines.com/en/system/Blazegraph%3BJena%3BSesame%3BVirtuoso%3B4store>.

Parmi les triples store présentés dans le tableau 7.2, notre choix s'est porté sur Virtuoso qui est une base de données hybride et un middleware qui supporte des données relationnelles, des données graphiques et RDF, des applications Web-serveur, etc.

En intégrant Virtuoso à SMW, seules les données sémantiques sont migrées dans Virtuoso. Ainsi comme le montre la figure 7.1, toutes les données textuelles restent dans la base de données MySQL qui est utilisée par MediaWiki et toutes les annotations faites dans le Wiki sont sauvegardées dans Virtuoso. Pour un meilleur passage à l'échelle, la base de données est scindée en trois entrepôts (graphes) : 1) un entrepôt où sont sauvegardées les données sémantiques (annotations issues du Wiki), 2) un entrepôt où est stockée l'ontologie USCO (voir section 5.2) et 3) un entrepôt où est stockée l'ontologie HuTO (voir chapitre 6). Puisque SMW permet d'importer des vocabulaires, les vocabulaires de HuTO et de USCO sont importés dans la plateforme. Ceci permet aux utilisateurs d'annoter les ressources du Wiki en utilisant les vocabulaires de HuTO et de USCO. En somme, les trois entrepôts forment la base de connaissances de notre plateforme ce qui nous permet avec un moteur d'inférence et une interface de requêtes d'interroger cette base de connaissances et d'extraire des informations implicites.

7.2.3 Couche sémantique

Le rôle principal de la couche sémantique est de lier la base de connaissances : 1) à Exhibit pour une meilleure visualisation, 2) aux ressources externes pour alimenter la base de connaissances et 3) aux règles définies dans la section 6.4 pour déduire des relations implicites dans la base de connaissances. Toutes ces actions sont faites à travers le framework Jena.

Jena est un framework open source pour le Web sémantique. Il fournit une API pour extraire et écrire des données dans des graphes RDF. Les graphes sont représentés comme un «modèle» abstrait. Un modèle peut provenir des données à partir de fichiers, de bases de données, des URIs ou une combinaison de ceux-ci. Un modèle peut également être interrogé par un SPARQL Endpoint.

Ainsi, pour permettre un accès aux données stockées dans Virtuoso, Virtuoso fournit un **Jena RDF Data Provider** qui est un service pour le framework Jena et qui permet aux applications Web sémantique d'interroger directement une base de connaissances Virtuoso.

Query Engine (fig. 7.1) est une interface de SPARQL Endpoint fournie par Virtuoso. Cette interface permet aux utilisateurs d'accéder directement aux données de Virtuoso. Mais dans notre configuration, cette interface n'est pas accessible aux utilisateurs de la plateforme. Cependant grâce au framework Jena, des applications Web sont développées pour accéder aux données de la plateforme.

Rules Engine (fig. 7.1) permet d'exécuter les règles définies dans la section 6.4.

Rules Engine est un composant Java EE que nous avons développé et qui exécute ces règles. Cependant, nous avons utilisé dans ces règles des clauses INSERT de SPARQL à la place des clauses CONSTRUCT utilisées dans la section 6.4.

7.3 Ontologie socioculturelle du Sénégal

L'ontologie USCO (voir section 5.2) est une ontologie socioculturelle pour le partage et la co-construction de connaissances socioculturelles. Ainsi, l'ontologie développée est une ontologie de haut-niveau c'est-à-dire une ontologie qui définit les concepts les plus généraux du domaine. Dans cette section, nous allons montrer comment l'ontologie USCO peut être utilisée dans le contexte sénégalais (ontologie de domaine) pour permettre aux communautés utilisatrices de partager et de co-construire leurs connaissances socioculturelles.

Une ontologie de domaine représente une «vision du monde» d'un domaine particulier en décrivant les concepts, les relations entre ces concepts et les instances qui sont des choses existantes dans ce domaine. Ainsi, nous allons spécifier les concepts de l'ontologie USCO dans le contexte sénégalais qui est un pays de l'Afrique de l'Ouest et une ancienne colonie française.

Les principaux événements organisés par les communautés sénégalaises sont liés aux groupes ethniques ou aux groupes religieux - plus de 98% des sénégalais sont soit musulmans (95%) soit catholiques (3%) et le pays compte à peu près 10 groupes ethniques. Dans la section 5.2.3, l'alignement entre le concept *usco:Community* de l'ontologie USCO et le concept *schema.org:Organization* du vocabulaire Schema.org avec l'opérateur `owl:equivalentClass` a été présenté. Ceci signifie que tous les sous-concepts du concept *schema.org:Organization* sont des sous-concepts du concept *usco:Community*. Cependant, le concept *schema.org:Organization* n'a pas de sous-concepts liés aux groupes ethniques et aux groupes religieux. Ainsi dans le contexte sénégalais, nous avons créé les concepts *usco:ReligiousCommunity* et *usco:EthnicCommunity* qui sont des sous concepts du concept *usco:Community*. Mis à part ces deux concepts, les autres groupes sociaux peuvent être retrouvés comme des sous concepts de *schema.org:Organization*. Par exemple, il existe le concept *schema.org:TeamsSport* qui est un sous concept de *schema.org:SportsOrganization*³ et qui représente un groupe social pratiquant du sport. Dans le contexte sénégalais, *Schema.org:TeamsSport* représente par exemple les équipes de lutte traditionnelle⁴.

Un alignement a été proposé aussi entre le concept *usco:Event* de l'ontologie USCO et le concept *schema.org:Event* (voir section 5.2). Nous avons aussi créé les concepts *usco:ReligiousEvent* et *usco:EthnicEvent* comme des sous concepts de *usco:Event*. D'autres types d'événements significatifs ont été trouvés dans le

³*schema.org:SportsOrganization* est un sous concept de *schema.org:Organization*.

⁴La lutte traditionnelle est le sport national. Elle regroupe de nombreuses équipes organisées par ethnies, villes, etc. Nous avons aussi un championnat national dont la durée est de 9 mois.

vocabulaire de Schema.org. Parmi ces sous concepts, nous avons, par exemple, *schema.org:SaleEvent* qui permet de représenter, dans le contexte sénégalais, l'événement Louma qui est un marché de vente hebdomadaire organisé principalement durant le week-end à travers le pays.

Le vocabulaire de DBpedia a été utilisé pour faire un autre alignement avec l'ontologie USCO (voir section 5.2). L'alignement a été fait avec les concepts *usco:Locality* et *usco:Infrastructure*. Le vocabulaire de DBpedia est assez riche pour que tous les concepts utilisés dans le contexte sénégalais puissent être retrouvés dans le vocabulaire de DBpedia.

L'architecture proposée permet d'utiliser des ressources externes pour alimenter notre base de connaissances. Ces ressources peuvent provenir du Linked Open Data c'est-à-dire des données ouvertes pour lesquelles une interface SPARQL Endpoint existe. Actuellement, nous utilisons la version anglaise de DBpedia pour extraire des données relatives au Sénégal. En effet, DBpedia propose une base de connaissances qui regroupe plusieurs thématiques et des données spécifiques à des pays. Ainsi, l'alignement entre les concepts *DBpedia:PopulatedPlace* et *usco:Locality* nous permet d'extraire à partir de la base de connaissances de DBpedia toutes les villes du Sénégal grâce à la requête suivante sur la base de connaissances de DBpedia :

```
PREFIX ontology: <http://dbpedia.org/ontology/>
PREFIX resource: <http://dbpedia.org/resource/>
SELECT *
WHERE { ?l ontology:type resource:Regions_of_Senegal .
        ?o ontology:isPartOf ?l }
```

Exemple 7.1 : Requête permettant d'avoir les régions du Sénégal et leurs subdivisions administratives qui sont dans la base de connaissances DBpedia⁵.

La quatrième ligne de la requête ci-dessus renvoie toutes les régions du Sénégal. Comme nous partageons avec l'ontologie DBpedia la même propriété *isPartOf* qui relie une Localité (*DBpedia:PopulatedPlace*) avec ses sous parties, la cinquième ligne donne les subdivisions des régions extraites dans la base de connaissances DBpedia. Par exemple dans la base de connaissances de DBpedia, nous avons les relations suivantes pour la Région de Dakar:

```
resource:Grand_Yoff ontology:isPartOf resource:Dakar_Departement
resource:Dakar_Departement ontology:isPartOf resource:Dakar_Region
resource:Dakar_Region ontology:isPartOf Senegal
```

Exemple 7.2 : Divisions administratives de la Région de Dakar dans DBpedia.

⁵Dans cet exemple nous avons une clause SELECT mais en pratique c'est une clause INSERT en mettant une double instanciation comme dans la figure 5.4.

Comme le concept de plus haut niveau de *DBpedia:PopulatedPlace* lié au Sénégal est *DBpedia:Region* et sa subdivision est *DBpedia:Department*, l'équivalence entre les concepts *usco:Locality* et *DBpedia:PopulatedPlace* permet de peupler certaines instances de *usco:Locality* dans notre base de connaissances. De ce fait si aucun sous concept de *DBpedia:PopulatedPlace* ne correspond à une division administrative du Sénégal, le concept est créé. Comme la base de connaissances de DBpedia est régulièrement mise à jour, ces requêtes sont exécutées de temps en temps pour avoir des informations à jour. Ainsi, les instances du concept *usco:Locality* sont partiellement mises à jour automatiquement à partir de la base de connaissances de DBpedia et manuellement par les utilisateurs qui peuvent également mettre à jour les informations extraites de la base de connaissances de DBpedia dans notre base de connaissances.

Cette façon de faire a été également utilisée pour le concept *usco:Infrastructure*. En effet, le concept *usco:Infrastructure* a été aligné avec 11 concepts de l'ontologie DBpedia. Ainsi, le concept *usco:Infrastructure* est partiellement mis à jour automatiquement à partir de DBpedia.

```
PREFIX ontology: <http://dbpedia.org/ontology/>
PREFIX resource: <http://dbpedia.org/resource/>
SELECT *
WHERE { ?1 ontology:location resource:Senegal .
        ?1 a ontology:ArchitecturalStructure}
```

Exemple 7.3 : Extraction des objets instances de *DBpedia:ArchitecturalStructure* liés au Sénégal.

La requête ci-dessus renvoie toutes les instances du concept *DBpedia:ArchitecturalStructure* de la base de connaissances de DBpedia liées au Sénégal.

La figure 7.2 représente l'ontologie USCO dans le contexte sénégalais. Ce vocabulaire est importé dans la plateforme WestAfricapedia (voir section 7.5) pour permettre aux utilisateurs d'annoter les ressources en utilisant l'ontologie Socioculturelle du Sénégal. Les concepts spécifiques du Sénégal sont en gris.

7.4 Concepts temporels de la plateforme

Puisque la plateforme est développée dans le contexte du Sénégal, les concepts temporels spécifiques au Sénégal doivent être modélisés pour permettre aux utilisateurs de pouvoir les utiliser pour annoter temporellement les ressources de la plateforme. L'ontologie HuTO permet de définir des notions temporelles qualitatives (fig. 6.3) qui sont des notions temporelles spécifiques à une culture ou à une zone géographique. Par conséquent, les concepts temporels que nous avons défini sont des notions temporelles qualitatives spécifiques à la position

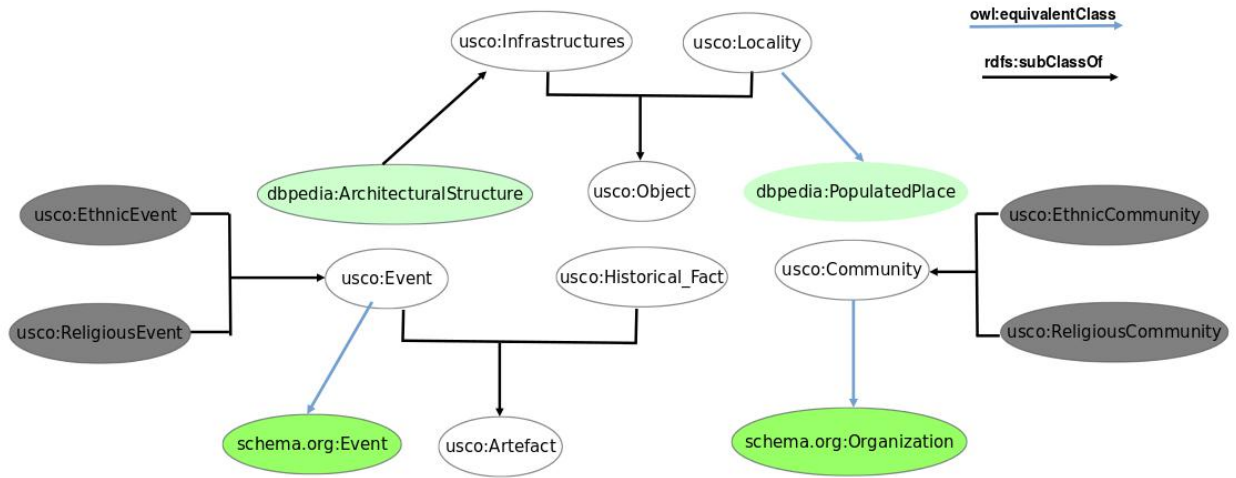


Figure 7.2: Ontologie Socioculturelle du Sénégal.

géographique du Sénégal et à la culture sénégalaise.

Le Sénégal est un pays au climat tropical composé de deux saison : une saison des pluies et une saison sèche (concepts *WetSeason* et *DrySeason* dans la figure 6.3). Cependant en fonction de la position géographique, les intervalles changent de définition. Ainsi, pour le Sénégal la saison des pluies va du mois de Juin au mois d'Octobre, le reste de l'année étant la saison sèche. De ce fait, les concepts *SenegalWetSeason* et *SenegalDrySeason* sont définis comme des sous concepts respectivement de *WetSeason* et *DrySeason* et sont annotés temporellement comme suit :

```
[a huto:TemporalAnnotation;
  huto:hasTemporalExp
    [a huto:Cycle;
      huto:every [a huto:Year];
      huto:exp [a huto:Contiguous;
        huto:hasBegin [a huto:Date;
          huto:hasMonth [a huto:June]];
        huto:hasEnd [a huto:Date;
          huto:hasMonth [a huto:October]]]]];
  huto:uri <SenegalWetSeason>].
```

```
[a huto:TemporalAnnotation;
  huto:hasTemporalExp
    [a huto:Cycle;
      huto:every [a huto:Year];
      huto:exp [a huto:Contiguous;
```

```

        huto:hasBegin [a huto:Date;
                      huto:hasMonth [a huto:November]];
        huto:hasEnd [a huto:Date;
                    huto:hasMonth [a huto:May]]];
    huto:uri <SenegalDrySeason>].

```

Exemple 7.4 : Définition des notions *SenegalWetSeason* et *SenegalDrySeason*.

Les définitions précédentes permettent d'annoter les ressources de la plateforme en utilisant des expressions comme *durant la saison des pluies*. De même, d'autres notions temporelles peuvent être définies comme *DakarNight* et être utilisées par les utilisateurs pour annoter les ressources.

Pour une utilisation plus conviviale de la plateforme, en particulier pour l'annotation temporelle des ressources, les utilisateurs peuvent donner les dates en format `xsd:dateTime` ou `xsd:duration` et une transformation est faite pour les mettre en format HuTO. Ainsi, différents scripts sont utilisés pour transformer les notations `xsd:dateTime` et `xsd:duration` au format HuTO correspondant. L'exemple 7.5 est une transformation d'une annotation temporelle répétitive (`huto:Cycle`) d'une ressource où l'heure est donnée en format `xsd:dateTime`.

```
PREFIX huto: <http://ns.inria.fr/huto/>
```

```
PREFIX property: <http://westaficapedia.sn/index.php/Special:URIResolver/Property-3A>
```

```

INSERT{ [a huto:TemporalAnnotation ;
        huto:hasTemporalExp
          [a huto:Cycle ;
          huto:every [a huto:Week] ;
          huto:exp [a huto:Contiguous ;
                  huto:hasDate [a huto:Date ;
                                huto:hasMinute [a huto:Minute;
                                                huto:value ?minute] ;
                                huto:hasHour   [a huto:Hour;
                                                huto:value ?hour] ;
                                huto:hasDay    [a ?d]]]] ;
        huto:uri ?x].
}

```

```

WHERE {   ?x property:hour      ?h ;
         property:day         ?d ;
         property:periodicite ?p .
}

```

```
FILTER (?p = "Weekly")
```

```
BIND (MINUTES(?h) as ?minute)
```

```

    BIND (HOURS(?h) as ?hour)
}

```

Exemple 7.5 : Transformation d'une annotation temporelle non convexe (cyclique) en format HuTO.

L'exemple 7.6 est un exemple de transformation d'une annotation temporelle convexe d'une ressource exprimée avec la propriété `huto:Begin`. La valeur de la propriété `property:hasBegin` est donnée en format `xsd:dateTime` et la requête ci-dessous permet de ré-écrire l'annotation temporelle en format HuTO. Cette transformation permet de regrouper également dans des graphes nommés toutes les ressources qui ont la même période d'occurrence.

```

PREFIX huto: <http://ns.inria.fr/huto/>
PREFIX usco: <http://www.westafricapedia.sn/ontology/usco/>
PREFIX property: <http://westafricapedia.sn/index.php/Special:URIResolver/Property-3A>
INSERT {
  [a huto:TemporalAnnotation ;
    huto:hasTemporalExp
      [a huto:Contiguous ;
        huto:hasBegin [a huto:Date ;
          huto:hasMinute [a huto:Minute ;
            huto:value ?minuteBegin] ;
          huto:hasHour [a huto:Hour ;
            huto:value ?hourBegin] ;
          huto:hasDay [a huto:Day ;
            huto:value ?dayBegin] ;
          huto:hasMonth [a huto:Month ;
            huto:number ?monthBegin] ;
          huto:hasYear [a huto:Year ;
            huto:value ?yearBegin ] ] ;
        huto:hasEnd [a huto:Date ;
          huto:hasMinute [a huto:Minute ;
            huto:value ?minuteEnd] ;
          huto:hasHour [a huto:Hour ;
            huto:value ?hourEnd] ;
          huto:hasDay [a huto:Day ;
            huto:value ?dayEnd] ;
          huto:hasMonth [a huto:Month ;
            huto:number ?monthEnd] ;
          huto:hasYear [a huto:Year ;
            huto:value ?yearEnd]]] ;
    huto:graph ?g].

graph ?g { ?x usco:occur ?p }

```

```

}
WHERE{
    ?x usco:occur ?p ;
        property:hasBegin ?dateTimeBegin .

    BIND (MINUTES(?dateTimeBegin ) as ?minuteBegin)
    BIND (HOURS(?dateTimeBegin ) as ?hourBegin)
    BIND (DAY(?dateTimeBegin ) as ?dayBegin)
    BIND (MONTH(?dateTimeBegin ) as ?monthBegin)
    BIND (YEAR(?dateTimeBegin ) as ?yearBegin)

    OPTIONAL { ?x huto:hasEnd ?dateTimeEnd .
        BIND (MINUTES(?dateTimeEnd ) as ?minuteEnd)
        BIND (HOURS(?dateTimeEnd ) as ?hourEnd)
        BIND (DAY(?dateTimeEnd ) as ?dayEnd)
        BIND (MONTH(?dateTimeEnd ) as ?monthEnd)
        BIND (YEAR(?dateTimeEnd ) as ?yearEnd)
    }

    BIND(IF(bound(?dateTimeEnd), uri(concat('http://www.westafricapedia.sn/named-graphs/',
        md5(concat(?dateTimeBegin, ?dateTimeEnd)), "/")),
        uri(concat('http://www.westafricapedia.sn/named-graphs/', md5(?dateTimeBegin),
        "/")))) AS ?g)
}

```

Exemple 7.6 : Transformation d'une annotation temporelle convexe en format HuTO.

7.5 Prototype

La plateforme WestAfricapedia repose principalement sur MediaWiki (MW) et Semantic MediaWiki (SMW) qui intègre un ensemble d'extensions (outils) pour une meilleure convivialité de la plateforme. Comme tout Wiki, MW permet aux utilisateurs de co-construire un corpus. Pour rendre ce corpus accessible aux agents logiciels, l'extension SMW a été développée pour permettre aux utilisateurs d'annoter ce corpus. Dans la suite de cette section, le terme SMW fait référence à la fois à MW et à ses extensions. En effet SMW ne peut pas être utilisé sans MW et les autres extensions utilisées sont assujetties à SMW ou à MW.

La plateforme a été développée sur un Laptop HP Notebook 64 bits avec un système d'exploitation Ubuntu LTS 14.

7.5.1 Outils installés

Nous présentons les outils utilisés pour développer la plateforme WestAfricapedia. Pour les détails des installations il faut se référer à l'annexe B.

Pour le développement de la plateforme WestAfricapedia, les huit outils suivants ont été installés :

1. MW : c'est le socle de la plateforme. MW est un Wiki qui permet aux utilisateurs de collaborer pour créer un corpus.
2. SMW : c'est une extension de MW. Il permet d'annoter les ressources de MW. Ainsi, son installation est assujettie à celle de MW.
3. Semantic Forms : annoter des données est une tâche difficile à accomplir pour des utilisateurs qui ne sont pas familiers avec les syntaxes à utiliser. Ainsi, l'extension **Semantic Forms** permet de définir des formulaires qui sont pré-définis en fonction des vocabulaires utilisés dans le Wiki. Les formulaires peuvent en outre être utilisés pour interroger les données du Wiki.
4. ParserFunctions : l'extension **ParserFunctions** améliore l'analyseur wikitexte avec des fonctions utiles, principalement des fonctions liées à la logique et à la manipulation de chaînes de caractères. Ainsi, **ParserFunctions** fournit onze fonctions d'analyse supplémentaires pour compléter les «mots magiques», qui sont déjà présents dans MediaWiki.
5. Variables : l'extension **Variables** permet de définir des variables dans des pages et de les réutiliser dans d'autres pages. Combinée avec l'extension **ParserFunctions**, l'extension permet de définir des fonctions qui peuvent être utilisées par d'autres pages.
6. ImageMap : l'extension **ImageMap** permet d'avoir des images cartes cliquables. Une image carte est une liste de coordonnées dans une image spécifique, où existent plusieurs zones cliquables, chacune pointant vers un lien spécifique (contrairement à un lien vers une image normale, situation dans laquelle l'ensemble de la zone pointe vers une seule destination).
7. Virtuoso : SMW utilise par défaut la base de données installée avec MW. L'utilisation du triple store Virtuoso permet de migrer les données sémantiques dans Virtuoso et de garder les données textuelles dans la base de données de MW. Virtuoso permet aussi de disposer d'un SPARQL Endpoint qui permet d'exposer ses données dans le Linked Open Data.
8. Exhibit : Exhibit permet une meilleure visualisation sur les données annotées (RDF). Ainsi, dans la plateforme une vue géolocalisée dans une carte des infrastructures et une vue sur une page HTML simple sont proposées.

7.5.2 Implémentation

La plateforme WestAfricapedia utilise les vocabulaires des ontologies USCO et HuTO. Ainsi, la seconde étape après l'installation des outils a été d'importer ces vocabulaires dans la plateforme. Pour cela, nous avons spécifié dans l'espace de noms MediaWiki les différents vocabulaires importés. Par exemple, pour le vocabulaire HuTO, en se connectant comme administrateur du Wiki, nous avons créé la page MediaWiki:Smw_import_huto dans laquelle l'URL du vocabulaire (<http://ns.inria.fr/huto/>) est précisée ainsi que toutes les annotations relatives à HuTO qui seront utilisées comme suit :

```
http://westaficapedia.sn/usco/  
Event|Category  
Community|Category  
Infrastructure|Category  
organize|Type:Page  
Nom|Type:Text
```

Si l'annotation à utiliser est une classe il faut la déclarer comme `Category`, si c'est une relation il faut la déclarer comme `Type:Page` et si c'est un attribut il faut la déclarer comme `Type:T` avec T les types usuels de SMW (`Text`, `Number`, `Url`, etc). Notons aussi que les vocabulaires de DBpedia et de Schema.org sont importés puisque l'ontologie USCO est alignée avec ces deux vocabulaires.

Une fois les vocabulaires spécifiés, il reste à définir les catégories et les propriétés des pages qui sont utilisées dans la plateforme.

Catégories et propriétés

Les catégories sont des fonctionnalités logicielles de MW qui permettent d'indexer automatiquement les pages du Wiki. Ces fonctionnalités ont été étendues par SMW pour matérialiser la notion de classe RDF. Ainsi, pour définir une catégorie il faut créer la page dans l'espace de noms `Category` et donner le nom de la catégorie. Par exemple, dans la plateforme, la page `Category:EvenementReligieux` a été créée. Pour cela, nous avons spécifié le code suivant :

```
[[Imported from::usco:ReligiousEvent]]
```

L'annotation ci-dessus dans la page `Category:EvenementReligieux` permet d'instancier toutes les pages catégorisées dans `Category:EvenementReligieux` comme instance de `http://westafricaPedia.sn/usco/ReligiousEvent`.

Les propriétés sont les méthodes de base pour ajouter des données sémantiques dans SMW. Elles peuvent être considérées comme des « catégories pour les valeurs » sur les pages d'un Wiki. Ainsi, pour définir une propriété il faut créer la page dans l'espace de noms `Property` et donner le nom de la propriété. Par exemple, dans la

plateforme, la page `Property:Partie_de` a été créée. Pour cela, nous avons spécifié le code suivant :

```
[[Imported from::usco:isPartOf]]
```

L'annotation ci-dessus dans la page `Property:Partie_de` signifie que l'utilisation de la propriété `Partie_de` dans le Wiki correspond à la propriété `http://westafricaPedia.sn/usco/isPartOf`.

WestAfricapedia est une plateforme collaborative où les utilisateurs ont la possibilité d'annoter les ressources de la plateforme. Cette annotation doit être faite grâce aux vocabulaires importés. Ceci implique que les utilisateurs doivent connaître les vocabulaires utilisés dans la plateforme. Ceci n'est pas évident pour tous les utilisateurs mais heureusement SMW propose des mécanismes, **Templates** et **Semantic Forms** (formulaires) qui permettent aux utilisateurs d'annoter les ressources sans nécessairement connaître les vocabulaires importés.

Templates et Formulaires

Les Templates sont des pages standard du Wiki dont le contenu est conçu pour être utilisé dans d'autres pages. Ainsi, pour définir un template il faut créer la page dans l'espace de noms `Template` et donner le nom du template. Par exemple, dans la plateforme, la page `Template:InfrastructureTemplate` a été créée. Pour cela, nous avons spécifié le code suivant :

```
<includeonly>
{| class="wikitable"
{{#if:{{{Nom|}}}|! Nom
{{WriteTemplate|dbpedia:name|{{{Nom|}}}}}
|-
{{#if:{{{Annee de construction|}}}|! Annee de construction
{{WriteTemplate|Annee Construction|{{{Annee de construction|}}}}}
|-
{{#if:{{{Taille|}}}|! Taille (m)
{{WriteTemplate|dbpedia:altitude|{{{Taille|}}}}}
|-
{{#if:{{{Superficie|}}}|! Superficie (m2)
{{WriteTemplate|dbpedia:area|{{{Superficie|}}}}}
|-
! Se trouve
| {{#if:{{{Se trouve|}}}|[[usco:localize::{{{Se trouve|autocomplete on
|          category=Localite}}]]}}
|-
|-{{#if:{{{Latitude|}}}|{{#if:{{{Longitude|}}}|[[LatitudeLongitude::{{{Latitude|}},
{{{Longitude|}}]]}}}}
```

```

|}
[[Category:Infrastructure]]
</includeonly>

```

Le code ci-dessus est utilisé par les pages catégorisées dans la catégorie **Infrastructure** (dernière ligne). Ce template utilise certaines propriétés du vocabulaire de DBpedia (`dbpedia:nom`, `dbpedia:area`, etc.). Il intègre aussi les coordonnées géographiques de l'infrastructure et la propriété *usco:localize* de l'ontologie USCO. Ce code est intégré dans un formulaire qui sera rempli par l'utilisateur. De ce fait, l'utilisateur n'a pas besoin de connaître les vocabulaires utilisés dans la plateforme.

L'installation de **Semantic Forms** ajoute un nouvel espace de noms pour créer des formulaires. Ainsi, pour définir un formulaire il faut créer la page dans l'espace de noms **Form** et donner le nom du formulaire. Par exemple, dans la plateforme, la page **Form:InfrastructureForm** a été créée. Pour cela, nous avons spécifié le code suivant :

```

<includeonly>
{{{for template|InfrastructureTemplate}}}
{| class="formtable"
! Nom:
| {{{field|Nom}}}
|-
! Annee de construction:
| {{{field|Annee de construction}}}
|-
! Taille (m):
| {{{field|Taille}}}
|-
! Superficie (m2):
| {{{field|Superficie}}}
|-
! Latitude:
| {{{field|Latitude}}}
|-
! Longitude:
| {{{field|Longitude|}}}
|-
! Se trouve:
| {{{field|Se trouve|autocomplete on category=Localite}}}
|}
</includeonly>

```

La première ligne de code ci-dessus, spécifie que ce formulaire est lié au template

`Template:InfrastructureTemplate`. Ainsi, les autres lignes correspondent aux paramètres du template à remplir.

Implémentation Java EE

Pour l'utilisation des données par Exhibit, des programmes Java EE ont été développés. Le modèle MVC a été adopté qui propose une solution générale au problème de la structuration d'une application. Le modèle MVC définit des règles qui déterminent dans quelle couche de l'architecture et dans quelle classe de cette couche doit être intégrée une fonctionnalité spécifique. Une application conforme à ces règles est plus facile à comprendre, à gérer et à modifier.

Dans notre implémentation nous avons utilisé la présentation de Exhibit Map et Exhibit Simple. Exhibit Map permet de placer dans une carte Google Map toutes les infrastructures géolocalisées dans la base de connaissances. Exhibit Simple affiche sur une page HTML les infrastructures de la base de connaissances avec leur photo éventuellement.

Une servlet pour la page Exhibit simple a été développée et permet d'afficher les infrastructures annotées. Elles sont classées par région et peuvent être triées par type ou par ville. Pour obtenir les informations nous avons exécuté la requête SPARQL suivante :

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX usco: <http://westaficapedia.sn/usco/>
SELECT DISTINCT ?link ?label ?city ?type ?imageUrl
WHERE {
    ?cityURI rdf:type usco:Locality .
    ?link usco:localize ?cityURI .
    ?link rdf:type/rdfs:subClassOf* ?t .
    ?t rdfs:label ?type .
    ?cityURI rdfs:label ?city .
    ?link rdfs:label ?label .
    OPTIONAL {?link usco:UrlImage ?imageUrl}
}
```

La requête ci-dessus permet d'extraire toutes les localités de la base de connaissances et leur type. Cette requête est utilisée dans un script qui extrait les données de la base de connaissances grâce au SPARQL Endpoint. Au préalable, nous devons changer le format du résultat de la requête SPARQL parce que Exhibit utilise un format JSON différent de celui des résultats d'une requête SPARQL en format JSON (fig. 7.3). Ainsi dans la figure 7.3, la ligne 4 englobe la requête ci-dessus et la ligne 6 exécute la requête dans le SPARQL Endpoint (`this.virtuosoHost` est l'URL du SPARQL Endpoint). La ligne 15, donne le résultat en format XML. Pour transformer le format XML en format JSON Exhibit, nous avons utilisé un fichier

XSLT pour avoir le bon format JSON utilisé par Exhibit (ligne 17 à ligne 23). Ainsi, une servlet est créée qui utilise ce code pour générer la page Exhibit simple.

```
1 public void generateJsonFile(String chemin, String fichierXml, String fichierJson, String fichierXsl) throws FileNk
2     String sparqlQueryString = this.query;
3
4     Query query = QueryFactory.create(sparqlQueryString);
5
6     QueryExecution qexec = QueryExecutionFactory.sparqlService(this.virtuosoHost, query);
7
8     ResultSet results = qexec.execSelect() ;
9
10    String xml = fichierXml;
11    String fhtml = fichierJson;
12
13    OutputStream out = new FileOutputStream(xml);
14
15    ResultSetFormatter.outputAsXML(out, results);
16
17    TransformerFactory tFactory = TransformerFactory.newInstance();
18    try {
19        Transformer transformer = tFactory.newTransformer(new StreamSource(fichierXsl));
20        transformer.transform(new StreamSource(xml), new StreamResult(fhtml));
21    } catch (TransformerConfigurationException ex) {
22        Logger.getLogger(UpdateTDB.class.getName()).log(Level.SEVERE, null, ex);
23    }
24
25    try {
26        out.close();
27    } catch (IOException ex) {
28        Logger.getLogger(UpdateTDB.class.getName()).log(Level.SEVERE, null, ex);
29    }
30    qexec.close() ;
31 }
```

Figure 7.3: Transformation du résultat de la requête en un format JSON Exhibit.

La ligne 9 de la figure 7.4 charge le fichier des données qui a été généré par le code de la figure 7.3. La ligne 11 charge l'API de Exhibit qui permet de traiter les données de la ligne 8. La ligne 20 montre le type de vue choisie pour représenter les données (`viewPanel`). Les lignes 22 à 27 présentent les objets à représenter dans la page HTML. Ces objets sont les attributs qui composent le résultat de la requête SPARQL. Les lignes 35 à 40 présentent les filtres utilisés dans la page HTML. Ainsi, ce code JSP est utilisé pour générer la figure 7.9.

```

4 <!DOCTYPE html>
5 <html>
6 <head>
7 <title> Exhibit Infranstures </title>
8 <!-- Fichier JSON chargé-->
9 <link href="/Exhibit/json/exhibit.json" type="application/json" rel="exhibit-data" />
10
11 <script src="http://api.simile-widgets.org/exhibit/3.0.0rc1/exhibit-api.js" type="text/javascript"></script>
12 <%@ include file="css/exhibit.jsp" %>
13 <%@ include file="css/style.jsp" %>
14 </head>
15 <body>
16 <div id="main-content">
17 <table id="frame" style="width: 100%;>
18 <tr>
19 <td id="main-content">
20 <div ex:role="viewPanel">
21 <!-- thumbnail -->
22 <div ex:role="view" ex:viewClass="Thumbnail" ex:showAll="true" ex:orders=".city"
23 ex:possibleOrders=".label, .link, .city">
24 <div ex:role="lens" class="thumbnail" style="display: none;">
25 <img ex:src-content=".imageUrl" />
26 <div><span ex:content=".label"></span></div>
27 <div>
28 <span ex:content=".city" class="city"></span>
29 </div>
30 </div>
31 </div>
32 </div>
33 </td>
34
35 <!-- Recherche -->
36 <td id="sidebar">
37 <div ex:role="facet" ex:facetClass="TextSearch" facetLabel="Search"></div>
38 <hr />
39 <div ex:role="facet" ex:expression=".city" ex:facetLabel="Ville" ex:height="15em"></div>
40 <div ex:role="facet" ex:expression=".type" ex:facetLabel="Site" ex:height="20em"></div>
41
42 <div style="text-align: center; margin: 2em;">
43 <div ex:role="logo" ex:color="Azure"></div>
44 <a href="/Exhibit/json/data.json" target="_blank">Get data</a>
45 </div>
46

```

Figure 7.4: Code de la page Exhibit Simple.

7.5.3 Maquette

Dans cette partie, nous allons présenter le résultat du développement en nous basant sur des captures d'écran. Nous allons commencer par la capture de la page d'accueil qui est l'interface d'entrée de la plateforme (fig. 7.5). La page d'accueil présente une carte du Sénégal avec ses 14 régions administratives. Chaque zone est cliquable et renvoie à la page correspondant à la région sélectionnée. Ce découpage est possible grâce à l'extension ImageMap.

La bande à gauche de la figure 7.5 représente le **Sidebar** du Wiki c'est-à-dire un espace qui permet d'accéder aux fonctionnalités majeures du Wiki. La section **Tools** permet d'accéder aux fonctions administrateur de la plateforme et les outils de navigation de MW. La section **Add data** permet aux utilisateurs de créer de nouvelles pages dans la plateforme en utilisant les formulaires définis dans la section précédente. Relativement à l'ontologie USCO, la section **Add data** permet de créer les quatre types d'entités principales utilisées dans la plateforme :

1. Création d'une communauté.
2. Création d'une infrastructure.
3. Création d'une activité.

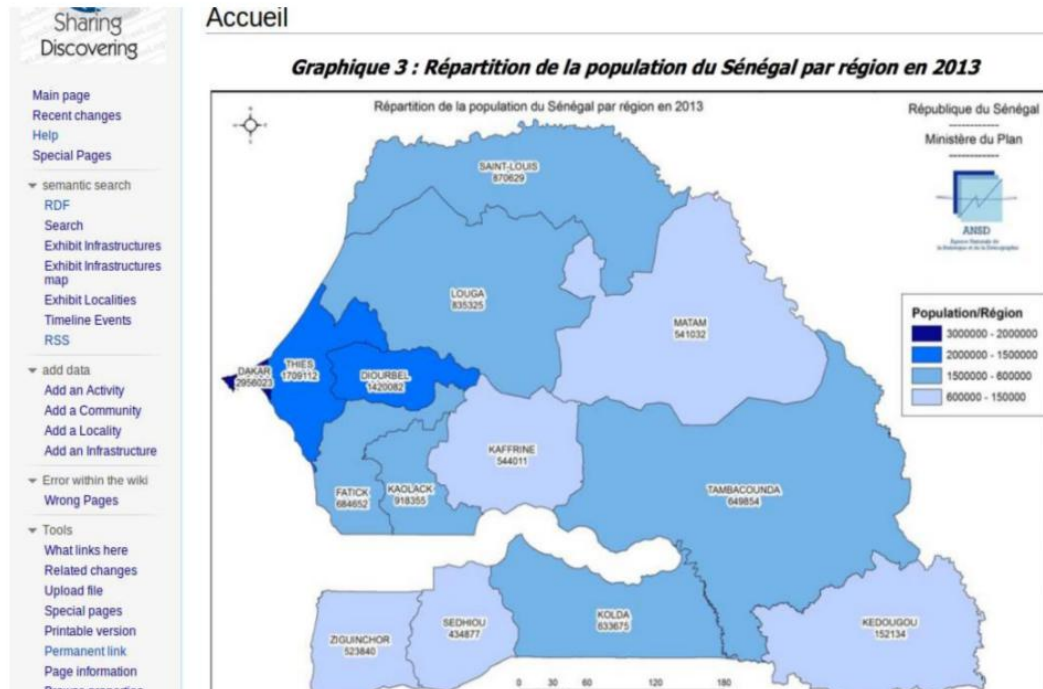


Figure 7.5: Page d'accueil de la plateforme.

4. Création d'une localité.

La section **Semantic search** permet d'utiliser les outils de Exhibit et un lien pour extraire toutes les données sémantiques dans un fichier RDF.

La figure 7.6 montre le résultat affiché suite à un clic sur la **Région de Louga**. La partie droite de la figure 7.6 montre les annotations issues du formulaire rempli pas les utilisateurs sur la page de la **Région de Louga** et la partie gauche de la figure montre un environnement comme celui proposé par Wikipedia où les utilisateurs peuvent participer à la création du corpus.

La figure 7.7 représente la page pour créer ou mettre à jour une page catégorisée comme **Infrastructure**. Le formulaire permet d'annoter la ressource. En fonction du type, d'autres annotations sont proposées. Il y a aussi la possibilité de donner plusieurs types à une infrastructure. La partie **Free Text** correspond au corpus que les utilisateurs vont créer en plus des annotations du formulaire. Grâce au formulaire, l'utilisateur n'a pas besoin de connaître les vocabulaires utilisés pour l'annotation des ressources. Ceci est fait à travers les **Templates** définis.

La figure 7.8 représente la page pour créer ou mettre à jour une page catégorisée comme **Activité**. En fonction de la périodicité de l'événement d'autres champs

- 9.1.1 L'agriculture
 - 9.1.1.1 Les atouts
 - 9.1.1.2 Les contraintes
- 9.1.2 L'élevage
 - 9.1.2.1 Les atouts
 - 9.1.2.2 Les contraintes
- 9.1.3 L'artisanat
- 10 Commerce et Services
- 11 Voies de communication et télécommunications
 - 11.1 Voies de communication
 - 11.2 Télécommunications
- 12 Le Secteur de l'industrie et PME/PMI
- 13 LES ACTEURS DU DÉVELOPPEMENT
 - 13.1 Les acteurs institutionnels
 - 13.1.1 La gouvernance locale
 - 13.1.2 Le Conseil Rural
 - 13.1.3 Le Sous-Préfet
 - 13.1.4 Conseil Régional
 - 13.1.5 ONG et projets
 - 13.1.6 Associations et Organisations Communautaire de Base

Présentation Historique

La région de Louga, autrefois province du Ndiambour, a été mise en place par la loi n° 76 61 du 26 Juin 1976 en deux entités administratives régionales distinctes : la région de **Diourbel** et de **Matam** (29 424 km² et 29 424 km²). Elle est limitée à l'Ouest par la région de Saint-Louis, à l'Est par la région de Matam et au Sud par les régions de Kolda et de Ziguinchor. Le cercle de Louga qui correspondait au NDIambour, en 1923, un autre découpage qui faisait apparaître 9 cantons (Mérinaghen, Mbaouar, Thilmakha, Guédiawaye, Diourbel, Matam, Louga, Kolda, Ziguinchor) rattachés à la subdivision de Louga et 4 autres cantons (Mbaouar, Thilmakha, Guédiawaye, Diourbel) rattachés à la subdivision de Matam. La dimension territoriale très grande, composée seulement de 6 cantons (le Djoloff, le cercle de Louga, le cercle de Matam, le cercle de Thiés, le cercle de Saint-Louis, le cercle de Kolda) vont donner lieu à la création des départements de Louga, Kébémér et Lioumassine. Les collectivités locales territorialement plus homogènes et moins étendues que les communes. Avec l'érection en 1990 de l'agglomération de Dahra en commune, la région de Louga fait partie de l'initiative 'Villes du Millénaire Africaines', récemment initiée par l'UNEP, qui comprend 7 villes d'Afrique subsaharienne, vise à identifier les opportunités économiques des villes en disséminant les résultats des recherches auprès des acteurs du développement urbain. La sélection de Louga était motivée par la présence de l'artisanat (habillement, meubles, articles en cuir, etc.), de la transformation des produits agricoles, en particulier sur la côte.

Région de Louga

Nom	Région de Louga
Nombre d'habitants	803485
Type de Localité	Région
Loi de création	loi n° 76 61 du 26 Juin 1976
Loi de modification	
taux d'urbanisation	
Localité mère	
Langues dominante	Wolof, Pular
Relief	Plat
Taux emigration	
Ethnies	
Limite ouest	
Limite est	Matam
Limite sud	Thiès
Limite nord	Saint-Louis

Contents [hide]

- 1 Présentation Historique
- 2 Présentation administrative
- 3 Aspects démographiques
 - 3.1 Caractéristiques ethniques et religieuses
 - 3.2 Mobilités des populations
 - 3.2.1 La migration intra-régionale
 - 3.2.2 La migration interrégionale
 - 3.2.3 Les migrations internationales
 - 3.3 Caractéristiques éco-géographiques
- 4 Les infrastructures sanitaires

Figure 7.6: Page de la **Région de Louga**.

sont proposés pour une annotation temporelle adéquate. Ainsi, la requête SPARQL de l'exemple 7.5 sera utilisée pour transformer l'annotation temporelle de la figure 7.8 au format HuTO.

La figure 7.9 représente la page JSP de Exhibit Simple sur les **Infrastructures**. Dans la partie gauche nous avons les items définis dans le code de la figure 7.4. Optionnellement l'image de l'infrastructure est proposée en plus du nom qui pointe sur l'URL de la page Wiki et l'année de construction. Dans la partie droite de la figure nous avons les filtres. Nous avons défini deux types de filtres : 1) un filtre par rapport aux villes (localisation) et 2) un filtre par rapport aux types. Ainsi, le clic sur un filtre affiche seulement les infrastructures relatives à ce filtre.

La figure 7.10 représente la page JSP de Exhibit Map sur les **Infrastructures**. À gauche nous avons une carte du Sénégal avec toutes les infrastructures géolocalisées qui sont dans la base de connaissances. Un clic sur un marqueur fait apparaître une fenêtre pop-up où il y a la photo de l'infrastructure optionnellement, son nom qui est un lien vers sa page Wiki et l'année de construction. Les mêmes filtres que ceux de Exhibit Simple sont proposés.

Nom: Aéroport international de Dakar-Léopold Sédar Senghor

Année de construction: 1920

Taille (m):

Superficie (m2): 8000000

Latitude: 14.740853

Longitude: -17.490406

Se trouve: Dakar

Donner un type: CentreHospitalier

Nombre de Mod: BatimentCultuel

Nombre d'Infirm: BatimentIndustriel

Nombre de Sage: BatimentReligieux

Nombre de Lits: BatimentResidentiel

Nombre de Char: CentreHospitalier

Un autre type: ServiceAdministratif

Free text:

L'Aéroport international de Dakar-Léopold Sédar Senghor ou Aéroport de Dakar est un aéroport situé à proximité du village de Yoff, commune située au nord-ouest de la capitale [[Dakar]]. Il s'agit du principal aéroport du [[Senegal|Sénégal]].

== Histoire ==
 Construit à proximité du village de Yoff, il a longtemps porté le nom de Dakar-Yoff, aujourd'hui remplacé par celui de l'ancien président du Sénégal, Léopold Sédar Senghor. Par la suite à mi respecto, les sabbats été décollados.

Sa partie sud abrite la base aérienne 160 Dakar-Quakan de l'Armée de l'Air française, qui ferma en 2011, et une base de l'armée de l'air du Sénégal. Sa superficie est 8000000 m², sa latitude est 14.740853 et sa longitude est -17.490406

Pour désorganiser Dakar, un nouvel aéroport est en construction à une quarantaine de kilomètres de la capitale, l'[[aéroport international Blaise Diagne]].

Figure 7.7: Création ou mise à jour d'une page de type **Infrastructure**.

Nom: Concert

Type d'Activite:

Se produit: Pont Faidherbe

Organise par:

Edition:

Invite:

Périodicité: Hebdomadaire

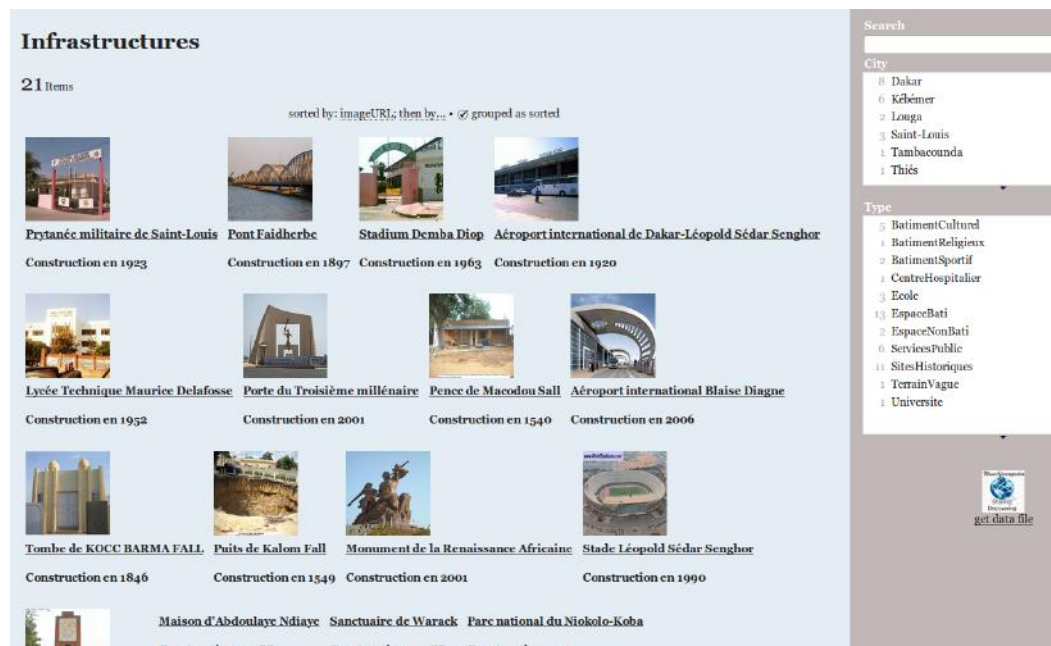
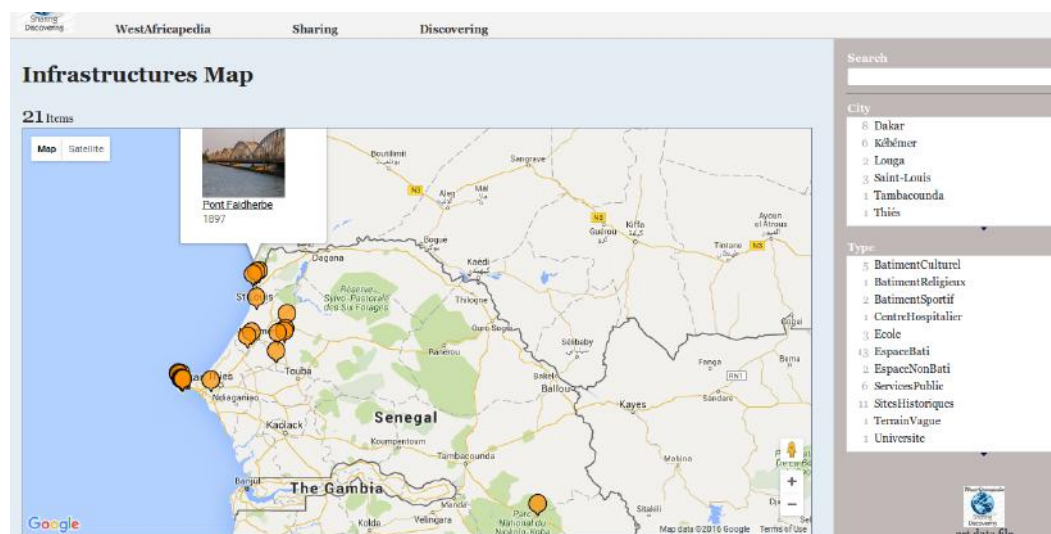
Jour de la Semaine :

Heure: Samedi

Free text:

Figure 7.8: Création ou mise à jour d'une page de type **Activité**.

Notons aussi que dans le cadre du partage et de la co-construction de connaissances, la plateforme offre différentes fonctionnalités qui sont héritées de MW. En effet, toute page créée dans le Wiki est associée à une page connexe appelée **page de discussion** qui permet aux utilisateurs d'échanger par rapport à tout changement dans la partie présentant le résultat de la collaboration. Ceci permet

Figure 7.9: Page Exhibit Simple sur les **Infrastructure**.Figure 7.10: Page Exhibit Map sur les **Infrastructure**.

une meilleure co-élaboration de connaissances. Il y a aussi l'option de l'historicité des mises à jour sur les pages. Ainsi, nous avons la possibilité de regarder spécifiquement la contribution de chaque utilisateur. Il y aussi la possibilité de restaurer les versions précédentes d'une page dans le cas de vandalisme. En effet, comme tout utilisateur a la possibilité d'ajouter ou de supprimer une entrée, cette option

est vitale pour la sauvegarde des échanges. Il y a aussi la possibilité de surveiller l'évolution des entrées d'une page. Ainsi à chaque nouvelle entrée, l'utilisateur enregistré reçoit une notification. Une autre fonctionnalité importante est le fait de protéger une page. Ainsi une fois qu'un consensus est trouvé sur le contenu d'une page, l'administrateur a la possibilité de protéger cette page qui ne pourra plus être mise à jour. Cependant, l'espace de discussion est toujours ouvert pour continuer à échanger.

En résumé, la plateforme offre deux possibilités majeures :

1. La possibilité à des utilisateurs de partager et construire leur patrimoine culturel à travers un processus de collaboration. Ce travail se fait en annotant les ressources de la plateforme : communautés, infrastructure, localité et événement. Il se fait aussi en créant un corpus par rapport aux thématiques des pages.
2. La possibilité de rechercher des informations. Cette recherche peut être sémantique en utilisant une vue sur une carte Google Map du Sénégal où sont positionnées les différentes infrastructures de la plateforme géolocalisées avec un système de filtre par rapport au type de l'infrastructure et au lieu de localisation. Elle peut aussi être textuelle en utilisant les outils de SMW.

7.6 Synthèse

Dans ce chapitre, nous avons présenté le prototype développé pour mettre en application les contributions scientifiques de cette thèse pour la mise en œuvre d'un cadre permettant à des communautés de partager et de co-construire leur patrimoine socioculturel.

En premier, nous avons présenté l'architecture générale de la plateforme. Ensuite, nous avons présenté l'architecture fonctionnelle de la plateforme composée de trois couches. Une couche IHM qui est utilisée par les utilisateurs pour participer à la co-construction de connaissances et visualiser les données de la plateforme. Cette couche repose de Semantic MediaWiki pour la co-construction et le partage de connaissances et sur Exhibit pour la visualisation des données. La deuxième couche est une couche sémantique qui permet d'extraire les données sémantiques et de les présenter aux utilisateurs. Cette couche nous permet aussi d'utiliser des données externes comme DBpedia pour alimenter la plateforme. Enfin, la couche persistance rassemble les techniques utilisées pour sauvegarder les données. Dans cette couche, nous utilisons une base de données MySQL pour sauvegarder les données textuelles et une base de connaissances Virtuoso pour sauvegarder les données sémantiques.

La deuxième partie de ce chapitre a présenté l'utilisation de l'ontologie USCO dans le contexte sénégalais. En effet, l'ontologie USCO développée dans la section 5.2 est une ontologie de haut-niveau. Dans le contexte sénégalais, quatre nouveaux concepts ont été ajoutés. Ces nouveaux concepts n'ont pas été trouvés dans les

autres ontologies alignées avec USCO. Ces quatre concepts sont *usco:EthnicEvent*, *usco:ReligiousEvent*, *usco:EthnicCommunity* et *usco:ReligiousCommunity*.

La troisième partie de ce chapitre a présenté l'ontologie HuTO (chapitre 6) dans le contexte sénégalais. Comme HuTO permet une modélisation temporelle qualitative c'est-à-dire des modélisations de notions temporelles relatives à une position géographique ou à une culture, nous avons défini des notions temporelles spécifiques au Sénégal comme *SenegalWetSeason* qui pourront être utilisées comme référence temporelle pour annoter les ressources de la plateforme.

La dernière partie de ce chapitre a présenté le prototype de la plateforme. Premièrement, nous avons présenté les différents outils utilisés à savoir Semantic MediaWiki et ses extensions, Exhibit et Virtuoso. Puis, la configuration faite pour l'importation des vocabulaires utilisés dans la plateforme (USCO, HuTO, DBpedia et Schema.org) et certains codes JAVA JEE pour l'utilisation de Exhibit ont été décrits. En dernier lieu, nous avons présenté la maquette de la plateforme à travers des captures d'écran des pages principales de la plateforme.

Conclusion générale et perspectives

Sommaire

8.1 Introduction	157
8.2 Bilan	157
8.3 Publications	161
8.4 Perspectives	162

8.1 Introduction

Dans cette thèse, nous avons abordé la problématique de la création d'un environnement de partage et de co-construction de connaissances socioculturelles. L'objectif principal a été de fournir un cadre qui permette aux communautés sénégalaises de vulgariser et de pérenniser leur culture nationale à travers un partage des coutumes et de l'histoire des différentes localités du Sénégal. Ce cadre peut être assimilé à la construction d'une plateforme qui serait l'interface entre une **mémoire d'entreprise** (ou plus exactement une **mémoire d'organisation**) et un **réseau social**, mais appliquée au contexte d'un pays. Pour une meilleure organisation et une meilleure recherche d'information, la plateforme est sous-tendue par plusieurs ontologies.

Dans cette thèse, nous avons cherché à proposer des ontologies pour permettre aux communautés de représenter leur environnement socioculturel et de pouvoir annoter les ressources partagées grâce à une collaboration des utilisateurs de la plateforme.

Dans la suite de cette section, nous dressons d'abord un bilan de nos travaux en soulignant les principales contributions que nous avons apportées, puis nous dégageons les perspectives pour nos futurs travaux.

8.2 Bilan

L'objectif premier des contributions que nous avons apportées dans cette thèse va dans le sens de permettre aux utilisateurs de la plateforme WestAfricapedia de

pourvoir collaborer et partager leur connaissances socioculturelles. Ainsi les contributions faites tournent autour de trois axes.

Proposition d'une méthodologie de construction d'une ontologie socioculturelle

La première contribution de cette thèse est la proposition d'une méthodologie de construction d'une ontologie socioculturelle : l'ontologie USCO (**Upper-level SocioCultural Ontology**). En effet, pour disposer d'un cadre de partage et de co-construction de connaissances socioculturelles il faut au préalable identifier et caractériser les différents aspects socioculturels d'un environnement. L'ontologie USCO permet d'identifier les aspects socioculturels d'un environnement et les relations qui existent entre ces derniers.

L'ontologie USCO repose sur l'extension de l'approche socioculturelle de la théorie historico-culturelle de l'Activité (**Cultural Historical Activity Theory** - CHAT) à une communauté. En effet, La CHAT est une approche psychologique pour comprendre l'évolution humaine dans le contexte culturel, institutionnel et historique de son environnement [17, 18]. Cependant, il existe deux approches de la CHAT. Une approche socioculturelle [20] qui s'articule autour d'un processus de médiation comme unité de base de l'analyse [21, 22, 23]. Une autre approche, parfois appelée théorie de l'activité, qui considère l'activité comme unité de base de l'analyse [16, 19, 24]. L'approche socioculturelle a été adoptée dans cette thèse car elle est plus adaptée au domaine de l'apprentissage et de la co-construction de connaissances grâce à sa ZPD (Zone Proximale de Développement) [28, 23, 30, 20]. Cependant, l'approche socioculturelle est centrée sur la personne et ne traite que l'action humaine. Dans cette thèse, nous nous intéressons à un cadre qui permet de modéliser les connaissances socioculturelles des communautés sénégalaises. Dans ce sens, nous nous sommes intéressés à l'action de la communauté de façon globale plutôt qu'à l'action des entités qui composent cette communauté. Ainsi, nous avons proposé une extension de l'approche socioculturelle à une communauté. Elle a servi comme méta-modèle pour modéliser les connaissances socioculturelles des communautés sénégalaises.

L'extension de l'approche socioculturelle nous a permis d'identifier trois concepts fondamentaux de l'ontologie socioculturelle USCO. Les trois concepts sont *usco:Artefact* et *usco:Object* qui sont hérités de l'approche socioculturelle et le concept *usco:Community*. Comme le processus de construction d'une ontologie est assez lourd pour conceptualiser tout un domaine, nous avons aligné notre ontologies avec des vocabulaire existants. Ainsi, à travers le Linked Open Data, les vocabulaires Schema.org et DBpedia ont été choisis. Schema.org propose une large collection de schémas pour des données structurées et DBpedia fournit une ontologie et une base de connaissances extraites de l'encyclopédie collaborative Wikipedia. Ces alignements nous ont permis de bénéficier de toutes les descriptions faites sur les objets alignés.

L'ontologie USCO est utilisée pour permettre à des communautés de partager et

de co-construire leur patrimoine socioculturel. Cela se fait à travers les descriptions faites des événements socioculturels organisés par ces communautés, les descriptions faites des ressources à la disposition de ces communautés mais également les descriptions faites sur ces communautés elles-mêmes.

L'extension de l'approche socioculturelle à une communauté fait apparaître un type de communauté virtuelle dans notre plateforme qu'il est important de positionner dans la littérature. Également la manipulation de ces données socioculturelles fait intervenir beaucoup de notions temporelles et particulièrement des notions temporelles cycliques, spécifiques à une culture et à une position géographique. Aussi, les deux autres contributions de cette thèse s'inscrivent dans le domaine du Web social pour positionner nos communautés utilisatrices et dans le domaine du Web sémantique pour l'annotation temporelle des ressources.

Communauté de co-élaboration de connaissances culturelles

La deuxième contribution de cette thèse est l'adaptation du concept de communauté de co-élaboration de connaissances (**Knowledge-building Community - KBC**) [50, 52] dans le domaine socioculturel. En effet, une KBC est un lieu où les individus sont engagés dans un processus de création, en discutant et en testant les connaissances et où leur «principal travail» est la production collaborative et l'amélioration de la connaissance et des idées [50]. Cette façon de concevoir les communautés a été identifiée dans le domaine de l'éducation [51, 52]. Nous parlons donc de communauté de co-élaboration de connaissances éducationnelles (**Educational Knowledge-building Community - EKBC**). Pour cette thèse nous avons identifié des communautés de co-élaboration de connaissances culturelles (**Cultural Knowledge-building Community - CKBC**) qui sont des KBC dans le domaine socioculturel.

Une CKBC engage les individus dans un processus social et cognitif d'appropriation, de transmission, de production et de création de connaissances caractérisant une communauté. Ce processus permet la construction de connaissances à partir de sources (ressources) sociales et culturelles et à partir de l'intégration dans des connaissances pré-existantes (connaissances des générations précédentes). Ainsi, une communauté dans la plateforme WestAfricapedia est un groupe de personnes partageant une histoire, une culture, une ethnicité ou bien un intérêt et souhaitant échanger ou collaborer grâce au Web pour partager leurs connaissances concernant ce domaine. Nos communautés sont caractérisées par trois axes : 1) un intérêt socioculturel¹ commun, 2) un échange, une collaboration et un partage entre les membres et 3) l'utilisation d'Internet pour interagir.

Puisque le domaine culturel et la connaissance culturelle ne sont pas assujettis à un groupe seulement, nous devons disposer d'une technologie qui permette à toute personne capable de participer à cette co-élaboration de pouvoir le faire. La meilleure technologie qui implémente cet aspect important est le Wiki. En

¹Aspect lié à la fois à la société et à la culture.

effet, l'idée principale d'un Wiki est que n'importe quelle page peut être consultée mais également éditée par tout utilisateur. Le moteur de Wiki qui a été choisi est MediaWiki (MW) qui est utilisé par Wikipedia. MW fournit des fonctionnalités qui facilitent une bonne co-élaboration de connaissances comme une recherche thématique facile, la facilité de contribution, la traçabilité des contributions, etc.

Les contributions dans le domaine du Web social ont été d'identifier, en premier, les CKBC dans le domaine socioculturel et de proposer 4 principes de co-élaboration de connaissances socioculturelles. Deuxièmement, nous avons aussi proposé des caractéristiques technologiques (plateforme) pour faciliter ce travail collaboratif.

HuTO : une ontologie temporelle narrative

La troisième contribution de cette thèse est la proposition de l'ontologie HuTO² (Human Time Ontology) qui est une ontologie narrative pour l'annotation temporelle des ressources du Web sémantique. L'ontologie HuTO apporte des contributions dans deux axes de recherche sur l'annotation temporelle des ressources du Web sémantique. En effet, dans l'état de l'art nous avons identifié deux axes de recherche où des travaux spécifiques ont été faits dans ce sens (voir chapitre 3) : 1) un axe de recherche que nous avons appelé **modélisation d'expressions temporelles** dont l'objectif est de proposer des formalismes qui permettent de modéliser une date, un intervalle, des notions temporelles répétitives, relatives ou absolues, etc. et 2) un axe de recherche que nous avons appelé **annotation temporelle des données** dont l'objectif est de proposer des formalismes qui permettent la représentation de notions temporelles de façon à annoter des connaissances (exprimées sous forme de triplet en RDF) et cela en conservant l'évolution des données (changement de valeur) dans le temps.

La principale contribution de l'ontologie HuTO dans le domaine de la modélisation d'expressions temporelles est la modélisation de l'expression humaine du temps. En effet, le temps humain, lié à la culture [13], permet de représenter le temps déictique qui est un temps en relation avec le temps du discours (**hier, aujourd'hui**), la modélisation qualitative du temps qui permet de modéliser des notions temporelles relatives à la culture ou à la position géographique et la possibilité d'utiliser une ressource comme un marqueur temporel. L'ontologie HuTO permet aussi d'étendre ses unités temporelles pour intégrer d'autres unités temporelles comme celles du calendrier Maya ou Aztèque.

La principale contribution de l'ontologie HuTO dans le domaine de l'annotation temporelle des données est de garder la sémantique des triplets RDF tout en ajoutant une dimension temporelle et de ne pas étendre les langages du Web sémantique. Ceci nous a permis d'associer une dimension temporelle aux connaissances du domaine. Cette approche permet de garder l'évolution temporelle des données que ce soit une ressource, un triplet ou un ensemble de triplets et facilite également

²<http://ns.inria.fr/huto>

la recherche d'informations qu'elle soit temporelle ou non.

Hormis ces contributions, nous avons intégré dans cette thèse pour l'utilisation de l'ontologie HuTO des requêtes types et des transformations RDF2Text qui permettent de mieux représenter les résultats des requêtes. Ces requêtes types paramétrées permettent de déterminer : 1) les ressources répétitives par rapport à une fréquence donnée, 2) les ressources qui se produisent relativement (avant ou après) à une ressource donnée, 3) les ressources qui se produisent à une date donnée, 4) la date d'occurrence d'une ressource spécifique, 5) les ressources entre deux intervalles et 5) les ressources sur un nombre de jours donnés.

Les transformations RDF2Text utilisent le langage STTL [96, 97] et permettent un rendu plus lisible et compréhensible des requêtes SPARQL sur des données annotées grâce à HuTO.

Une validation du système

Nous avons démontré la viabilité des contributions de cette thèse, en les inscrivant dans une architecture logicielle concrète. Il s'agit plus précisément d'une plateforme de partage et de co-construction de connaissances. Le prototype que nous avons présenté dans le chapitre 7 implémente, en respectant scrupuleusement les principes théoriques que nous avons proposés dans cette thèse, (i) l'utilisation de l'ontologie USCO pour permettre à des communautés de partager et de construire leur patrimoine culturel à travers l'annotation des différentes ressources et des activités qui composent leur environnement, (ii) la création d'un environnement où les membres des communautés de co-élaboration de connaissances culturelles (CKBC) peuvent échanger pour participer au processus de création de connaissances collectives et (iii) l'utilisation de l'ontologie HuTO pour annoter temporellement les ressources de la plateforme bien que cette annotation soit faite en arrière plan (pas par les utilisateurs directement mais par des scripts).

8.3 Publications

La liste suivante énumère les publications concernant le travail que nous présentons dans cette thèse :

1. **Diallo P F**, Corby O, Mirbel I, Lo M, Ndiaye S (2016) Ontologies-based Platform for Sociocultural Knowledge Management. In Journal on Data Semantics (65). Spring DOI : 10.1007/s13740-016-0065-4.
2. **Diallo P F**, Corby O, Mirbel I, Lo M, Ndiaye S (2015) HuTO: une Ontologie Temporelle Narrative pour les Applications du Web Sémantique. In acte conférence d'Ingénierie des Connaissances 2015 du 29 Juin au 03 Juillet à Rennes - France. Disponible sur http://arxiv.org/find/cs/1/au:+Diallo_P/0/1/0/all/0/1

3. **Diallo P F**, Corby O, Lo M, Mirbel I, Ndiaye S (2014) Sociocultural Ontology: Upper-level and Domain Ontologies. in Actes JFO, Hammamet, Tunisie Novembre 2014, pp 15-27. <https://hal.inria.fr/hal-0107975>

Notons que la méthodologie présentée dans cette thèse pour la construction d'une ontologie socioculturelle a été utilisée dans le domaine Camerounais (Afrique Centrale). Dans ce contexte, l'approche Leont'ev-Engeström de la CHAT a été utilisée car notre objectif a été de modéliser le dynamisme au sein des communautés. Cependant, le concept **usco:Infrastructure** a été réutilisé de même que la technique d'alignement utilisé dans cette thèse. Ces travaux ont permis les publications suivantes :

1. Kaladzavi G, **Diallo P F**, Kolyang and Lo M (2015) Virtual Activity Ontology Modeling : the Case of Sociocultural Knowledge Sharing. In proceedings of the fifth international conference on social media technologies, communication, and informatics SOTICS, barcelona , spain isbn: 978-1-61208-443-5.
2. Kaladzavi G, **Diallo P F**, Kolyang and Lo M (2015) OntoSoc: Sociocultural Knowledge Ontology. In International Journal of Web Semantic Technology (ijwest). Vol .6, n o .2, april 2015. <http://irccj.org/ijwest/ijwest2015.html>

8.4 Perspectives

Les recherches que nous avons menées dans le cadre de cette thèse peuvent être approfondies afin d'améliorer nos approches. Ces perspectives peuvent être étalées dans le court et le long terme.

Court terme

Nos perspectives à court terme sont essentiellement orientées sur les besoins de la plateforme. Tout d'abord, nous pensons qu'il serait intéressant d'étudier comment le Linked Open Data pourrait encore nous aider à automatiser partiellement la mise à jour de certaines ressources. En l'état actuel, seul DBpedia est utilisé pour cette tâche. L'idée est de voir comment utiliser d'autres ressources comme Wikidata et YAGO par exemple comme sources de données. Le travail de recherche serait de voir comment l'ontologie USCO pourrait être alignée ou évoluer pour intégrer ces ressources.

Le deuxième travail à court terme serait l'intégration d'un serveur Corese [95] dans la plateforme. En effet, nous avons développé dans cette thèse des transformations RDF2Text pour les requêtes temporelles en SPARQL. Ces transformations utilisent le langage STTL [96, 97] qui est intégré dans Corese. Ainsi, l'intégration d'un serveur Corese permettra de visualiser dans la plateforme les transformations RDF2Text développées.

Long terme

À plus long terme, nous envisageons d'étudier les autres relations d'Allen [57, 58] hormis les relations *after* et *before* qui sont déjà intégrées dans HuTO. Le principal travail serait de voir comment les caractéristiques logiques de ces 11 relations pourraient être gérées par des règles comme celles définies dans la section 6.4.

Un autre axe de recherche serait d'intégrer dans HuTO la modélisation de l'incertitude dans les notions temporelles comme dans l'expression *à la fin des années 1980*. En effet, dans le domaine culturel les dates de certains événements ne sont pas connues de façon exacte mais sont données de façon approximative. Le principal travail n'est pas seulement d'intégrer ces notations dans HuTO mais aussi de pouvoir raisonner dessus afin que ces dates puissent être aussi dans les résultats des requêtes temporelles.

Un autre axe de recherche est de voir comment les téléphones mobiles et les smartphones pourraient être utilisés pour participer à la co-construction de connaissances dans la plateforme (*access mobile for web of data*). L'idée est de voir comment nous pouvons explorer les spécificités (avantages et contraintes) de partage de connaissances avec les mobiles. L'application mobile ciblée devra permettre aux communautés et à leurs utilisateurs d'enrichir et d'accéder plus facilement la base de connaissances en utilisant le contexte de l'utilisateur (par exemple la géolocalisation, les caractéristiques matérielles, etc) et de traiter ses limitations (accès par exemple intermittent, ressources limitées).

Annexe : HuTO

A.1 Concepts généraux de HuTO

La figure A.1 donne une vue générale des principaux concepts de HuTO.

A.2 Règles

Actuellement, nous avons défini 50 règles de normalisations et de raisonnement et deux règles de corrections. Nous présentons ici seulement quelques règles mais les règles complètes sont disponible sur <http://ns.inria.fr/huto/rules/>. Notons aussi que toutes les règles sont définies que pour un calendrier grégorien.

Ajoute le nombre de jours du mois, le numéro du mois et la parité du mois

```
PREFIX huto: <http://ns.inria.fr/huto/>
CONSTRUCT{ ?x huto:number ?m ;
            huto:numberOfDay ?d ;
            huto:even ?e }
WHERE{
  ?x rdf:type ?o
  ?o rdfs:subClassOf huto:Month ;
    huto:number ?m ;
    huto:even ?e
  OPTIONAL{?x rdf:type/huto:numberOfDay ?d}
}
```

Ajoute les mois civils (les mois grégoriens)

```
PREFIX huto: <http://ns.inria.fr/huto/>
CONSTRUCT{ ?x a ?c }
WHERE{ ?x huto:number ?m ;
        rdf:type huto:Month .
  ?c huto:number ?m ;
        rdfs:subClassOf huto:Month .
}
```

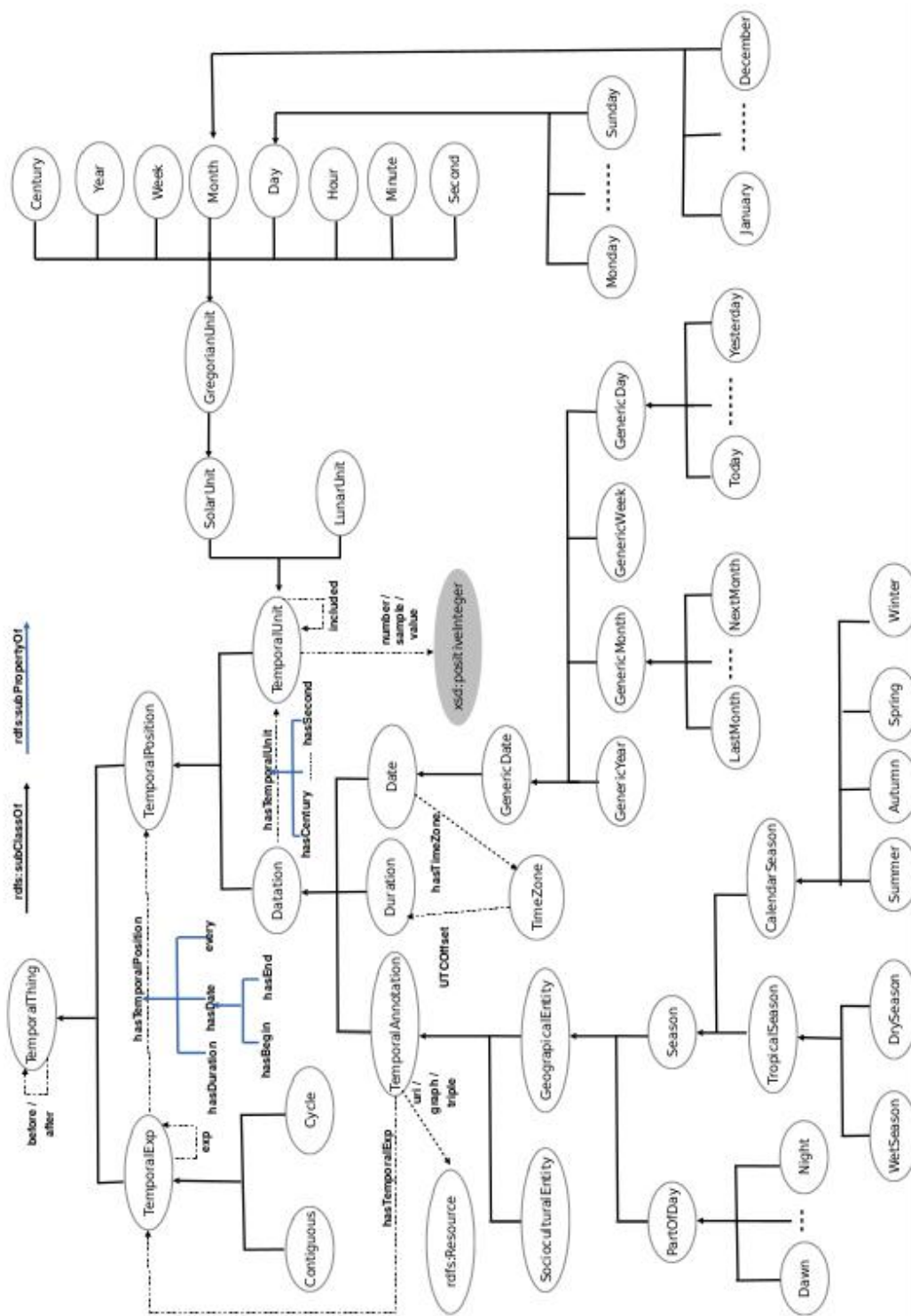


Figure A.1: Principaux concepts de HuTO.

Les genres de règles précédentes permettent de déterminer le nombre de jours pour le mois, la parité des mois et des années. Toutes ces règles permettent d'ajouter des informations appropriées qui sont nécessaires pour répondre à des questions comme *quelles les ressources qui se produisent durant les mois impairs, quelles ressources se produisent entre le 22/01/2015 et le 02/02/2015*, etc.

les propriétés after et before sont inverses

```
PREFIX huto: <http://ns.inria.fr/huto/>
CONSTRUCT{ ?y huto:before ?x}
WHERE{ ?x huto:after ?y }
```

la propriétés after est transitive (règle de propagation)

```
PREFIX huto: <http://ns.inria.fr/huto/>
CONSTRUCT{ ?x huto:after ?z }
WHERE{ ?x huto:after+ ?y .
       ?y huto:after+ ?z }
```

Si la relation d'Allen est définie entre des concepts *TemporalAnnotation* alors cette règles ajoute la même information aux concepts *TemporalExp* concernés (cas de la propriété after).

```
PREFIX huto: <http://ns.inria.fr/huto/>
CONSTRUCT{ ?s huto:after ?o }
WHERE{
  [a huto:TemporalAnnotation ;
   (huto:uri|huto:triple|huto:graph) ?p1 ;
   huto:hasTemporalExp ?s]
  huto:after
  [a huto:TemporalAnnotation ;
   (huto:uri|huto:triple|huto:graph) ?p2 ;
   huto:hasTemporalExp ?o]
}
```

Si la relation d'Allen est définie entre des concepts *TemporalExp* alors cette règles ajoute la même information aux concepts *TemporalAnnotation* concernés (cas de la propriété after).

```
PREFIX huto: <http://ns.inria.fr/huto/>
CONSTRUCT{ ?x huto:after ?y }
```

```

WHERE{
  ?s huto:after ?o

  ?x a huto:TemporalAnnotation ;
    (huto:uri|huto:triple|huto:graph) ?p1 ;
    huto:hasTemporalExp ?s .

  ?y a huto:TemporalAnnotation ;
    (huto:uri|huto:triple|huto:graph) ?p2 ;
    huto:hasTemporalExp ?o .
}

```

Des règles équivalentes ont été définies pour la propriété *before* qui a les mêmes caractéristiques que la propriété *after*.

Les caractéristiques logiques des relations de Allen. Les Annotations faites avec les propriétés *hasBegin* et *hasEnd* (cas de la propriété *after*).

En ce qui concerne les caractéristiques logiques des relations Allen, nous vérifions les ressources qui les satisfont. Les deux règles ci-dessous vérifient les ressources qui satisfont deux à deux les caractéristiques logiques de la propriété *after*. Les règles équivalentes sont fournies pour la propriété *before* qui a les mêmes caractéristiques que la propriété *after*.

```

PREFIX huto: <http://ns.inria.fr/huto/>
CONSTRUCT{ ?s huto:after ?o }
WHERE{
  ?s huto:hasTemporalExp/huto:hasEnd ?te .
  ?o huto:hasTemporalExp/huto:hasBegin ?tb .

  ?te huto:hasYear/huto:value ?y1 .
  ?tb huto:hasYear/huto:value ?y2 .

  FILTER(?y1 > ?y2 ||
    (?y1 = ?y2 ||
      EXISTS{
        ?te huto:hasMonth/huto:number ?m1 .
        ?tb huto:hasMonth/huto:number ?m2 .

        FILTER(?m1 > ?m2 ||
          (?m1 = ?m2 ||
            EXISTS{
              ?te huto:hasDay/huto:value ?d1 .

```

```

        ?tb huto:hasDay/huto:value ?d2 .
        FILTER(?d1 > ?d2)
    })
)
})
}

```

Les caractéristiques logiques des relations de Allen. Le ressources exprimées avec la propriété *hasDate* (cas de *after*).

```

PREFIX huto: <http://ns.inria.fr/huto/>
CONSTRUCT{ ?s huto:after ?o }
WHERE{
    ?s huto:hasTemporalExp/huto:hasDate ?t1 .
    ?o huto:hasTemporalExp/huto:hasDate ?t2 .

    ?t1 huto:hasYear/huto:value ?y1 .
    ?t2 huto:hasYear/huto:value ?y2 .

    FILTER(?y1 > ?y2 ||
        (?y1 = ?y2 ||
            EXISTS{
                ?t1 huto:hasMonth/huto:number ?m1 .
                ?t2 huto:hasMonth/huto:number ?m2 .

                FILTER(?m1 > ?m2 ||
                    (?m1 = ?m2 ||
                        EXISTS{
                            ?t1 huto:hasDay/huto:value ?d1 .
                            ?t2 huto:hasDay/huto:value ?d2 .
                            FILTER(?d1 > ?d2)
                        })))
            })))
}

```

Les caractéristiques logiques des relations de Allen. Appliquer aux ressources qui ont des dates ordonnées (case of *after*).

```

PREFIX huto: <http://ns.inria.fr/huto/>
CONSTRUCT{ ?resource1 huto:after ?resource2 }
WHERE{
    ?s1 huto:hasTemporalExp ?date1 ;
        (huto:uri|huto:graph|huto:triple) ?resource1 .

    ?s2 huto:hasTemporalExp ?date2 ;

```

```
(huto:uri|huto:graph|huto:triple) ?resource2 .

?date1 huto:after ?date2
}
```

Normalisation des intervalles.

La règles ci-dessous permet le premier ou le dernier jour du mois. Par exemple, quand on utilise l'expression **début janvier** sans préciser le jour, cela signifie le **1er Janvier** ou lorsqu'on utilise **fin en Janvier** sans préciser le jour, cela signifie le **31 Janvier**.

```
PREFIX huto: <http://ns.inria.fr/huto/>
CONSTRUCT{ ?e huto:value ?d }
WHERE{
  ?x huto:hasMonth ?e
  ?e rdf:type ?o ;
    huto:numberOfDay ?n .
  ?o rdfs:subClassOf+ huto:Month .

  BIND(IF(EXISTS{?s huto:hasBegin ?x}, 1,
    IF(EXISTS{?s huto:hasEnd ?x}, ?n, 0))
    as ?d)

  FILTER NOT EXISTS{?x huto:hasDay ?p}
  FILTER(EXISTS{?s huto:hasBegin ?x} ||
    EXISTS{?s huto:hasEnd ?x})
}
```

Dans notre approche de modélisation, nous considérons toute entité comme un intervalle. Ainsi, la règle précédente normalise à la fois les intervalles exprimés avec la propriété *huto:hasBegin* et la propriété *huto:hasEnd*.

Normalisation des intervalles exprimés par la relation *huto:hasDuration*

Les deux règles suivantes sont utilisées pour normaliser les intervalles exprimés avec les propriétés *huto:hasBegin*. Les règles ajoutent la date de fin appropriées. Notons qu'il existe six règles pour ajouter la date de fin car chaque règle correspond à un format spécifique de la propriété *huto:hasDuration* .

Notons que des règles équivalentes existent pour les intervalles exprimés par une durée avec la propriété *huto:hasEnd* .

```
PREFIX huto: <http://ns.inria.fr/huto/>
```

```

CONSTRUCT{
  ?x huto:hasEnd
    [ ?j ?l ;
      huto:hasDay [a huto:Day ;
                    huto:value ?d3] ;
      huto:hasMonth [a huto:Month ;
                     huto:number ?m2] ;
      huto:hasYear [a huto:Year ;
                    huto:value ?y1]]}

Where{
  ?x huto:hasBegin ?p ;
    huto:hasDuration/huto:hasDay/huto:number ?d.
  ?p huto:hasMonth/huto:number ?m ;
    huto:hasMonth/huto:numberOfDay ?nb ;
    huto:hasDay/huto:value ?d1 ;
    huto:hasYear/huto:value ?y ;
    ?j ?l .

  BIND(?d + ?d1 as ?d2)
  BIND(IF(?d2 > ?nb, ?d2 - ?nb, ?d2) as ?d3)
  BIND(IF(?d2 > ?nb, ?m + 1, ?m) as ?m3)
  BIND(IF(?m3 > 12, ?m3 - 12, ?m3) as ?m2)
  BIND(IF(?m > 12, ?y + 1, ?y) as ?y1)

  FILTER(?j != huto:hasMonth || ?j != huto:hasDay
         || ?j != huto:hasYear)
  FILTER NOT EXISTS{?x huto:hasEnd ?h}
}

```

```

PREFIX huto: <http://ns.inria.fr/huto/>
CONSTRUCT{
  ?x huto:hasEnd
    [ ?j ?l ;
      huto:hasDay [a huto:Day ;
                    huto:value ?d3] ;
      huto:hasMonth [a huto:Month ;
                     huto:number ?m2]]}

WHERE{
  ?x huto:hasBegin ?p ;
    huto:hasDuration/huto:hasDay/huto:number ?d1.
  ?p huto:hasMonth/huto:number ?m ;
    huto:hasMonth/huto:numberOfDay ?nb ;
    huto:hasDay/huto:value ?d ;

```

```

    ?j ?l .

    BIND(?d1 + ?d as ?d2)
    BIND(IF(?d2 > ?nb, ?d2 - ?nb, ?d2) as ?d3)
    BIND(IF(?d2 > ?nb, ?m + 1, ?m) as ?m2)

    FILTER(?j!=huto:hasMonth || ?j!=huto:hasDay)
    FILTER NOT EXISTS{?p huto:hasYear ?k}
    FILTER NOT EXISTS{?x huto:hasEnd ?h}
}

```

Requêtes Corrections

Les requêtes ci-dessous vérifient la cohérence entre les concepts *Contiguous* et *Cycle*.

```

PREFIX huto: <http://ns.inria.fr/huto/>
DELETE{ ?p ?o ?k
        ?m ?f ?j
      }
INSERT{ ?p ?f ?j
        ?m ?o ?k
      }
WHERE{
    ?x a huto:Cycle ;
        huto:every ?p ;
        huto:exp/huto:hasDate ?m .

    ?p rdf:type ?y ;
        ?o ?k .

    ?m ?f ?j .
    ?j rdf:type ?z .

    ?y huto:included ?t .
    ?z rdfs:subClassOf* ?t .

    FILTER(?y != ?t)
}

```

```

PREFIX huto: <http://ns.inria.fr/huto/>
DELETE { ?p ?o ?k
        ?m ?f ?j }
INSERT { ?p ?f ?j
        ?m ?o ?k }

```

```

WHERE {
    ?x a huto:Contiguous ;
        huto:hasDate ?p ;
        huto:exp/huto:every ?m .

    ?p ?o ?k .
    ?k rdf:type ?y .

    ?m rdf:type ?z ;
        ?f ?j .

    ?y huto:included ?t .
    ?z rdfs:subClassOf* ?t .

    FILTER (?y != ?t)
}

```

A.3 Requêtes Types

Dans la sous section 6.5.2, nous avons présenté nos requêtes-types sur les éléments temporels. Ici nous allons présenter les sept requêtes-types :

- 1) Les ressources qui sont vraies pendant un intervalle donné.

```

PREFIX huto: <http://ns.inria.fr/huto/>.
PREFIX data: <http://example.org/data/>
DESCRIBE ?x
where {
    {# Ponctual resources expressed
    ?x huto:hasTemporalExp/huto:hasDate
        [ huto:hasYear/huto:value ?year ;
          huto:hasMonth/huto:number ?month;
          huto:hasDay/huto:value ?day]

    FILTER( (?yearBeg < ?year && ?year < ?yearEnd) ||
            (?yearBeg = ?year && ?year != ?yearEnd &&
             (?monthBeg < ?month || (?monthBeg = ?month && ?dayBeg <= ?day))) ||
            (?yearBeg = ?year && ?year = ?yearEnd &&
             ((?monthBeg < ?month && ?month < ?monthEnd) ||
              (?monthBeg = ?month && ?dayBeg <= ?day) ||
              (?month = ?monthEnd && ?day <= ?dayEnd))) ||
            (?year = ?yearEnd && (?month = ?monthEnd && ?day <= ?dayEnd)))
    }
UNION

```

```

    {# Punctual resources expressed with hasBegin and hasEnd properties
    ?x huto:hasTemporalExp/huto:hasBegin
      [ huto:hasYear/huto:value ?y1 ;
        huto:hasMonth/huto:number ?m1;
        huto:hasDay/huto:value ?d1];
      huto:hasTemporalExp/huto:hasEnd
    [ huto:hasYear/huto:value ?y2 ;
      huto:hasMonth/huto:number ?m2;
      huto:hasDay/huto:value ?d2]

FILTER((?yearBeg < ?year && ?year < ?yearEnd) ||
  (?yearBeg = ?year && ?year != ?yearEnd &&
(?monthBeg < ?month || (?monthBeg = ?month && ?dayBeg <= ?day))) ||
  (?yearBeg = ?year && ?year = ?yearEnd &&
  ((?monthBeg < ?month && ?month < ?monthEnd) ||
  (?monthBeg = ?month && ?dayBeg <= ?day) ||
  (?month = ?monthEnd && ?day <= ?dayEnd))) ||
  (?year = ?yearEnd && ((?month < ?monthEnd) ||
  (?month = ?monthEnd && ?day <= ?dayEnd))))
}
UNION
{# Weekly resources
  ?x huto:hasTemporalExp/rdf:type huto:Cycle;
  huto:hasTemporalExp/huto:every/rdf:type huto:Week
  }
UNION
{# Monthly resources
  { ?x huto:hasTemporalExp/huto:exp/huto:hasDate/huto:hasDay/huto:value ?day
    FILTER((?dayBeg <= ?day && ?monthBeg < ?monthEnd) ||
      (?dayBeg <= ?day && ?day <= ?dayEnd))
  }
  UNION
  {# Monthly resources expressed with hasBegin and hasEnd properties
    ?x huto:exp/huto:hasBegin/huto:hasDay/huto:value ?d1;
    huto:exp/huto:hasEnd/huto:hasDay/huto:value ?d2
    FILTER(?dayBeg <= ?d1 && ?d2 <= ?dayEnd)
  }
  ?x huto:hasTemporalExp/rdf:type huto:Cycle;
  huto:hasTemporalExp/huto:every/rdf:type huto:Month
  }
UNION
  {# Yearly resources
  ?x huto:hasTemporalExp/rdf:type huto:Cycle ;
    huto:hasTemporalExp/huto:every/rdf:type huto:Year;

```

```

huto:hasTemporalExp/huto:exp/huto:hasDate/huto:hasMonth/huto:number ?month;
huto:hasTemporalExp/huto:exp/huto:hasDate/huto:hasDay/huto:value ?day

FILTER((?monthBeg < ?month && ?month < ?monthEnd) ||
        (?monthBeg = ?month && ?day <= ?dayEnd) ||
(?monthBeg = ?month && ?monthBeg = ?monthEnd &&
  ?dayBeg <= ?day && ?day <= ?dayEnd) ||
  (?monthEnd = ?month && ?day <= ?dayEnd))
}
UNION
  {# Yearly resources expressed with hasBegin and hasEnd properties
  ?x huto:hasTemporalExp/rdf:type huto:Cycle ;
  huto:hasTemporalExp/huto:every/rdf:type huto:Year;
  huto:hasTemporalExp/huto:exp/huto:hasBegin/huto:hasMonth/huto:number ?m1;
  huto:hasTemporalExp/huto:exp/huto:hasBegin/huto:hasDay/huto:value ?d1;
  huto:hasTemporalExp/huto:exp/huto:hasEnd/huto:hasMonth/huto:number ?m2;
  huto:hasTemporalExp/huto:exp/huto:hasEnd/huto:hasDay/huto:value ?d2

  FILTER((?monthBeg < ?m1 && ?m2 < ?monthEnd) ||
          (?monthBeg = ?m1 && ?dayBeg <= ?d1 &&
            (?m2 < ?monthEnd || (?m2 = ?monthEnd && ?d2 <= ?dayEnd))))
  }
}

```

2) Les ressources qui sont vraies sur une période donnée.

```

PREFIX huto: <http://ns.inria.fr/huto/>.
PREFIX data: <http://example.org/data/>
TEMPLATE {st:apply-templates-with(st:turtle, ?x)}
where {
  {# Punctual resources
  ?x huto:hasTemporalExp/huto:hasDate [ huto:hasYear/huto:value ?year ;
                                       huto:hasMonth/huto:number ?month ;
                                       huto:hasDay/huto:value ?day] .

  FILTER(?day <=?maxDay)
  }
UNION
  {# Punctual resources expressed with hasBegin and hasEnd properties
  ?x huto:hasTemporalExp/huto:hasBegin
    [ huto:hasYear/huto:value ?y1 ;
      huto:hasMonth/huto:number ?m1 ;
      huto:hasDay/huto:value ?d1] ;
    huto:hasTemporalExp/huto:hasEnd
    [ huto:hasYear/huto:value ?y2 ;

```

```

        huto:hasMonth/huto:number ?m2] .

    FILTER((?y1 = ?year && ?year < ?y2 && (?m1 < ?month ||
        (?m1 = ?month && ?d1 <= ?maxDay))) ||
        (?y1 = ?y2 && ?y1 = ?year && ((?m1 = ?month && ?d1 <= ?maxDay) ||
        (?m1 < ?month && ?month <= ?m2 ))) ||
        (?y1 < ?year && (?year < ?y2 || (?year = ?y2 && ?month <= ?m2 ))))
    }
UNION
    {# Weekly resources
        ?x huto:hasTemporalExp/rdf:type huto:Cycle ;
        huto:hasTemporalExp/huto:every/rdf:type huto:Week .
    }
UNION
    {# Monthly resources
        { ?x huto:hasTemporalExp/huto:exp/huto:hasDate/huto:hasDay/huto:value ?day .
        FILTER(?day <= ?maxDay)
        }
    UNION
        {# Expressed with hasBegin and hasEnd properties
            ?x huto:hasTemporalExp/huto:exp/huto:hasBegin/huto:hasDay/huto:value ?d1 ;
            huto:hasTemporalExp/huto:exp/huto:hasEnd/huto:hasDay/huto:value ?d2 .
            FILTER(?d1 <= ?maxDay)
        }
    }

    ?x huto:hasTemporalExp/rdf:type huto:Cycle ;
    huto:hasTemporalExp/huto:every/rdf:type huto:Month

    # To calculate the parity of ?month
    BIND((?month / 2) as ?e1)
    BIND(ROUND(?e1) = ?e1 as ?even)

    # To handle expressions like every even month (there is an implicit exception)
    FILTER NOT EXISTS{ ?x huto:hasTemporalExp/huto:every/huto:even ?mEven
        FILTER (?even != ?mEven)}
    }
UNION
    {# Yearly resources
        { ?x huto:hasTemporalExp/huto:exp/huto:hasDate/huto:hasMonth/huto:number ?month .
        OPTIONAL{?x huto:hasTemporalExp/huto:exp/huto:hasDate/huto:hasDay/huto:value ?day}
        FILTER(!BOUND(?day) || ?day <= ?maxDay)
        }
    UNION
        {# Yearly resources expressed with hasBegin and hasEnd properties

```

```

    ?x huto:hasTemporalExp/huto:exp/huto:hasBegin/huto:hasMonth/huto:number ?m1 ;
        huto:hasTemporalExp/huto:exp/huto:hasBegin/huto:hasDay/huto:value ?d1 ;
        huto:hasTemporalExp/huto:exp/huto:hasEnd/huto:hasMonth/huto:number ?m2 .
    FILTER((?m1 < ?month && ?month <= ?m2) || (?m1 = ?month && ?d1 <= ?maxDay))
}

?x huto:hasTemporalExp/rdf:type huto:Cycle ;
    huto:hasTemporalExp/huto:every/rdf:type huto:Year;

# To calculate the parity of ?year
BIND((?year / 2) as ?e1)
BIND(ROUND(?e1) = ?e1 as ?even)

# To handle expressions like every even year (there is an implicit exception)
FILTER NOT EXISTS { ?x huto:hasTemporalExp/huto:every/huto:even ?yEven
                    FILTER (?even != ?yEven)}
}
}

```

3) Les ressources répétitives par rapport à une fréquence donnée.

```

PREFIX huto: <http://ns.inria.fr/huto/> .
DESCRIBE ?s
WHERE {
    BIND ( CONCAT ('huto:', ?period) as ?frequency)

    ?s huto:hasTemporalExp ?o .
    ?o rdf:type huto:Cycle ;
        huto:every ?q .
    ?q rdf:type ?frequency .

    FILTER NOT EXISTS { ?q huto:sample ?t }
}

```

4) Les ressources dont l'occurrence est avant une ressource donnée..

```

PREFIX huto: <http://ns.inria.fr/huto/> .
PREFIX data: <http://example.org/data/> .
DESCRIBE ?z
WHERE {
    { ?x huto:uri ?resource }
    UNION
    { ?x huto:triple/(rdf:subject|rdf:object) ?resource }
    UNION
    { ?x huto:graph ?g .
      graph ?g { ?resource ?p ?o }
    }
}

```

```

        UNION
        { ?s ?p ?resource }}
    }
    FILTER NOT EXISTS { ?j ?k ?x }

    ?x huto:hasTemporalExp/(huto:hasDate|huto:hasBegin) ?t1 .
    ?z huto:hasTemporalExp/(huto:hasDate|huto:hasBegin) ?t2 .

    ?t1 huto:hasYear/huto:value ?y1 .
    ?t2 huto:hasYear/huto:value ?y2 .

    FILTER ( ?y1 > ?y2 ||
            (?y1 = ?y2 &&
             EXISTS {
                 ?t1 huto:hasMonth/huto:number ?m1 .
                 ?t2 huto:hasMonth/huto:number ?m2 .

                 FILTER (?m1 > ?m2 ||
                         (?m1 = ?m2 &&
                          EXISTS {
                              ?t1 huto:hasDay/huto:value ?d1 .
                              ?t2 huto:hasDay/huto:value ?d2 .
                              FILTER(?d1 > ?d2)
                          })
                        })
                })
            )
    }
}

```

5) Les ressources dont l'occurrence est après une ressource donnée.

```

PREFIX huto: <http://ns.inria.fr/huto/> .
PREFIX data: <http://example.org/data/> .
DESCRIBE ?z
WHERE {
    { ?x huto:uri ?resource }
    UNION
    { ?x huto:triple/(rdf:subject|rdf:object) ?resource }
    UNION
    { ?x huto:graph ?g .
      graph ?g{
        { ?resource ?p ?o }
        UNION
        { ?s ?p ?resource }}
    }
    FILTER NOT EXISTS { ?j ?k ?x }
}

```

```

?x huto:hasTemporalExp/(huto:hasDate|huto:hasBegin) ?t1 .
?z huto:hasTemporalExp/(huto:hasDate|huto:hasBegin) ?t2 .

?t1 huto:hasYear/huto:value ?y1 .
?t2 huto:hasYear/huto:value ?y2 .

FILTER ( ?y1 < ?y2 ||
         (?y1 = ?y2 &&
          EXISTS {
            ?t1 huto:hasMonth/huto:number ?m1 .
            ?t2 huto:hasMonth/huto:number ?m2 .

            FILTER (?m1 < ?m2 ||
                    (?m1 = ?m2 &&
                     EXISTS {
                       ?t1 huto:hasDay/huto:value ?d1 .
                       ?t2 huto:hasDay/huto:value ?d2 .
                       FILTER(?d1 < ?d2)
                     })
                  })
          )
        )
}

```

6) Détermination de la temporalité d'une ressource donnée.

```

PREFIX huto: <http://ns.inria.fr/huto/>
PREFIX data: <http://example.org/data/>
DESCRIBE ?x
WHERE{
  { ?x huto:uri ?resource }
  UNION
  { ?x huto:triple/(rdf:subject|rdf:object) ?resource }
  UNION
  { ?x huto:graph ?g .
    graph ?g{ { ?resource ?p ?o }
              UNION
              { ?s ?p ?resource }
            }
  }
  FILTER NOT EXISTS { ?j ?k ?x }
}

```

7) les ressources vraies à une date donnée.

```

PREFIX huto: <http://ns.inria.fr/huto/>.
PREFIX data: <http://example.org/data/>
DESCRIBE ?x
WHERE {
  {# Punctual resources
    ?x huto:hasTemporalExp/huto:hasDate [ huto:hasYear/huto:value ?year ;
                                           huto:hasMonth/huto:number ?month ;
                                           huto:hasDay/huto:value ?day]
  }
UNION
  {# Punctual resources expressed with hasBegin and hasEnd properties
    ?x huto:hasTemporalExp/huto:hasBegin
      [ huto:hasYear/huto:value ?y1 ;
        huto:hasMonth/huto:number ?m1 ;
        huto:hasDay/huto:value ?d1 ] ;
      huto:hasTemporalExp/huto:hasEnd
      [ huto:hasYear/huto:value ?y2 ;
        huto:hasMonth/huto:number ?m2 ;
        huto:hasDay/huto:value ?d2 ] .

    FILTER((?y1 < ?year && ?year < ?y2) ||
           (?y1 = ?year && (?m1 < ?month ||
                          (?m1 = ?month && ?d1 <= ?day))) ||
           (?y1 = ?year && ?year = ?y2 && ( (?m1 < ?month && ?month < ?m2) ||
           (?m1 = ?month && ?d1 <= ?day) || (?month = ?m2 && ?day <= ?d2) ||
           (?m1 = ?month && ?month = ?m2 && ?d1 <= ?day && ?day <= ?d2))))
  }
UNION
  {# Weekly resources
    ?x huto:hasTemporalExp/rdf:type huto:Cycle ;
      huto:hasTemporalExp/huto:every/rdf:type huto:Week ;
      huto:hasTemporalExp/huto:exp/huto:hasDate/huto:hasDay/rdf:type ?weekday .
      ?weekday rdfs:label ?dayOfWeek .
  }
UNION
  {# Weekly resources expressed with hasBegin and hasEnd properties
    ?x huto:hasTemporalExp/rdf:type huto:Cycle ;
      huto:hasTemporalExp/huto:every/rdf:type huto:Week ;
      huto:hasTemporalExp/huto:exp/huto:hasBegin/huto:hasDay/rdf:type ?weekday1 ;
      huto:hasTemporalExp/huto:exp/huto:hasEnd/huto:hasDay/rdf:type ?weekday2 .

    ?w rdfs:label ?dayOfWeek ;
      huto:position ?p .
    ?weekday1 huto:position ?p1 .
  }

```

```

?weekday2 huto:position ?p2 .

FILTER(?p1 <= ?p && ?p <= ?p2)
}
UNION
{# Monthly resources
{
?x huto:hasTemporalExp/huto:exp/huto:hasDate/huto:hasDay/huto:value ?day
}
}
UNION
{# Monthly resources expressed with hasBegin and hasEnd properties
?x huto:hasTemporalExp/huto:exp/huto:hasBegin/huto:hasDay/huto:value ?d1 ;
huto:hasTemporalExp/huto:exp/huto:hasEnd/huto:hasDay/huto:value ?d2 ;

FILTER(?d1 <= ?day && ?day <= ?d2)
}

?x huto:hasTemporalExp/rdf:type huto:Cycle ;
huto:hasTemporalExp/huto:every/rdf:type huto:Month .

# To calculate the parity of ?month
BIND((?month / 2) as ?e1)
BIND(ROUND(?e1) = ?e1 as ?even)

# To handle expressions like every even month (there is an implicit exception)
FILTER NOT EXISTS{ ?x huto:hasTemporalExp/huto:every/huto:even ?mEven
FILTER (?even != ?mEven)}
}
UNION
{# Yearly resources
{ ?x huto:hasTemporalExp/huto:exp/huto:hasDate/huto:hasMonth/huto:number ?month ;
huto:hasTemporalExp/huto:exp/huto:hasDate/huto:hasDay/huto:value ?day .
}
}
UNION
{# Yearly resources expressed with hasBegin and hasEnd properties
?x huto:hasTemporalExp/huto:exp/huto:hasBegin/huto:hasMonth/huto:number ?m1 ;
huto:hasTemporalExp/huto:exp/huto:hasBegin/huto:hasDay/huto:value ?d1 ;
huto:hasTemporalExp/huto:exp/huto:hasEnd/huto:hasMonth/huto:number ?m2 ;
huto:hasTemporalExp/huto:exp/huto:hasEnd/huto:hasDay/huto:value ?d2 .

FILTER((?m1 < ?month && ?month < ?m2) ||
(?m1 = ?month && ?month != ?m2 && ?d1 <= ?day) ||
(?m1 = ?month && ?month = ?m2 && ?d1 <= ?day && ?day <= ?d2) ||
(?month = ?m2 && ?day <= ?d2))
}

```

```
}

?x huto:hasTemporalExp/rdf:type huto:Cycle ;
   huto:hasTemporalExp/huto:every/rdf:type huto:Year ;

# To calculate the parity of ?year
BIND((?year / 2) as ?e1)
BIND(ROUND(?e1) = ?e1 as ?even)

# To handle expressions like every even year (there is an implicit exception)
FILTER NOT EXISTS { ?x huto:hasTemporalExp/huto:every/huto:even ?yEven
                    FILTER (?even != ?yEven)}
}
}
```

Annexe : Installation des Outils

Dans cette partie, nous présentons les détails d'installation des outils utilisés pour développer la plateforme WestAfricapedia. SMW a une forte communauté d'utilisateurs et toutes les extensions développées ont leur propre documentation. Ainsi, pour plus de détails de l'installation des extensions utilisées dans notre plateforme le lecteur se réfèra à la page de documentation de l'extension qui sera donnée en référence.

MW 1.26.2

La version 1.26.2 de MW du 21 Décembre 2015 a été installée. C'est un fichier zip¹ qu'il faut extraire dans un serveur (`/var/www/html/mediawiki`). Au préalable, une base de données relationnelle doit être installée où sont stockées toutes les données de MW. Dans notre cas, nous avons utilisé MySQL 5.5.47. Nous avons créé une base de donnée appelée `SemWiki`. Après cela, il faut accéder à l'URL du serveur via un navigateur et suivre les instruction du guide d'installation (`localhost/mediawiki`). Quand l'installation² se termine, le téléchargement du fichier `localSettings.php` proposé doit être fait et sauvegardé à la racine du Wiki (`/var/www/html/mediawiki`).

SMW 2.3.1

SMW propose deux types d'installation³ : une installation en utilisant *Composer* qui est un gestionnaire de dépendance PHP permettant d'installer des packages et une installation avec Tarball qui est utilisée si les commandes SHELL ne sont pas disponibles. Nous avons utilisé l'installation avec le gestionnaire composer⁴. Après l'installation, il suffit d'ajouter la ligne suivante en bas du fichier `localSettings.php` téléchargé lors de l'installation de MW pour que MW puisse intégrer SMW où `example.org` est l'URL du site:

```
enableSemantics( 'example.org' );
```

Une fois l'installation terminée, il faut se connecter comme administrateur du site et aller sur `SpecialPages` (pages spéciales) pour créer les tables utilisées par SMW dans la base de données utilisées par MW.

¹<https://releases.wikimedia.org/mediawiki/1.26/mediawiki-1.26.2.tar.gz>

²https://www.mediawiki.org/wiki/Manual:Installation_guide

³<https://www.semantic-mediawiki.org/wiki/Help:Installation>

⁴https://www.semantic-mediawiki.org/wiki/Help:Installation/Using_Composer_with_MediaWiki_1.25%2B

Semantic Forms 3.5

La version 3.5 de Semantic form a été installée. Après installation⁵, il suffit seulement d'ajouter la ligne suivante en bas du fichier `localSettings.php` téléchargé lors de l'installation de MW :

```
include_once "$IP/extensions/SemanticForms/SemanticForms.php";
```

ParserFunctions 1.6.0

La version 1.6.0 de l'extension ParserFunctions a été installée. Après installation⁶, il suffit seulement d'ajouter la ligne suivante en bas du fichier `localSettings.php` téléchargé lors de l'installation de MW :

```
wfLoadExtension( 'ParserFunctions' );
```

Variables 2.1.0

La version 2.1.0 de l'extension Variables a été installée. Après installation⁷, il suffit d'ajouter la ligne suivante en bas du fichier `localSettings.php` téléchargé lors de l'installation de MW :

```
require_once "$IP/extensions/Variables/Variables.php";
```

ImageMap 1.26

La version 2.16 de l'extension ImageMap a été installée. Après installation⁸, il suffit seulement d'ajouter la ligne suivante en bas du fichier `localSettings.php` téléchargé lors de l'installation de MW :

```
wfLoadExtension( 'ImageMap' );
```

Virtuoso 7.2.2

La version open source de Virtuoso⁹ contient toutes les informations pour son installation. Dans notre configuration, Virtuoso est accessible à l'URL `http://localhost:8890`.

Une fois installé, il faut paramétrer SMW pour qu'il puisse écrire les données annotées dans la base de connaissances de Virtuoso grâce aux lignes suivantes qu'il faut mettre dans le fichier `localSettings.php` :

⁵https://www.mediawiki.org/Wiki/Extension:Semantic_Forms/Download_and_installation

⁶<https://www.mediawiki.org/Wiki/Extension:ParserFunctions>

⁷<https://www.mediawiki.org/Wiki/Extension:Variables>

⁸<https://www.mediawiki.org/Wiki/Extension:ImageMap>

⁹<https://github.com/openlink/virtuoso-opensource/releases/download/v7.2.2.1/virtuoso-opensource-7.2.2.tar.gz>

```

$smwgDefaultStore = 'SMWSparqlStore';
$smwgSparqlDatabaseConnector = 'virtuoso';
$smwgSparqlQueryEndpoint = 'http://localhost:8890/sparql/';
$smwgSparqlUpdateEndpoint = 'http://localhost:8890/sparql/';
$smwgSparqlDataEndpoint = '';
$smwgSparqlDefaultGraph = 'http://westafricapedia.sn/';

```

Ainsi, la première ligne ci-dessus indique à SMW d'utiliser un triple store pour sauvegarder les données sémantiques. La deuxième ligne indique le nom du triple store utilisé. La troisième ligne, spécifie l'URL pour les requêtes de lecture sur le triple store. La quatrième ligne, spécifie l'URL pour les requêtes d'écriture sur le triple store. La cinquième ligne, spécifie l'URL pour les données pour des services HTTP de SPARQL qu'il n'est pas obligatoire de spécifier. Enfin la sixième ligne, spécifie l'URL du graphe utilisé dans la base de connaissances pour sauvegarder les données provenant de SMW. En effet, dans notre architecture (figure 7.1) la base de connaissances a été scindée en trois entrepôts (graphes) : 1) un pour sauvegarder les données annotées de SMW (<http://westafricapedia.sn/>), 2) un pour sauvegarder l'ontologie USCO (<http://usco.sn/>) et 3) un pour sauvegarder l'ontologie HuTO (<http://huto.sn/>).

Après installation de Virtuoso, il faut le configurer pour qu'il puisse accepter des requêtes de mise à jour. Pour cela, il faut se connecter en tant qu'administrateur et exécuter les requêtes ISQL suivantes :

```

GRANT EXECUTE ON DB.DBA.SPARQL_INSERT_DICT_CONTENT TO "SPARQL";
GRANT EXECUTE ON DB.DBA.SPARQL_DELETE_DICT_CONTENT TO "SPARQL";
GRANT EXECUTE ON SPARQL_DELETE_DICT_CONTENT to "SPARQL";
GRANT EXECUTE ON SPARQL_DELETE_DICT_CONTENT to SPARQL_UPDATE;
GRANT EXECUTE ON DB.DBA.SPARUL_LOAD_SERVICE_DATA TO "SPARQL";
GRANT EXECUTE ON DB.DBA.SPARQL_SD_PROBE TO "SPARQL";
GRANT EXECUTE ON DB.DBA.L_O_LOOK TO "SPARQL";
GRANT EXECUTE ON DB.DBA.RDF_QUAD to SPARQL_UPDATE;

```


Bibliography

- [1] World Summit on the Information Society (2014) Building inclusive Knowledge Societies. A review of UNESCO's action in implementing the WSIS outcomes. ADVANCE VERSION, pages 66. (Cité en pages 1 et 2.)
- [2] Berners-Lee T., Hendler J., Lassila O. (2001). The Semantic Web. Scientific American, Vol 284, n 5, pages 34-43. (Cité en pages 2 et 4.)
- [3] Gruber T. (2007). Collective Knowledge Systems: Where the Social Web meets the Semantic Web. Web Semantics: Science, Services and Agents on the World Wide Web, doi:10.1016/j.websem.2007.11.011. (Cité en pages 3 et 5.)
- [4] Gruber T. (1993). A translation Approach to portable ontology specifications. In Knowledge Acquisition Volume 5, Issue 2, pp. 199-220. (Cité en page 4.)
- [5] Gómez-Pérez A., Fernandez-Lopez M., Corcho O. (2004) Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. Springer-Verlag New York, Inc. Secaucus, NJ, USA, ISBN:1846283965, pages 411. (Cité en pages 4 et 5.)
- [6] Grüninger M and Fox M S (1995). The Role of Competency Questions in Enterprise Engineering. IFIP WG 5.7 Workshop on Benchmarking. Theory and Practice. Trondheim, Norway. (Cité en page 4.)
- [7] Uschold M and King M (1995). Towards a Methodology for Building Ontologies. Workshop on Basic Ontological Issues in Knowledge Sharing. (Cité en page 4.)
- [8] Gómez-Pérez A., Juristo N and Pazos J (1995). Evaluation and assessment of knowledge sharing technology. In N. J. Mars (editor), Towards Very Large Knowledge Bases. KBKS 95. IOS Press, Amsterdam, pp.,289-296. (Cité en pages 4 et 5.)
- [9] Swartout B, Ramesh P, Knight K and Russ T (1997). Toward Distributed Use of Large-Scale Ontologies. Symposium on Ontological Engineering of AAAI. Stanford (California). Mars. (Cité en pages 4 et 5.)
- [10] Aussenac-Gilles N, Despres S and Szulman S (2008). The Terminae Method and Platform for Ontology Engineering from Texts. In Bridging the Gap between Text and Knowledge - Selected Contributions to Ontology Learning and Population from Text, IOS press, pp.192-223. (Cité en page 4.)
- [11] Noy N F and McGuinness . L (2001). Ontology Development 101: a guide to creating your first ontology. Technical Report KSL-01-05. (Cité en pages 5 et 64.)
- [12] Diallo P F, Ndiaye S M, and Lo M (2011). Study of Sociocultural Ontology. The First International Conference on Social Eco-Informatics., 1(5):69-74. (Cité en page 5.)

- [13] Hall E.T. (1984) *La danse de la vie - Temps culturel, Temps vécu*. Paris, Éditions du Seuil, pp. 23-72. (Cité en pages 5, 67, 68, 92 et 160.)
- [14] Passant A. (2009). *Technologies du Web Sémantique pour l'Entreprise 2.0*. PhD Thesis, Université Paris-IV Sorbonne. (Cité en page 6.)
- [15] Scott S.E. and Palincsar A.S. (2013) *Sociocultural Theory*. In *Psychology of Classroom Learning: An Encyclopedia*. Farmington Hills, MI: Gale Group, publication 2009. Updated version on Jul 15, 2013, <http://www.education.com/reference/article/sociocultural-theory/> (Cité en page 14.)
- [16] Engeström Y. (1987) *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki, pages 269. FI: Orienta-Konsultit. (Cité en pages xiii, 14, 15, 17, 18, 35 et 158.)
- [17] Vygotsky L.S. (1978) *Mind in society: the development of higher psychological processes*. pages 159. Cambridge: Harvard University Press. (Cité en pages xiii, 5, 14, 15, 16, 35, 75, 82 et 158.)
- [18] Leont'ev A.N. (1979) *The development of the mind*. pages 428, ISBN 978-0-9805428-6-8. Erythrós Press and Media. (Cité en pages 5, 14, 16, 17, 35 et 158.)
- [19] Engeström Y. (2000) *Activity theory as a framework for analyzing and redesigning work*. In *Ergonomics*, pp. 960-974. (Cité en pages 14, 15, 17 et 158.)
- [20] Cole M. (2010) *Cultural-Historical Activity Theory*. In *Qualitative Research*, Elsevier Ltd. (Cité en pages 15, 19, 36, 74, 75 et 158.)
- [21] Wertsch J.V., Del Rio P. and Alvarez A. (1995) *Sociocultural studies: History, action, and mediation*. In Wertsch, P. del Rio and Alvarez, A. (eds.) *Sociocultural Studies of Mind*, pp 1-34. New York: Cambridge University Press (Cité en pages 15 et 158.)
- [22] Cole M. (1995) *Socio-cultural-historical psychology: some general remarks and a proposal for a new kind of cultural genetic methodology*. In Wertsch, P. del Rio and Alvarez, A. (eds.) *Sociocultural Studies of Mind*, pp 187-214. New York: Cambridge University Press (Cité en pages 15 et 158.)
- [23] John-Steiner V. and Mahn H. (1996) *Sociocultural Approaches to Learning and Development: A Vygotskian Framework*. In *Educational Psychologist*, 31 (3/4), pp. 191-206. (Cité en pages 15, 18, 36, 74 et 158.)
- [24] Kaptelinin V. (2014) *Activity Theory*. In: Soegaard, Mads and Dam, Rikke Friis (eds.). *The Encyclopedia of Human-Computer Interaction*, 2nd Ed.. Aarhus, Denmark: The Interaction Design Foundation. (Cité en pages 15, 17, 19, 35, 64, 74, 75 et 158.)

- [25] Kuuti K. (1995) Activity theory as a potential framework for human-computer interaction research. In: Nardi, Bonnie A. (ed.). *Context and Consciousness: Activity Theory and Human-Computer Interaction*. The MIT Press pp. 17-44 (Cité en pages 15, 18, 19, 20, 36 et 74.)
- [26] Yamagata-Lynch L.C. (2010) Understanding Cultural Historical Activity Theory. In *Activity Systems Analysis Methods: Understanding Complex Learning Environments*, DOI 10.1007/978-1-4419-6321-5-2, Springer Science Business Media. pp. 13-26. (Cité en pages 15, 17 et 75.)
- [27] Kaptelinin V. (2005) The Object of Activity: Making Sense of the Sense-Maker. In: *Mind, Multure and Activity*. Regents of the University of California on behalf of the Laboratory of Comparative Human Cognition. pp. 4-18. (Cité en page 18.)
- [28] Brown A.L. (1992) Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, 2(2), pp. 141-178. (Cité en pages 19, 36, 74 et 158.)
- [29] Brown A.L., Ash D., Rutherford M., Nakagawa K., Gordon A. and Campione J.C. (1993) Distributed expertise in the classroom. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations*, pp. 188-228. New York: Cambridge University Press. (Cité en pages 19 et 36.)
- [30] Teemant A. (2005) Evaluating Socio-Cultural Pedagogy in a Distance Teacher Education Program. In *Teacher Education Quarterly* Vol. 32, No. 3, *Assessing Innovation in Teacher Education* (Summer 2005), pp. 49-62. (Cité en pages 19, 36, 74 et 158.)
- [31] Palincsar A.S. (1999) Applying a sociocultural lens to the work of a transition community, *Discourse Processes*, 27:2, pp. 161-171. Disponible sur <http://dx.doi.org/10.1080/01638539909545056>. (Cité en pages 19 et 36.)
- [32] Koszalka T.A. and Chun-Ping W. (2004) Cultural Historical Activity Theory [CHAT] analysis of technology integration: Case study of two teachers. In *Association for Educational Communications and Technology*, 27th, Chicago, IL, October 19-23, pp. 492-502. (Cité en pages 19, 20, 36 et 74.)
- [33] Bjørkli C.A., Røed B.K., Bjelland H.V., Gould K.S. and Hoff T. (2007) La théorie de l'activité comme aide à la conception d'équipement de navigation. *Activités*, 4 (1), pp. 170-178, <http://www.activites.org/v4n1/v4n1.pdf>. (Cité en pages 19, 20, 36 et 74.)
- [34] Bødker S. and Klokose C.N. (2011) The Human-Artifact Model: An Activity Theoretical Approach to Artifact Ecologies, *Human-Computer Interaction*, 26:4, 315-371. Disponible sur <http://dx.doi.org/10.1080/07370024.2011.626709> (Cité en pages 19, 20, 36 et 74.)

- [35] Wilson T.D. (2008) Activity Theory and Information Seeking. In Annual Review of Information Science and Technology Volume 42, Issue 1, pp. 119-161. (Cité en page 20.)
- [36] Toomela A. (2008) Activity theory is a dead end for methodological thinking in cultural psychology too. *Culture and Psychology*, 14, 289-303. (Cité en page 20.)
- [37] Engeström Y. (2009) The Future Of Activity Theory: a Rough Draft. In the book *Learning and Expanding with Activity Theory*, edited by Annalisa Sannino, Harry Daniels and Kris Gutierrez, pp. 303-357. ISBN-13 978-0-511-65159-5, Cambridge University Press. (Cité en page 20.)
- [38] Sannino A., Daniels H. and Gutiérrez K. D. (2009) Activity Theory Between Historical Engagement and Future-Making Practice. In the book *Learning and Expanding with Activity Theory*, edited by Annalisa Sannino, Harry Daniels and Kris Gutierrez, pp. 1-15. ISBN-13 978-0-511-65159-5, Cambridge University Press. (Cité en page 20.)
- [39] White N. (2005) How Some Folks Have Tried to Describe Community. Full Circle Assoc. Disponible sur <http://www.fullcirc.com/community/definingcommunity.htm>. (Cité en page 22.)
- [40] Rheingold H. (1993) *The virtual community: Homesteading on the electronic frontier*. Reading, MA: Addison-Wesley. Disponible sur <http://www.rheingold.com/vc/book/> (Cité en page 22.)
- [41] Hagel, H. and Armstrong A. (1997) *Net gain: Expanding markets through virtual communities*. Harvard Business School Press, Boston, MA, USA. (Cité en pages 22, 23, 28 et 36.)
- [42] Jones Q. and Rafaeli S. (2000) Time to split, virtually: 'Discourse architecture' and 'community building' create vibrant virtual publics. *Electronic Markets*, 10(4) 214-223. (Cité en pages 22, 23, 28, 29, 31 et 36.)
- [43] Lee F.S.L., Douglas V. and Moez L. (2003) Virtual Community Informatics: A Review and Research Agenda. In *Journal of Information Technology Theory and Application (JITTA)*: Vol. 5: Iss. 1, Article 5. Available at: <http://aisel.aisnet.org/jitta/vol5/iss1/5> (Cité en pages 22, 23 et 28.)
- [44] Porter C.S.E. (2004) A Typology of Virtual Communities: A Multi-Disciplinary Foundation for Future Research. In *Journal Computer-Mediated Communication* 01/2004; 10. Disponible sur <http://jcmc.indiana.edu/vol10/issue1/porter.html> (Cité en pages xiii, 24, 26, 27, 29, 30 et 36.)
- [45] Lazar J. and Preece J. (1998) Classification Schema for Online Communities. Proceedings of in the 1998 Association for Information Systems, Americas Conference, 84-86. (Cité en pages 24, 26, 27 et 36.)

- [46] Markus U. (2002) *Characterizing the virtual community* (5th edition). SAP Design Guild. (Cit  en pages xiii, 25, 26, 28 et 36.)
- [47] Stephens M. (2012). *Recent Research on Virtual Communities*. Disponible sur http://webjunction.org/documents/webjunction/Recent_Research_on_Virtual_Communities.html (Cit  en pages 29 et 30.)
- [48] Stanoevska-Slabeva K. and Schmid B.F. (2001). *A Typology of Online Communities and Community Supporting Platforms*. In *Proceedings of the 34th Hawaii International Conference on System Sciences*. (Cit  en pages 29 et 30.)
- [49] Wenger E. (2006). *Communities of practice a brief introduction*. Disponible sur <http://www.ewenger.com/theory/> (Janvier 2016). (Cit  en page 30.)
- [50] Scardamalia M. and Bereiter C. (1994). *Computer support for knowledge-building communities*. In the *Journal of the Learning Sciences*, 3(3), pp. 265-283. Disponible sur http://hrast.pef.uni-lj.si/~joze/podiplomci/prs/clanki03/CSILE_Scardamaila.htm (Janvier 2016). (Cit  en pages 8, 30, 65, 74, 82, 90 et 159.)
- [51] Scardamalia M. and Bereiter C. (2006). *Knowledge building: Theory, pedagogy, and technology*. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* pp. 97-118. New York, NY: Cambridge University Press. Disponible sur <http://ikit.org/fulltext/KBTheory.pdf> (Janvier 2016). (Cit  en pages 30, 65, 74, 82, 90 et 159.)
- [52] Scardamalia M. and Bereiter C. (2010) *A Brief History of Knowledge Building*. *Canadian Journal of Learning and Technology*, 36(1). (Cit  en pages 8, 65, 74, 82, 90 et 159.)
- [53] Mitchell E. (2007). *Three types of community*. Disponible sur <http://www.edmitchell.co.uk/blog/2007/11/16/three-types-of-community/> (Janvier 2016). (Cit  en page 31.)
- [54] Preece J. (2001). *Sociability and usability: Twenty years of chatting online*. In *Behavior and Information Technology Journal*, 20, 5, 347-356. (Cit  en pages 32, 33, 34 et 35.)
- [55] Ridings C.M. and Gefen D. (2004) *Virtual Community Attraction: Why People Hang Out Online*. In *Journal of Computer-Mediated Communication* 2004, <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2004.tb00229.x/full> . (Cit  en pages 32, 34, 35 et 88.)
- [56] Ivic I (1994). *Lev. S. Vygotsky*. In *Prospects:the quarterly review of comparative education* (Paris, UNESCO: International Bureau of Education), vol. XXIV, no. 3/4, p. 471-485. (Cit  en pages 35 et 64.)

- [57] Allen J F (1981). An Interval-Based Representation of Temporal Knowledge. In Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI'81), p. 221-226, Canada. (Cité en pages 37, 39, 100, 122 et 163.)
- [58] Allen J F (1983). Maintaining knowledge about temporal intervals. In Communications of the ACM, 26(11): p. 832-843. (Cité en pages 37, 39, 100, 122 et 163.)
- [59] Ladkin P. B. (1986). Primitives and Units for Time Specification. In the Fifth National Conference on Artificial Intelligence, Philadelphia, Pennsylvania (USA). (Cité en pages 37, 38 et 39.)
- [60] Ladkin P. B. (1987). The Logic of Time Representation. PhD Thesis at the University of California at Berkeley, November. (Cité en pages 37, 38 et 39.)
- [61] Studer R, Benjamins V.R., Fensel D. (1998). Knowledge Engineering: Principles and Methods. IEEE Transactions on Data and Knowledge Engineering 25(1-2), pp. 161-197. (Cité en page 38.)
- [62] Baumann R., Loebe F. and Herre H. (2012). Ontology of Time in GFO. 7th International Conference on Formal Ontology in Information Systems, July 24-27. (Cité en pages 40 et 60.)
- [63] Lenat D. B. and Guha R. V. (1990). Building Large Knowledge-based Systems: Representation and Inference in the Cyc Project. Addison-Wesley, Boston, Massachusetts. (Cité en pages 40, 41 et 60.)
- [64] Witbrock M., Pittman K., Moszkowicz J., Beck A. and Lenat D. B. (2015). Cyc and the Big C: Reading that Produces and Uses Hypotheses about Complex Molecular Biology Mechanisms. Scholarly Big Data: AI Perspectives, Challenges, and Ideas: Papers from the 2015 Association for the Advancement of Artificial Intelligence (AAAI) Workshop, pp 27-32. (Cité en pages 40, 41 et 60.)
- [65] ICOM/CIDOC CRM Special Interest Group (2015). Definition of the CIDOC Conceptual Reference Model. Version 6.0. (Cité en pages 40, 42 et 60.)
- [66] Pease A. (2007). The Suggested Upper Merged Ontology (SUMO) at Age 7: Progress and Promise. At Ontolog. (Cité en pages 40, 43 et 60.)
- [67] Masolo C., Borgo S., Gangemi A., Guarino N., Oltramari A. (2003) WonderWeb Deliverable D18 Ontology Library (final). IST Project 2001-33052 WonderWeb: Ontology Infrastructure for the Semantic Web. (Cité en pages 40, 43, 60 et 93.)
- [68] Tao C., Solbrig H. R. and Chute C. G. (2011). CNTRO 2.0: A Harmonized Semantic Web Ontology for Temporal Relation Inferencing in Clinical Narratives. AMIA Jt Summits Transl Sci Proc. 2011; 2011: 64-68. Published online Mar 7, 2011. Available at <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3248753/> (Janvier 2016). (Cité en pages 40, 49, 60 et 67.)

- [69] Tao C., Wei-Qi Wei, Solbrig H. R., Savova G. and Chute C. G. (2010). CN-TRO: A Semantic Web Ontology for Temporal Relation Inferencing in Clinical Narratives. *AMIA Annu Symp Proc.* 2010; 2010: 787-791. Published online Nov 13, 2010. Available at <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041418/> (Janvier 2016). (Cité en pages 40, 49, 60 et 67.)
- [70] Sauri R., Littman J., Knippen B., Gaizauskas R., Setzer A. et Pustejovsky J. (2006). *TimeML Annotation Guidelines Version 1.2.1.* (Cité en pages 40, 50, 60 et 67.)
- [71] W3C Working Group Note 12 April 2006. Defining N-ary Relations on the Semantic Web. Available at <http://www.w3.org/TR/swbp-n-aryRelations/> (September, 2015). (Cité en pages 44, 53, 54 et 61.)
- [72] Mylka A, Sauermann L., Sintek M., van Elst L. (2012) NEPOMUK Calendar Ontology. *OSCAF Recommendation 13.06.2012.* (Cité en pages 44 et 45.)
- [73] Colpaert P., De Vocht L., Verstockt S., Dimou A., Buyle R., Mannens E. and Van de Walle, R. (2014). *Ontology Based Improvement of Opening Hours in E-governments.* Published in: *Semantic Computing (ICSC), 2014 IEEE International Conference*, pp 167-170. (Cité en pages 44 et 45.)
- [74] Jewell M. O., Lawrence F. K. and Tuffield M. M. (2005) *OntoMedia: An Ontology for the Representation of Heterogeneous Media.* In, *Multimedia Information Retrieval Workshop, ACM SIGIR.* (Cité en page 45.)
- [75] Lawrence F. K. (2010) *OntoMedia: Telling Stories to Your Computer.* *First International AMICUS Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts.* (Cité en page 45.)
- [76] Pan F. and Hobbs J. R. (2005). *Temporal Aggregates in OWL-Time.* *FLAIRS Conference* p. 560-565. (Cité en pages 45, 46, 60 et 67.)
- [77] Pan F. (2007). *Representing complex temporal phenomena for the semantic web and natural language.* PhD Thesis at the University of Southern California. December. (Cité en pages 45, 60 et 67.)
- [78] Tappolet J. and Bernstein A. (2009) *Applied Temporal RDF: Efficient Temporal Querying of RDF Data with SPARQL.* In *The Semantic Web: Research and Applications*, Springer. (Cité en page 53.)
- [79] Gutierrez C., Hurtado C., and VAisman A. (2005) *Temporal RDF.* In *European Conference on the Semantic Web (ECSW'05) (Best paper award)*, p 93-107. (Cité en pages 56, 57, 61, 68 et 106.)
- [80] Welty C. and Fikes R. (2006). *A Reusable Ontology for Fluents in OWL.* In *Frontiers in Artificial Intelligence and Applications*, 150:226-236. (Cité en pages 56, 58, 61 et 68.)

- [81] Baratis E., Petrakis G. M., Batsakis S., Maris N. and Papadakis N. (2009). TOQL: Temporal Ontology Querying Language. In International Symposium on Spatial and Temporal Databases, pp. 338-354. (Cité en pages 56, 59 et 61.)
- [82] Batsakis S., and Petrakis G. M. (2011). SOWL: A Framework for Handling Spatio-temporal Information in OWL 2.0. RuleML Europe 2011: 242-249. (Cité en pages 56, 59 et 61.)
- [83] Frasinca F., Milea V. and Kaymak U. (2010) tOWL: Integrating Time in OWL. In book Semantic Web Information Management, pp 225-246. Springer Berlin Heidelberg. (Cité en pages 56 et 59.)
- [84] Milea V., Frasinca F., Kaymak U. (2012) tOWL: A Temporal Web Ontology Language. IEEE Transactions on Systems, Man, and Cybernetics, Part B 42(1): 268-281. (Cité en pages 56 et 59.)
- [85] Rula A., Palmonari M., Ngomo A. C. N., Gerber D., Lehmann J. and Bühmann L. (2014). Hybrid Acquisition of Temporal Scopes for RDF Data. In The Semantic Web: Trends and Challenges, p. 488-503. (Cité en pages 56 et 60.)
- [86] Scheuermann A., Motta E., Mulholland P., Gangemi A. and Presutti V. (2013) An Empirical Perspective on Representing Time. In 7th International Conference on Knowledge Capture, K-CAP, p. 89-96. (Cité en pages 56 et 60.)
- [87] Artale A. and Franconi E. (2001) A survey of temporal extensions of description logics. In Annals of Mathematics and Artificial Intelligence 30, pp. 171-210, Kluwer Academic Publishers. Printed in the Netherlands. (Cité en pages 57, 60 et 61.)
- [88] Artale A., Kontchakov R., Wolter F. and Zakharyashev M. (2013) Temporal Description Logic for Ontology-Based Data Access. In Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, pp. 711-717. (Cité en pages 57, 60, 61 et 68.)
- [89] Vaisman A. and Hurtado C. (2006). Reasoning About Temporal Constraints in RDF. Principles and Practice of Semantic Web Reasoning. Lecture Notes in Computer Science Volume 4187, pp. 164-178. (Cité en page 61.)
- [90] Noy N.F. (2004) Semantic Integration: A Survey Of Ontology-Based Approaches. In ACM SIGMOD Record, Vol. 33, No. 4, December, pp. 65-70. (Cité en page 78.)
- [91] Scardamalia M. (2002) Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), Liberal education in a knowledge society, pp. 67-98. Chicago: Open Court. (Cité en page 83.)
- [92] Dione S. (2004) L'appel du Ndut ou l'initiation des garçons Seereer. IFAN/Enda-Editions, 2004. (Cité en page 88.)

-
- [93] Hayes P. (1996) A Catalog of Temporal Theories. Technical Report UIUC-BI-AI-96-01, University of Illinois, p61. (Cité en pages 93, 97 et 98.)
- [94] Ter Horst H.J. (2005) Completeness, Decidability and Complexity of Entailment for RDF Schema and a Semantic Extension Involving the OWL Vocabulary. In Journal of Web Semantics 3 pp. 79-115. (Cité en page 110.)
- [95] Corby O., Gaignard A., Faron-Zucker C. and Motagnat J. (2012) KGRAM Versatile Data Graphs Querying and Inference Engine. In Proc. IEEE/WIC/ACM International Conference on Web Intelligence, Macau. (Cité en pages 112, 116 et 162.)
- [96] Corby O. and Faron-Zucker C. (2015) STTL: A SPARQL-based Transformation Language for RDF. 11th International Conference on Web Information Systems and Technologies, May 2015, Lisbon, Portugal. (Cité en pages 116, 123, 161 et 162.)
- [97] Corby O. and Faron-Zucker C. (2015) Un navigateur pour les données liées du Web. Journées francophones d'Ingénierie des Connaissances, Jul 2015, Rennes, France. IC2015, AFIA. (Cité en pages 116, 123, 161 et 162.)
- [98] Diallo P.F., Corby O., Mirbel I. (2015) HuTO Vocabulary Specification. Disponible sur <http://ns.inria.fr/huto> (Cité en page 124.)