



HAL
open science

OntoTextes (Ontologies et Textes) - Bilan du projet de recherche financé au titre du BQR 2007/AO1

Nathalie Aussenac-Gilles, Marie-Pierre Gleizes, Ollivier Haemmerlé, Josiane Mothe

► To cite this version:

Nathalie Aussenac-Gilles, Marie-Pierre Gleizes, Ollivier Haemmerlé, Josiane Mothe. OntoTextes (Ontologies et Textes) - Bilan du projet de recherche financé au titre du BQR 2007/AO1. [Research Report] IRIT. 2007. hal-03811071

HAL Id: hal-03811071

<https://hal.science/hal-03811071>

Submitted on 12 Oct 2022

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

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



RAPPORT INTERNE IRIT

IRIT/RR--2007-24--FR

DECEMBRE 2007

NATHALIE AUSSENAC-GILLES, MARIE-PIERRE GLEIZES, OLLIVIER
HAEMMERLE, JOSIANE MOTHE,

ET

ALBERTO ASENSIO, VALERIE CAMPS, PIERRE GLIZE, NATHALIE
HERNANDEZ, KEVIN OTTENS, BERNARD ROTHENBURGER

*OntoTextes (Ontologies et
Textes) - Bilan du projet de
recherche financé au titre du
BQR 2007/A01*

Bilan du projet de recherche financé au titre du BQR 2007/AO1

OntoTextes (*Ontologies et Textes*)

Thème du projet de recherche (titre)

Modélisation de connaissances pour la gestion documentaire.

Acronyme : **OntoTextes** (*Ontologies et Textes*)

Equipes concernées

Responsable du projet

Nathalie AUSSENAC-GILLES, Chargée de recherches du CNRS en Informatique à l'IRIT, équipe CSC

aussenac@irit.fr. 05.61.55.82 93

Partenaires du projet

- Equipe IC3 (Ingénierie des Connaissances, de la Coopération et de la Cognition), composante « modélisation de connaissances et analyse de textes », dont Ollivier Haemmerlé, anciennement rattaché à l'équipe ISYCOM du laboratoire GRIMM de l'UTM.
- Equipe SIG (Systèmes d'information Généralisés), composante **EVI** (Exploration et Visualisation d'Information), responsable pour le projet : Josiane Mothe
- Equipe SMAC (Systèmes Multi-Agents Coopératifs) responsable Marie-Pierre Gleizes.

Auteurs du documents

Résumé

Depuis près de 10 ans, l'équipe IRIT-CSC (devenue IC3) a constitué une compétence en construction de modèles de connaissances, tels que les ontologies, à partir de textes, et ce en utilisant des logiciels d'analyse du langage naturel. Cette approche s'avère particulièrement pertinente lorsque les ontologies doivent servir dans des applications de gestion documentaire et de recherche d'information. Des collaborations régulières avec l'équipe IRIT-SIG ont montré que l'utilisation d'ontologies pour l'annotation conceptuelle de documents permet de mieux les organiser en classes, de les retrouver ou d'y rechercher des informations. Pour cela, les ontologies utilisées doivent être bien adaptées et au plus près du contenu des textes. De nouvelles perspectives s'ouvrent pour approfondir les deux facettes de cette recherche : l'enrichissement régulier d'ontologies à partir de textes ; l'utilisation de ces ontologies pour comparer, annoter et classer des

collections de textes. Ce projet a eu pour objectif de faciliter l'exploitation conjointe des travaux des différentes équipes concernées en vue d'intégrer plus rapidement les nouveaux membres du laboratoire. Deux des trois réalisations envisagées ont été réalisées, et les partenaires ont pu monter ensemble une proposition de projet pour l'ANR –RNTL, le projet Dynamo, qui a été retenu en juillet 2007.

Contexte - Objectifs initiaux

Les questions fondamentales abordées par ces recherches couvrent les points suivants :

- L'adaptation des méthodes et outils de construction de ressources ontologiques et d'utilisation de ces ressources pour la gestion documentaire à un contexte évolutif, dans lequel les connaissances du domaine, les collections de documents et/ou les besoins des utilisateurs évoluent ;
- L'intégration des approches de l'ingénierie des connaissances, de la représentation des connaissances et de la recherche d'information pour la bonne utilisation de ces ressources ;
- La confrontation de différentes propositions élaborées sur des domaines spécialisés.

Dans ce contexte, l'axe de recherche envisagé dans ce projet était de spécifier comment adapter à des contextes dynamiques, les outils et méthodes d'ingénierie des ontologies et d'annotation sémantique. Ce projet s'est appuyé sur l'expérience des équipes partenaires et leurs travaux pour préparer la recherche future. Il se proposait donc d'expérimenter la complémentarité des approches et des logiciels disponibles sur trois questions pour lesquelles la convergence est immédiate afin de préparer la recherche pour les projets futurs.

Travail réalisé et résultats obtenus

Trois axes ont été envisagés : (i) apport des systèmes multi-agents à la maintenance des « dimensions » (hiérarchies de termes) guidant la navigation dans des collections documentaires, (ii) comparaison d'ontologies et de textes pour évaluer la couverture de communautés scientifiques dans le domaine spatial, (iii) annotation de données scientifiques d'observatoires virtuels par des concepts d'ontologies. Le 3^e point a été abandonné faute de disponibilité des données et faute de temps. En revanche, un travail proche a été réalisé sur des données pédagogiques, annotées à l'aide d'éléments d'ontologie. Finalement, trois études ont été réalisées :

- 1) Evaluation du système multi-agent Dynamo pour maintenir des hiérarchies de concepts guidant la navigation dans des collections documentaires
- 2) Annotation sémantique de ressources pédagogiques dans un entrepôt de données à l'aide d'OntoExplo
- 3) Comparaison d'ontologies et de textes

Chacune des études s'est déroulée en parallèle. Un bilan a été tiré des expériences et différentes questions scientifiques soulevées. Cette collaboration et les résultats positifs obtenus ont permis de monter un projet commun, Dynamo, en réponse à l'appel d'offre de l'ANR-RNTL et de rédiger des publications. Il constitue un cadre dans lequel poursuivre ces études.

Les résultats obtenus sont à la fois pratiques, techniques et organisationnels :

- Des avancées pratiques relatives l'une à la gestion dynamique de hiérarchies pour la classification de documents, toutes basées sur l'utilisation ou l'adaptation d'OntoExplo : utilisation de DynamO pour enrichir des hiérarchies utilisées dans OntoExplo, l'utilisation des outils et métriques d'OntoExplo pour comparer des ontologies entre elles et avec des textes au cas de l'étude de communautés virtuelles (OntoExplo + projet CNES) ;

réutilisation et adaptation d'OntoExplo pour assurer l'annotation sémantique de documents pédagogiques à l'aide d'une ontologie de thème et d'une ontologie documentaire ; adaptation d'OntoExplo pour permettre une interrogation des collections annotées par l'utilisation du formalisme des graphes conceptuels

- Une meilleure intégration des nouveaux membres de IC3 par la définition de problématiques de recherche communes et la mise en place de moyens pour une collaboration à plus long terme avec SMAC et SIG-EVI, grâce au financement du projet Dynamo par l'ANR

1) Evaluation des résultats de Dynamo : apport des SMA à la maintenance d'ontologies

Participants SMAC-IC3

Kévin Ottens, Pierre Glize, Valérie Camps, Nathalie Aussenac-Gilles, Nathalie Hernandez

Objectif

L'objectif est d'étudier les mécanismes possibles pour aider à la construction et à la maintenance de ressources termino-ontologiques lorsque celles-ci sont utilisées pour naviguer dans une collection documentaire. J. Mothe a montré avec le système DocCube l'intérêt d'une représentation ontologique pour définir des visualisations graphiques d'aide à la navigation. N. Hernandez a ensuite étudié la question de la maintenance de ces ontologies dans sa thèse (Hernandez, 2005). Elle a développé le logiciel OntoExplo pour gérer l'ensemble du processus, de la mise au point d'une ontologie adaptée à une collection jusqu'à la navigation dans la collection via l'ontologie. K. Ottens, dans sa thèse, a défini un système multi-agent adaptatif permettant d'extraire des éléments d'ontologies et surtout de faciliter la mise à jour d'ontologies à partir de textes (Ottens, 2007).

Notre étude dans le cadre de ce BQR consistait à :

- Évaluer les hiérarchies de concepts créées à partir de Dynamo en les comparant à des ressources existantes
- Intégrer les ressources créées à partir de Dynamo dans OntoExplo

Réalisation

Afin de comparer des ontologies construites par Dynamo à des ontologies préexistantes, un corpus de référence du domaine de l'astronomie a été fourni en entrée de l'outil Dynamo. Nous avons montré que l'utilisation du système Dynamo permettait à un utilisateur de construire une ontologie équivalente à l'ontologie de référence 4 fois plus rapidement que sans outil dédié. A partir de l'analyse syntaxique du corpus de référence, le système propose automatiquement à l'utilisateur des concepts et des relations hiérarchiques entre concepts. L'utilisateur peut à tout moment intervenir pour améliorer les placements automatiques proposés par le système. Dans notre expérimentation, 4 phases d'intervention seulement ont été nécessaires pour obtenir l'ontologie finale. L'ontologie construite a ensuite été utilisée dans le système de recherche d'information OntoExplo. Ce système permet d'accéder aux documents en naviguant dans l'ontologie qui a été utilisée pour annoter le corpus. Nous avons testé le système sur 2 versions de l'ontologie construite par Dynamo :

- Tout d'abord, nous avons considéré l'ontologie construite automatiquement sans intervention de l'utilisateur. Dans ce cas de figure, lors d'une recherche d'informations, l'utilisateur a besoin de temps pour appréhender l'organisation des concepts dans l'ontologie, notamment les concepts intermédiaires créés par le système pour structurer l'ontologie.
- Ensuite, nous avons considéré l'ontologie finale obtenue après l'intervention de l'utilisateur. Dans ce cas-là, l'utilisateur retrouve plus facilement l'information car il a participé à la structuration de l'ontologie.

Des évaluations en termes de rappel et de précision sont encore nécessaires. Néanmoins, cette première expérimentation permet d'envisager l'intégration de l'outil Dynamo dans un système de

recherche d'information reposant sur une ontologie, car il facilite la construction de cette dernière par l'interaction avec l'utilisateur qui va rechercher l'information.
Ce travail a été soumis pour publication dans une revue internationale.

2) Associer des méta-données à des ressources pédagogiques

Participants SIG-EVI, IC3

Alberto Asencio, Josiane Mothe, Bachelin Ralalason et Nathalie Hernandez

Objectif

L'objectif de cette partie du projet visait à étudier le caractère ré-utilisable du projet OntoExplo, développé dans le cadre de la thèse Nathalie Hernandez (Hernandez, 2005).

Le travail de thèse s'est attaché en particulier à proposer un modèle d'indexation et de navigation dans une collection de documents, à base d'une ontologie de domaine et d'une ontologie de tâche, en se basant sur une analyse linguistique des documents et un outil sous licence. Le cadre applicatif concernant des documents scientifiques dans le domaine de l'astronomie.

Notre étude dans la cadre de ce BQR s'intéressait :

- à montrer la ré-utilisabilité dans un cadre complètement différent. Le cadre choisi concerne l'indexation et la recherche de ressources pédagogiques,
- à étudier une version « libre » du processus, impliquant donc un travail sur la modularité des codes sources impliqués dans l'application initiale,
- à réaliser une analyse d'impact concernant l'adaptation des codes, en particulier lors de la mise à jour de la collection de documents

Réalisation

Ces objectifs ont été remplis. Nous avons proposé les spécifications d'une plate-forme permettant la collection, l'indexation et la recherche d'objets pédagogiques réalisés pour des formations à distance, en visant leur réutilisation dans de nouvelles formations.

Le développement selon ces spécifications est pratiquement terminé. Nous avons choisi une plate forme dédiée à l'apprentissage en hôtellerie/restauration pour valider l'approche. Le système « OntoExplo » (Hernandez, 2005) a été utilisé pour mettre en place la modélisation du contexte sémantique de la recherche des objets pédagogiques. Ce système représente l'information en utilisant deux ontologies, l'une liée aux thèmes abordés dans les objets et l'autre aux méta-données intéressant l'utilisateur. Les ontologies sont implémentées par le langage OWL (OWL). Un profil d'application issu du standard LOM(LOM) spécifie la syntaxe et la sémantique des méta-données décrivant les objets pédagogiques. Cette approche est enrichie par l'ontologie du domaine, complétant ainsi le contexte de la recherche.

La navigation dans les ontologies est implantée en utilisant le module d'exploration de corpus d' « OntoExplo », tandis que le module d'indexation est chargé d'annoter automatiquement les objets pédagogiques à partir des concepts et instances des deux ontologies. La génération du format nécessaire pour l'indexation sémantique a été développée en utilisant une procédure d'analyse à cascade, qui inclut l'étiqueteur morphosyntaxique et l'analyseur de surface de l'outil « TreeTagger » (<http://www.ims.uni-stuttgart.de/projekte/complex/TreeTagger/>), ainsi que du code Perl pour filtrer les résultats.

D'autre part, la technologie « XForms » (<http://www.w3.org/MarkUp/Forms/>) est utilisée pour implémenter l'interface qui permet aux utilisateurs l'introduction des méta-données par le biais d'un formulaire. « XForms » permet la validation des champs requis par le profil d'application et la génération d'un document XML (une instance du modèle issu du standard LOM) à partir duquel est réalisée l'annotation des méta-données. De plus, l'utilisation de XML pour définir l'interface des métadonnées, en séparant la forme, le contenu et la logique, facilite les modifications et la réutilisation du profil d'application.

Après l'indexation, les résultats sont stockés dans une base de données. Finalement, pour adapter « OntoExplo » à un environnement Web, une *applet* Java implémentant l'exploration des objets pédagogiques et un *servlet* Java réalisant l'indexation ont été développés.

La modularité du code et son adaptation lors de mise à fait ressortir le besoin d'être capable d'indexer un document isolé plutôt qu'une collection entière. Ceci pose le problème du paramétrage de l'indexation qui se base sur le principe tf.idf qui prend en compte la fréquence d'apparition absolue d'un terme

4) Comparaison d'ontologies, de thésaurus et de collection de textes à l'aide d'OntoExplo

Participants IC3

Bernard Rothenburger, Nacim Chikhi et Nathalie Hernandez

Objectif

Le but de cette partie concerne la comparaison de connaissances contenues dans des ontologies et/ou des textes. La comparaison d'éléments d'ontologies (concepts et relations) ainsi que la mesure de la correspondance entre une ontologie et les connaissances contenues dans un corpus de textes a été étudiée indépendamment par les équipes ISYCOM, SIG et IC3. O. Haemmerlé a montré que les graphes conceptuels flous pouvaient servir de base pour associer des ontologies. Au sein d'IC3, des métriques définies pour la comparaison d'ontologies ont été appliquées pour mesurer la distance entre deux concepts au sein d'une ontologie. Un certain nombre d'outils ont été proposés par B. Rothenburger dans ce cadre (Rothenburger et al. 2007). OntoExplo comporte également des modules dédiés à ces comparaisons (Hernandez, 2005). Un des but de cette partie est de comparer ces deux ensembles d'outillage afin d'évaluer leur complémentarité.

Ces résultats sont mis en œuvre dans le cadre d'une action de R&T du CNES. L'objectif d'une composante de ce projet est de définir les moyens d'assurer l'intelligibilité des archives de données scientifiques qui sont produites dans les cadre des 'Observatoires Virtuels'. On a alors besoin de mesures telles que celles étudiées dans cette partie pour juger des différences terminologiques et conceptuelles entre communautés scientifiques, virtuelles et humaines, du domaine spatial. Ces communautés produisent des ontologies, des corpus et des données scientifiques. Par le biais de la comparaison d'ontologies et de textes, on souhaite savoir si une communauté scientifique comprend une ontologie, si deux ontologies de deux communautés sont proches, si les sites web de ces communautés sont comparables via la terminologie utilisée ou les concepts présents, etc. Les outils disponibles dans OntoExplo seront adaptés pour être utilisés dans ce contexte.

Réalisation

L'objectif final du projet avec le CNES est de prévenir les risques d'inintelligibilité des descriptions d'archives de données scientifiques lorsque celles-ci sont accédées par des communautés d'utilisateurs différentes de celles prévues lors de la constitution de l'archive.

Une première tâche a été de décrire les moyens d'identifier des communautés. Nous nous sommes pour cela inspiré des outils de 'community mining'. Deux directions originales ont été privilégiées dans ce cadre : l'utilisation de méthodes de classification non encore utilisées dans ce cadre : la Non-negative Factorisation Method et la combinaison de l'analyse de liens et de l'analyse de contenu. Cette double approche a été mise en œuvre pour l'identification de communautés dans le domaine du 'space plasma' à partir d'extraits du fond astrophysique de l'Astrophysics Data System (ADS) de la NASA

La deuxième étape en cours consistera à qualifier du point de vue des connaissances les différentes communautés identifiées. Cette étape correspond à l'objectif décrit ci-dessus.

Un premier test a été effectué pour cette étape en utilisant :

- un ensemble d'ontologies concernant entre autre les instruments pour l'astrophysique
- deux fichiers de documents issus de l'ADS à partir de la requête "measure and instrument", le premier avec des abstracts antérieurs à 1985 le second avec des abstracts postérieurs à 2005.

Perspectives

- *Le projet DYNAMO (DYNAMic Ontology for information retrieval)* doit permettre d'améliorer la recherche d'information et la satisfaction des utilisateurs en prenant en compte la dynamique du contexte de recherche : les évolutions de la collection de documents consultés, des connaissances et de la terminologie du domaine, et enfin des besoins des utilisateurs. Son objectif est de concevoir une approche méthodologique et un ensemble d'outils logiciels qui prennent ainsi en compte la construction et la maintenance de ressources ontologiques à partir de documents et l'utilisation de ces ressources pour une indexation sémantique facilitant la recherche d'information. Deux approches complémentaires seront développées pour la maintenance : l'une, classique, s'appuiera sur une extraction terminologique et une maintenance manuelle ; l'autre fera appel à une extension du système multi-agent DynamO développé par K. Ottens pendant sa thèse. Les partenaires du projet sont les trois équipes de l'IRIT impliquées dans le BQR (SMAC, SIG-EVI et IC3), ainsi que deux entreprises (ARTAL Technologies et ACTIA) et deux autres laboratoires de recherche (LALIC et « Préhistoire et Technologie »). ACTIA et Préhistoire et Technologie fournissent des terrains d'application de recherche d'information dans des collections documentaires spécialisées régulièrement mises à jour. ACTIA et LALIC apportent également des compétences en annotation sémantique de documents, alors qu'ARTAL fournit une compétence technologique pour développer les logiciels prévus. Le projet va commencer en janvier 2008 pour une durée de 3 ans et permettra de financer deux thèses à l'IRIT.
- *Evolution de la plate-forme adaptée d'OntExplo permettant la collecte, l'indexation et la recherche d'objets pédagogiques.* Les spécifications de l'application vont donner lieu à un rapport interne IRIT. Nous envisageons également une publication dans un congrès dans le domaine du e-learning. Ce travail va être poursuivi selon deux axes :
 - l'évaluation de l'indexation de ressources pédagogique, en lien avec le projet BOPI soutenu par l'Institut Universitaire de Formation des Maîtres de Midi-Pyrénées
 - le problème que nous avons identifié sur l'indexation d'information isolée va être plus spécifiquement être étudié dans le projet Dynamo, soutenu par l'ANR.
- *Annotation de données du domaine spatial à l'aide de méta données :* L'outil AQWEB a été développé conjointement par les équipes INRA et LRI auxquelles appartenait O. Haemmerlé avant son recrutement à Toulouse [GAG 05] [DIB 06]. Cet outil, défini dans le cadre de deux projets RNTL (e.dot et WebContent), permet d'annoter automatiquement des tableaux de données en se fondant sur une ontologie existante. Cette approche sera appliquée aux données scientifiques disponibles dans les observatoires virtuels du domaine spatial dans le projet CNES.
- *Interrogation de méta-données :* cette étude a fait ressortir la nécessité de définir des modes d'interrogation adaptés de ces différentes méta-données. la solution envisagée consiste à s'appuyer sur une représentation des ontologies et des méta-données à l'aide de graphes conceptuels. Cette approche est développée au sein de IC3 par O. Haemmerlé, C. Comparot et N. Hernandez. Il s'agira de générer des graphes conceptuels d'annotations pour interroger des données annotées. Plusieurs applications sont envisagées, dont une en biomédecine.

- *Valorisation des modules d'annotation et d'interrogation dans la plate-forme OSIRIM de l'IRIT.* Cette plate-forme dédiée à la recherche d'informations permet de gérer de gros volumes de données et mettre à disposition de la communauté scientifique des résultats en matière de recherche d'information. Il nous semble indispensable d'étudier comment intégrer nos modules logiciels dans cette plate-forme.

Utilisation des moyens financiers obtenus (IRIT)

IRIT	Année 2007	Les moyens demandés ont permis aux équipes partenaires d'indemniser des stagiaires (eq. SIG-EVI 2000 euros) pour développer les modules logiciels nécessaires et de communiquer les résultats établis lors de conférences (eq. IC3 et SMAC, 4000 euros).
Équipement <i>Postes de travail</i>	0 €	
Fonctionnement <i>Documentation, missions pour analyses de terrain, indemnités de stages</i>	6000 €	
Total (en K€ TTC)	6000 €	

Références

Hernandez N. (2005). Ontologies de domaine pour la modélisation du contexte en Recherche d'Information. Thèse de l'Université Paul Sabatier. Décembre 2005.

Kévin Ottens, Un système multi-agent adaptatif pour la construction d'ontologies à partir de textes, thèse de l'Université Paul Sabatier, 2 octobre 2007

OWL Web Ontology Language. <http://www.w3.org/TR/owl-features/>

Learning Object Metadata (IEEE). http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf

Bernard Rothenburger, Nacim F. Chikhi, Nathalie Aussenac-Gilles. *Comment maintenir l'intelligibilité des archives scientifiques : une solution basée sur l'utilisation des ontologies et des textes.* Dans : *Atelier "ontologies et textes" Associé à TIA 2007, Sophia-Antipolis, 10/10/07*, N. Aussenac-Gilles, P. Laublet (Eds.).

Articles publiés dans le cadre du projet de recherche

Claude Chrisment, Ollivier Haemmerlé, Nathalie Hernandez, Josiane Mothe. *Méthodologie de transformation d'un thésaurus en une ontologie de domaine.* Dans / In : *Revue d'Intelligence Artificielle*, Hermès Science Publications, 2008 (à paraître).

Nacim Fateh Chikhi, Bernard Rothenburger, Nathalie Aussenac-Gilles. *A Comparison of Dimensionality reduction techniques for Web Structure Mining.* Dans / In : *IEEE/WIC/ACM International Conference on Web Intelligence, Silicon Valley (USA), 02/11/2007-05/11/2007*, Li Tsau Young (T.Y.) (Eds.), IEEE Computer Society, (support électronique), 2007.

Kurt Englmeier, F. Murtagh, Josiane Mothe. *Domain Ontology: Automatically Extracting and Structuring Community Language from Texts.* Dans: *IADIS Applied Computing, Salamanca, Spain, Espagne, 18/02/2007-20/02/2007*, IADIS (Eds.), IADIS Digital Library, p. 59-66, 2007.

Ollivier Haemmerlé, Nathalie Hernandez, Josiane Mothe. *Graphes conceptuels et ontologies de domaines pour la recherche d'information.* Dans / In : *Atelier Modélisation des Connaissances, 7ème conférence Extraction et Gestion des Connaissances, Namur, Belgique, 23/01/2007-23/01/2007, 2007 (à paraître).*

Nathalie Hernandez, Josiane Mothe, Bachelin Johnn Victorino Ralalason, Andriantiana Bertin Ramamonjisoa , Patricia Stolf. *A Model to Represent the Facets of Learning Objects,* Dans : *Interdisciplinary Journal of Knowledge and Learning Objects, Informing Science Institute, Santa Rosa - USA, V. 4, 2008 (à paraître).*

Kévin Ottens, Nathalie Aussenac-Gilles. *Un algorithme multi-agent de classification pour la construction d'ontologies dynamiques.* Dans : *EGC 2007 7e Journées Francophone Extraction et Gestion des Connaissances, Namur (B)*, G. Venturini (Ed.), 0-0, 2007.

Kévin Ottens, Nathalie Aussenac-Gilles, Marie-Pierre Gleizes, Valérie Camps. *Dynamic Ontology Co-Evolution from Texts: Principles and case Study.* Dans / In : *International Workshop on Emergent Semantics and Ontology Evolution at ISWC 2007 (ESOE 2007), Busan (South Korea), 12/11/2007*, Luke Chen, Peter Haase, Andreas Hotho, Ernie Ong (Eds.), (en ligne), novembre 2007 (à paraître). <http://km.aifb.uni-karlsruhe.de/ws/esoe2007/>

Annexe : ressources disponibles au début du projet

Le projet s'est appuyé sur les outils suivants déjà développés par les équipes partenaires

- **Dynamo** est un logiciel innovant qui exploite les résultats d'une analyse syntaxique de textes pour aider un cogniticien à construire une ontologie. Son originalité est de se baser sur des agents adaptatifs et de faciliter des mises à jour interactives du modèle [SIE 06]. Dans sa version actuelle, il permet de construire des hiérarchies de concepts, et l'analyseur syntaxique utilisé, Syntex, lui fournit un réseau de termes [OTT 07a]. A partir de chaque terme proposé par Syntex, un agent adaptatif « concept » est défini. Il exploite les liens entre termes au sein du réseau Syntex et leurs contextes d'utilisation pour se comparer aux autres agents. Ainsi, il se place en relation avec certains d'entre eux, se cherche un concept père et des fils. L'interface de consultation du réseau d'agents présente la hiérarchie au cogniticien qui peut l'ajuster, et le réseau d'agent s'adapte pour produire un nouvel arbre [OTT 07b].
- **OntoExplo** de N. Hernandez : OntoExplo est une plate-forme qui permet de gérer une ontologie en vue de guider la consultation d'une collection de documents [HER 05b]. Pour simplifier la tâche de construction de l'ontologie et assurer son adéquation à la collection, y compris lorsqu'elle est enrichie ou modifiée, OntoExplo propose deux approches de construction [HER 05a] : l'importation d'ontologies ou de thésaurus existants et l'apprentissage de concepts à partir des textes de la collection. Enfin, OntoExplo guide le cogniticien en lui fournissant une métrique qui évalue la pertinence de l'ontologie par rapport au contenu des textes de la collection, et permet de comparer deux ontologies.
- L'outil **AQWEB**, défini dans le cadre de deux projets RNTL (e.dot et WebContent) en collaboration avec l'INRA et le LRI, permet d'associer des méta-données (prises dans une ontologie) à des tableaux de données [GAG 05]. Les valeurs présentes dans les colonnes sont identifiées, ce qui permet de déterminer la nature des colonnes (leurs titres), pour finalement permettre de reconnaître les relations portant sur ces colonnes.

Les projets complémentaires à ce projet, menés séparément par les équipes impliquées, et dont OntoExplo permettra de fédérer les résultats :

- **CNES** : le CNES finance une étude de Recherche et Technologie confiée à l'INRIA (B. Rothenburger), à l'IRIT-CSC et au laboratoire CLL-ERSS. L'objectif en est de montrer l'apport des techniques linguistiques et statistiques d'analyse de textes à deux problématiques de gestion de connaissance pour les chercheurs du domaine spatial : l'accès aux données scientifiques sur les observatoires virtuels par des chercheurs de communautés scientifiques différentes ; repérage du risque et des changements de sens dans des corpus, et des ontologies. Ce travail fait l'objet de la thèse de N. Chikhi, équipe CSC
- **WebContent** : projet RNTL faisant collaborer 19 partenaires institutionnels et privés. Le but du projet WebContent est de réaliser une plate-forme logicielle capable d'accueillir tous les outils nécessaires pour exploiter efficacement et étendre le Web Sémantique. La plate-forme vise initialement un groupe d'applications autour de la veille technologique (veille économique dans le domaine aéronautique, risque microbiologique et chimique dans le domaine alimentaire, surveillance des événements sismiques).
- **DynamO** : ce projet, soutenu à son début par le programme interdisciplinaire TCAN, rassemble les équipes SMAC et CSC de l'IRIT, l'entreprise ACTIA, les URFIST nationales autour de la problématique de la définition d'ontologies dynamiques pour la recherche d'information et l'annotation sémantique. Le caractère dynamique consiste à définir les moyens d'adapter facilement l'ontologie au contenu des documents à annoter, de manière à couvrir le registre terminologique et conceptuel des documents. Une demande de financement par l'ANR, qui a échoué en 2006, sera faite en 2007. Le cœur du projet est le travail de thèse de K. Ottens, équipes SMAC et CSC.