



HAL
open science

La fiche terminologique comme outil de valorisation des connaissances : identification, structuration et modélisation d'un lexique pour une terminologie de la mine

Oussila Lafri

► To cite this version:

Oussila Lafri. La fiche terminologique comme outil de valorisation des connaissances : identification, structuration et modélisation d'un lexique pour une terminologie de la mine. Sciences de l'information et de la communication. 2017. dumas-01561481

HAL Id: dumas-01561481

<https://dumas.ccsd.cnrs.fr/dumas-01561481>

Submitted on 12 Jul 2017

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.



Oussila LAFRI
Master en Sciences de l'Information et de la Documentation
Première année

Mémoire de stage
Mission effectuée du 3 avril au 9 juin 2017

au
Laboratoire GERiico
Université Charles-de-Gaulle – Lille 3
Villeneuve d'Ascq

**La fiche terminologique comme outil de valorisation des
connaissances : identification, structuration et modélisation d'un
lexique pour une terminologie de la mine**

Sous la direction de :
Monsieur Stéphane Chaudiron (Tuteur Professionnel)
Monsieur Bernard Jacquemin (Tuteur Universitaire)

Soutenu le 29 Juin 2017 à l'UFR DECCID-SID
Université de Lille 3, Sciences Humaines et Sociales
BP 60 149, 59 650 Villeneuve d'Ascq Cedex

Année Universitaire 2016/2017

Remerciements

Je tiens, en premier lieu, à remercier Monsieur Stéphane Chaudiron, tuteur professionnel du stage et directeur du laboratoire GERiiCO, pour m'avoir accueillie, et pour avoir su partager ses compétences et sa rigueur scientifique, et aussi pour ses précieux conseils qui ont nourri toute ma réflexion.

Je tiens à remercier mon tuteur pédagogique, Monsieur Bernard Jacquemin, pour son accueil, sa patience, sa disponibilité, ses remarques pertinentes dans des moments de questionnements, de doute.

Je remercie également l'ensemble des enseignants-chercheurs du Master en Sciences de l'Information et Documentation pour m'avoir transmis toute la motivation nécessaire dans cette "aventure" de reprise d'études.

Mes remerciements s'adressent aussi aux étudiants du Master 1 que j'ai côtoyés tout au long de cette année universitaire pour leur soutien, leur aide "technique" et leurs sincères encouragements.

Ce stage a aussi été l'occasion de nouer des liens avec d'autres stagiaires qui ont, d'une certaine manière, contribué à faire avancer mon projet de stage, et pour cela, je tiens à remercier tout particulièrement Rémi Cardon pour toute la disponibilité dont il a fait preuve et pour toute son expertise.

Enfin, mes remerciements s'adressent tout particulièrement à ma famille et mes amis qui ont été de véritables soutiens tout au long de l'année.

Je terminerai en adressant mes profonds et sincères remerciements à Delphine Weill pour tout le travail de relecture fourni et ses conseils avisés.

Avant propos

Entreprendre un master vingt ans après une licence d'histoire est une gageure parce qu'il faut retourner sur les bancs d'une université qui a beaucoup changé. Le changement est d'autant plus important quand le master choisi est dans une discipline qui a été bouleversée par l'arrivée d'internet il y a justement une vingtaine d'années. Cependant, trois éléments ont été facilitateurs dans le cadre de ce master et de ce stage : la première chose est l'interdisciplinarité de la formation qui a permis de s'appuyer sur des connaissances acquises précédemment. La deuxième est l'intérêt porté à un stage qui entre dans la valorisation d'un patrimoine historique. La troisième est l'aptitude à structurer sa pensée qui augmente avec la vie et l'accroissement du volume des connaissances...Quand les données sont de plus en plus nombreuses, il devient impératif d'apprendre à les classer, à les structurer pour les retrouver et les utiliser à bon escient. Or, c'est justement ce dont il s'est agi dans ce stage

Résumé

Les langages documentaires (thésaurus, terminologies, ontologies...) sont des systèmes d'organisation de connaissances principalement utilisés dans des systèmes d'informations dont l'objectif est de collecter, conserver, traiter et partager de la connaissance d'un domaine. Le travail présenté ici s'inscrit dans le cadre d'un projet soutenu par l'Agence Nationale de la Recherche (ANR), Mémo-mines, qui vise à organiser la conversion de la mémoire minière en médiations numériques dans un souci de conservation et de valorisation du patrimoine minier. Dans ce cadre précis, collecter des ressources sémantiques liées à l'univers de la mine est le point de départ d'un des axes du projet de vouloir sauvegarder la richesse linguistique de cet environnement. C'est pourquoi la modélisation doit être au coeur de ce projet car il s'inscrit dans une démarche d'élaboration d'un modèle qui sera à la fois un outil de connaissance et de valorisation.

Mots-clés : langage documentaire, thésaurus, terminologie, ontologie, connaissance, système d'information, patrimoine, mémoire minière, modélisation

Abstract

The documentary languages (thesauri, terminologies, ontologies, etc.) are systems of knowledge organization mainly used in information systems whose aim is to collect, treat, and share about the knowledge of a domain. The work developed here is part of a project supported by the National Research Agency (ANR), Mémo-mines, whose purpose is to organize the conversion of mining memory into digital mediations in a concern for conservation and valorisation of the mining heritage. In that precise field, to collect semantic resources linked to the universe of the mine is the starting point of one of the project line wishing to save the linguistical wealthpart of this environment. That is why modelling has to beat the heart of this project because it is part of a process of elaboration of a model that will, at the same time be a tool of knowledge and of valorisation.

Keywords : documentary language, thesauri, terminology, ontology, knowledge, information system, heritage, mining memory, modelling

Sommaire

Remerciements.....	1
Avant propos.....	2
Résumé	3
Abstract.....	3
Introduction	6
1 - Les lieux et acteurs	8
1 - 1 - Un stage de Master 1 en Sciences de l'Information et de la Documentation (SID), au sein de l'UFR DECCID, Université de Lille 3.....	8
1 - 1 - 1 L'université de Lille 3.....	8
1 - 1 - 2 Le Master 1 SID dans l'UFR DECCID	9
1 - 1 - 3 De l'importance de l'interdisciplinarité : une formation alliant sciences humaines et technologie	10
1 - 2 - Présentation du lieu de stage	11
1 - 2 - 1 De GERiCO à GERiiCO.....	11
1 - 2 - 2 Le laboratoire et son organisation	12
1 - 2 - 3 GERiiCO et ses projets de recherche	13
1 - 3 - Le stage et son objet	14
1 - 3 - 1 La valorisation du patrimoine minier comme un véritable enjeu	14
1 - 3 - 2 Les partenaires du projet	16
1 - 3 - 3 Le stage et ses missions.....	16
2 - Les outils théoriques nécessaires à la modélisation	18
2 - 1 - La science de l'information : de la documentarisation à la redocumentarisation.....	18
2 - 1 - 1 De la naissance de la science de l'information	18
2 - 1 - 2 De l'information aux technologies de l'information	20
2 - 1 - 3 L'information au coeur du stage	22
2 - 2 - L'Organisation des Connaissances, une discipline de la Science de l'Information	23
2 - 2 - 1 De l'organisation des connaissances	23
2 - 2 - 2 Les Systèmes d'Organisation des Connaissances (SOC).....	24
2 - 2 - 3 Outils d'organisation et modélisation de l'information utilisés lors de ce stage	26
2 - 3 - De la linguistique en général et de la terminologie en particulier	27
2 - 3 - 1 La place et le rôle de la terminologie dans les Sciences de l'information et de la communication	27
2 - 3 - 2 La modélisation comme outil de la terminologie.....	29
2 - 3 - 3 La terminologie en pratique.....	30

3 - La modélisation d'un SOC au coeur de la démarche méthodologique	32
3 - 1 - Collecter des ressources sémantiques	32
3 - 1 - 1 Présentation de cette mission de stage.....	32
3 - 1 - 2 Méthodologie de la recherche d'information.....	33
3 - 1 - 3 Résultat d'une collecte non exhaustive	34
3 - 2 - Constitution d'un corpus : description des choix	36
3 - 2 - 1 Le rejet de certaines sources et ses justifications.....	36
3 - 2 - 2 Le corpus et ses traitements	37
3 - 2 - 3 Un cas pratique : Les mots de la mine	38
3 - 3 - Vers la robustesse d'un modèle	39
3 - 3 - 1 De l'importance de modéliser.....	39
3 - 3 - 2 Une analyse lexicale	40
3 - 3 - 3 De la modélisation au modèle : la fiche terminologique.....	42
Conclusion	45
Bibliographie	47
Webographie.....	48
ANNEXES	50

Introduction

La révolution industrielle survenue en Europe dès le XVIII^{ème} siècle a principalement débuté en Grande-Bretagne, en Belgique, en France et plus particulièrement dans le Nord. Cette épopée de l'industrialisation a profondément transformé les espaces en passant d'une société à dominante rurale à une société industrielle. C'est à la même période que se développe dans cette partie de l'Europe l'exploitation du charbon qui marquera profondément ces territoires. En effet, le monde de la mine et de son exploitation sont fortement associés à l'histoire liée à cette révolution industrielle. Cette industrie florissante va peu à peu connaître un véritable déclin industriel dès les années 60 pour totalement disparaître à partir des années 1990 pour la région Nord-Pas de Calais, par exemple. La fin de l'exploitation du charbon a provoqué certes une fragilité économique de ces régions mais elle révèle, par ailleurs une identité "minièr" forte qui est toujours aussi présente. Toutefois, la fin de cette industrie interroge sur les questions de conservation d'un patrimoine industriel riche. C'est toute la problématique de la patrimonialisation de ces anciennes friches, de ces paysages qui est clairement posée à savoir comment conserver, pérenniser un patrimoine. Cette démarche de patrimonialisation se définit d'abord comme un "processus socioculturel, juridique ou politique par lequel un espace, un bien, une espèce ou une pratique se transforment en objet du patrimoine naturel, culturel ou religieux digne de conservation et de restauration"¹. De plus, elle a pour vocation d'insuffler de nouvelles dynamiques territoriales dans des régions extrêmement fragilisées par la désindustrialisation et la fin de cette exploitation minière. Dans ce contexte, les enjeux sont devenus majeurs et notamment en termes de développement touristique et culturel qui sont devenus de "véritables outils de développement territorial."²

Aussi, dans ce contexte de patrimonialisation, l'Agence Nationale de la Recherche soutient un Projet de Recherche Collaborative (PRC) où les laboratoires De Visu (Université de Valenciennes et du Hainaut), GERiiCO (Université de Lille3) et

¹ <https://fr.wikipedia.org/wiki/Patrimonialisation>, [consulté en ligne le 10 juin 2017]

² LUSSO B., « Patrimonialisation et greffes culturelles sur des friches issues de l'industrie minière », *EchoGéo* [En ligne], 26 | 2013, mis en ligne le 19 décembre 2013, consulté le 14 juin 2017. URL : <http://echogeo.revues.org/13645> ; DOI : 10.4000/echogeo.13645

l'INALCO s'impliquent depuis janvier 2017 sur la question de la conservation des patrimoines culturels, et plus spécifiquement du patrimoine minier. C'est justement l'objet de ce stage qui porte sur la problématique de la conservation de faits ou d'événements constituant des patrimoines liés à la mémoire minière et bien trop souvent proches de leur disparition en œuvrant à la conversion de traces mémorielles par des processus innovants.

Plus précisément, ce stage dont la mission est d'identifier des ressources sémantiques liées au domaine de la mine a engagé une réflexion sur la meilleure façon de modéliser des connaissances. Pourquoi modéliser ? Justement parce que la modélisation est une démarche de construction d'un outil de compréhension des connaissances. Modéliser correspond aussi à une phase essentielle qu'est la conception d'un système d'organisation de connaissances. Sa mise en place et sa consolidation doivent aboutir à la structuration d'une terminologie dont l'objectif est d'améliorer ou encore de créer un dispositif d'information dont les éléments sont quelque peu délaissés, abandonnés voire en voie de disparition. L'outil terminologique tient un rôle indispensable pour améliorer et faciliter la communication spécialisée. En effet, les connaissances autour de la mine existent mais ne sont pas répertoriées et identifiées de manière structurée, ce qui peut rendre difficile leur accès. Dans ce cadre, et au vu du sujet proposé pour ce stage, il a fallu réfléchir à une modélisation d'une terminologie à partir de lexiques traitant du domaine de la mine. C'est ce qui a dû être fait avec attention pour que le travail de ce stage soit exploitable *a posteriori*.

Pour tenter de répondre à ces questions, il a été décidé de présenter tout d'abord les lieux et les acteurs de cette formation et de ce stage, parce que les circonstances sont apparues primordiales pour comprendre le questionnement de ce travail. Ensuite, les attendus universitaires ont permis de mettre à plat les outils théoriques utiles à la modélisation car la justesse de la réflexion nécessite une définition de concepts et notions utiles à la discipline et au questionnement de ce stage. Enfin sera reprise la méthodologie qui a permis d'aboutir à une modélisation en vue de la création d'un nouveau système d'organisation des connaissances.

1 - Les lieux et acteurs

1 - 1 - Un stage de Master 1 en Sciences de l'Information et de la Documentation (SID), au sein de l'UFR DECCID, Université de Lille 3

1 - 1 - 1 L'université de Lille 3

On ne peut présenter l'objet de ce stage sans introduire le diplôme dans le cadre duquel il a lieu. Ce stage de Master 1 en Information et Documentation est le point final d'une formation dispensée par l'UFR DECCID au sein de l'Université de Lille 3. Cette vénérable institution lilloise a été créée en 1896 grâce notamment à une convention entre la ville de Lille et l'État qui a permis à l'université de se constituer progressivement par la mise en place d'institutions spécialisées telles que les instituts de sciences naturelles, de physique, de chimie, d'archéologie classique et les facultés de droit et de lettres. Cependant, avec la loi du 12 novembre 1968 portant sur l'orientation de l'enseignement supérieur, l'université se restructure dans un contexte politique mouvementé par les événements de mai 1968 et permet la création de l'Université des Sciences et Techniques, plus communément appelé Lille 1, l'Université de Droit et de la Santé, Lille 2, et enfin l'Université des Sciences Humaines, Lettres et Arts, Lille 3.

Située dans la métropole lilloise, elle est, aujourd'hui, la plus grande université située au Nord de Paris et elle recouvre comme champs disciplinaires les sciences humaines et sociales, l'économie et gestion, ou encore les lettres, les arts et les langues.

L'Université de Lille 3 arrivée sur le site de Pont de Bois à Villeneuve d'Ascq en 1974, compte aujourd'hui près de 20 000 étudiants à qui sont proposés des enseignements riches et variés représentés par plus de 120 diplômes.

Elle est devenue un pôle universitaire fortement caractérisé par son dynamisme grâce à une activité de recherche riche, portée et animée par près de 500 enseignants-chercheurs. Le calendrier politique étant à la re-concentration des universités, une fusion des universités lilloises et de six grandes écoles est actuellement à l'ordre du jour afin d'augmenter encore la visibilité des UFR au niveau international et la force de recherche des laboratoires qui constitueront le fer de lance de la future Université Lille Nord de France.

1 - 1 - 2 Le Master 1 SID dans l'UFR DECCID

Le master SID est une des formations proposée par l'UFR DECCID. Cette Unité de Formation et de Recherche a été créée en 2012 en regroupant cinq départements. Sa spécificité est d'offrir aux étudiants des formations à la fois théoriques et professionnalisantes avec l'intervention de près de 400 intervenants professionnels.

Le département Culture offre aux étudiants, au travers de ses enseignements et sur l'ensemble de ses formations, la possibilité de comprendre les enjeux de la production et de la diffusion de l'art et de la culture dans nos sociétés contemporaines. Infocom est une antenne de Lille 3 basée à Roubaix depuis déjà plus de 20 ans et, qui forme des professionnels de l'information et de la communication. Le département de Sciences de l'éducation dispense des enseignements didactiques, pédagogiques et sur le fonctionnement des institutions de l'éducation et de la formation. C'est, comme pour le Département en Sciences de l'Information et de la Documentation, le croisement de plusieurs disciplines qui fait la force et la richesse de cette spécialité. Le département Sociologie et Développement social propose, quant à lui, des formations où se croisent un enseignement sociologique proprement universitaire et une démarche professionnelle liée aux questions du développement social.

Enfin, le département en Sciences de l'Information et de la Documentation représente la cinquième composante de l'UFR. Il offre plusieurs parcours de formations professionnalisantes telles les DEUST ou les licences professionnelles mais aussi des formations plus générales comme la licence documentation, par exemple. Ce département a pour objectif de former des spécialistes de l'information et de la documentation qui participeront à la médiation des savoirs dans un contexte fortement dominé par le numérique. Or, le numérique bouleverse profondément les cadres traditionnels de médiation des savoirs et le département propose justement des formations qui suivent cette évolution technique et, permettent aux étudiants de s'adapter aux techniques et technologies qui seront bien souvent désormais au coeur de leur futurs métiers. Mais le département n'est pas un département purement technique (d'où son insertion au sein d'une Université de Lettres et Sciences Humaines), et c'est son approche interdisciplinaire, et la diversité des parcours des étudiants qui offrent l'opportunité de développer des connaissances dans le domaine et de s'initier aux techniques des pratiques documentaires.

1 - 1 - 3 De l'importance de l'interdisciplinarité : une formation alliant sciences humaines et technologie

On l'a dit, la discipline allie donc des compétences disciplinaires liées, entre autres, à la documentation et à une information traditionnelles et des compétences plus techniques liées au développement exponentiel du média numérique. Cette interdisciplinarité est au cœur de la création de la discipline dans bien d'autres domaines, comme peut l'exprimer Jean Meyriat :

« Cette interdiscipline est une discipline : c'est ce qu'entend affirmer le fait même de lui donner un nom. Autrement dit, il y a une problématique propre à l'information et à la communication, et en se laissant guider par elle, on doit parvenir à dégager une théorie de ces phénomènes qui serait autre chose qu'une juxtaposition des éclairages latéraux fournis par d'autres disciplines. Cette interdiscipline est plurielle, comme le signifie son nom « les sciences de l'information et de la communication ». Il y a pluralité, d'objets, d'objectifs théoriques, de finalités professionnelles. Mais cette pluralité est interne à une unité que nous avons voulue affirmer [...] »

Jean Meyriat, Président d'honneur à la
SFSIC,

Sans vouloir retracer l'histoire de la discipline, il est, toutefois, judicieux de rappeler que les Sciences de l'Information et de la Communication sont nées de la volonté d'une poignée d'hommes tels que Robert Escarpit, Jean Meyriat, Roland Barthes d'obtenir une reconnaissance universitaire de cette discipline qu'ils arracheront en 1975 par la création de la 52ème section qui deviendra la 71ème en 1981. C'est donc une « discipline jeune (trente ans) grandie en France sous le signe spécifique de l'alliance de l'information et de la communication » (Jeanneret, Ollivier, 2004 : 15)³. Cependant, cette place réservée à la Science de l'Information et de la Communication (SIC) n'exclut pas la question essentielle et tant controversée de la documentation. Jean Meyriat a constamment défendu l'idée que la documentation est une notion totalement

³ JEANNERET Yves (coord.) et OLLIVIER Bruno (coord.) (2004). Les Sciences de l'Information et de la Communication : savoirs et pouvoirs. Hermès , mai, n° 38, 256 p. [consulté en ligne le 02 juin 2017]

inhérente à la science de l'information. C'est justement cette notion de documentation qui voit s'ajouter une compétence fortement liée à l'aspect technique de cette science même si elle ne doit pas être réduite à une activité purement technologique. Cela explique que la prise en charge et le traitement d'un document suppose d'avoir recours à un certain nombre de techniques et de savoir-faire.

1 - 2 - Présentation du lieu de stage

L'année de master 1 se clôture par un stage qui s'effectue à partir du mois d'avril et ce, pour une durée de six semaines. Pour ma part, il s'est déroulé au sein du laboratoire GERiiCO sur un projet de recherche et de valorisation du patrimoine minier.

1 - 2 - 1 De GERiCO à GERiiCO

GERiiCO n'est autre que l'acronyme du Groupe d'Etudes et de Recherche Interdisciplinaire en Information et COmmunication. Le laboratoire est reconnu comme un pôle de recherche à vocation internationale basé au sein de l'Université de Lille 3.

Depuis la naissance des Sciences de l'Information et de la Communication, le positionnement scientifique des laboratoires de recherche se pose la question de la légitimité de la discipline caractérisée par son interdisciplinarité et explique, d'une certaine manière, l'évolution de GERiiCO. Il est vrai que l'histoire du laboratoire se caractérise par de profonds changements survenus au sein même de la discipline. C'est sans doute une des raisons qui explique le changement de dénomination du laboratoire qui passe du Groupe d'Etude et de Recherche Interdisciplinaire en Communication à GERiiCO en intégrant comme nouvelle thématique de recherche : l'information.

Aujourd'hui, ce pôle de recherche, dirigé par Stéphane Chaudiron, rassemble, au 1er janvier 2017, 36 enseignants-chercheurs titulaires dont 7 professeurs et 29 maîtres de conférences, 3 professeurs émérites et 38 doctorants qui portent une attention particulière aux questions relatives à la médiation des savoirs dans notre société contemporaine considérablement dominée par le numérique.

Grâce au concours et au soutien de l'Université, le laboratoire publie une revue "Etudes de communication" à la fois en version papier et numérique. "L'utilisation du numérique contribue donc à une hybridation et une flexibilité des espaces de diffusion

participant à la pérennité de la revue *Études de communication*".⁴ Elle est également accessible sur le portail revue.org qui offre un accès intégral aux articles de plus de deux ans, et sur Cairn. Ses premières publications se sont faites dès 1982 sous le nom de Bulletin du CERTE (Centre de Recherche en Technique d'Expression), dont le nom a été changé en CERTEIC (Centre de Recherche en Techniques d'Expression en Information et Communication) en 1986 avant de devenir en 1992 ce qu'elle est, aujourd'hui, une revue qui a pour ambition de valoriser les travaux des chercheurs et de pouvoir les faire bénéficier d'une réelle visibilité internationale.

1 - 2 - 2 Le laboratoire et son organisation

Le laboratoire, par son approche interdisciplinaire, conjugue plusieurs compétences qui sont au carrefour des sciences de l'information et de la communication, des sciences sociales, des sciences du langage ou encore des sciences cognitives. GERiiCO et son équipe de chercheurs contribuent, par leur savoir-faire et leur expertise à valoriser de véritables thématiques de recherche dans les domaines suivants⁵ :

- L'analyse des actions et des phénomènes d'information et de communication dans les organisations (entreprises, musées, collectivités territoriales, hôpitaux...)
- L'étude de l'impact des TIC sur les médias traditionnels et l'émergence des nouveaux médias
- L'évaluation des logiciels de traitement avancé de l'information (recherche, filtrage, traduction automatique, extraction de termes, outils d'organisation de l'information...)
- L'analyse des usages et des pratiques informationnelles dans différents milieux (professionnel, scolaire...)
- L'analyse des mutations socio-économiques dans le secteur de l'information scientifique et technique
- L'étude des pratiques et des dispositifs de veille dans les entreprises (notamment le secteur des PMI-PME)

⁴ Bruno Chaudet, Romain Huet, Solen Tenier, Nina Malledant, Rachel Thoa et Damien Sauvaire, « La revue *Études de communication* », *Revue française des sciences de l'information et de la communication* [En ligne], 3 | 2013, mis en ligne le 04 novembre 2014, consulté le 03 juin 2017. URL : <http://rfsic.revues.org/655> ; DOI : 10.4000/rfsic.655

⁵ <http://GERiiCO.recherche.univ-lille3.fr/index.php?page=expertise-et-valorisation> - [consulté en ligne le 03 juin 2017]

- L'analyse des mutations des dispositifs de médiation et communication dans l'espace public urbain (médias associatifs, interprétation du patrimoine urbain, stratégies communicationnelles et renouvellement urbain)

GERiiCO s'établit et s'organise selon quatre axes majeurs de recherche, sous la responsabilité d'un enseignant-chercheur, qui permet la coordination des activités de chacun des chercheurs. Les orientations du laboratoire se déclinent comme suit :

- Axe 1 : Information et communication dans les organisations
- Axe 2 : Culture et médias dans l'espace public
- Axe 3 : Innovation par l'usage et dispositifs numériques
- Axe 4 : Circulation de l'information et organisation des connaissances

L'activité de recherche du laboratoire se réalise très souvent grâce à une riche collaboration et un long travail partenarial.

1 - 2 - 3 GERiiCO et ses projets de recherche

Dans un souci de développer ses activités de recherche, le laboratoire GERiiCO développe de véritables partenariats très souvent en répondant à des appels à projets. L'essentiel de ses projets reçoivent des soutiens financiers principalement de la région des Hauts de France, de l'Agence Nationale de la Recherche (ANR), des Ministères chargés de l'enseignement supérieur et de la culture...

Cependant, les partenaires du laboratoire ne sont pas uniquement des institutionnels car des collaborations peuvent se nouer également avec des acteurs du monde industriel comme des éditeurs de logiciels, par exemple ou encore des associations professionnelles.

Aujourd'hui, l'activité de recherche de GERiiCO se déploie sur les axes cités ci dessus, en menant des projets de niveau régional, national, et aussi européen. Sans vouloir énumérer l'ensemble des programmes de recherche menés, ceux qui suivent ne peuvent être passés sous silence :

- **Muséo** + est un projet collaboratif de conception d'une application ludo-éducative accessible par le biais d'une tablette numérique dont l'ambition est de favoriser l'accès à la culture pour tous. C'est un projet soutenu par la Région Nord-Pas de Calais dans le cadre du programme "Chercheurs-citoyens" mais

aussi par d'autres financements tels que Pictanovo (Pôle image de la région). Cette action de recherche est le fruit d'un travail partenarial entre : les laboratoires De Visu de l'Université de Valenciennes, GERiiCO et SCALAB de l'Université de Lille 3, et l'association Signe de sens.

- **Le Vase qui parle** est un projet soutenu par le Ministère de la Culture qui reprend dans un très grand format cette fameuse céramique grecque conservée au Palais des Beaux-arts de Lille. C'est un programme retenu suite à l'appel à projet "services culturels numériques innovants". Ce travail, purement collaboratif, a réuni divers partenaires comme le Service Commun de la Documentation (SCD), le Learning Center Archéologie/Egyptologie, le Département "Langues et cultures antiques" et les laboratoires HALMA-IPEL (UMR 8164) et GERiiCO.

Enfin, il faut citer le projet sur lequel j'ai été amenée à travailler : actuellement, suite à un appel à projet générique de l'Agence Nationale de la Recherche (ANR), un Projet de Recherche Collaboratif (PRC) a été retenu sur la "Conversion des traces mémorielles en médiations numériques : le cas de la mémoire minière".

1 - 3 - Le stage et son objet

Le stage s'est donc déroulé au sein du laboratoire GERiiCO sur ce projet retenu en 2016 par l'ANR et qui porte sur la valorisation du patrimoine minier.

1 - 3 - 1 La valorisation du patrimoine minier comme un véritable enjeu

L'industrie minière est un des secteurs les plus caractéristiques de la révolution industrielle du XIXème siècle en Europe de l'Ouest, elle a perduré encore jusqu'en 2015, année où la dernière mine de charbon en exploitation ferme, dans le Yorkshire en Angleterre. En France, la dernière mine française de charbon a fermé en avril 2004 à La Houve en Lorraine.⁶

Le patrimoine industriel minier est, aujourd'hui, un véritable enjeu territorial parce qu'il participe à la revalorisation d'un territoire d'abord sinistré par les fermetures industrielles dans un contexte de développement économique fortement axé sur le

⁶ https://fr.wikipedia.org/wiki/Mines_de_charbon_de_France#Histoire [consulté en ligne le 03 juin 2017]

développement touristique et culturel. Concernant la région des Hauts de France, le bassin houiller a considérablement marqué le paysage à la fois économique et naturel de la région. Aussi, depuis 2012, le bassin minier est inscrit sur la liste du patrimoine mondial de l'UNESCO où il "a été distingué au titre de « paysage culturel», « œuvre conjuguée de l'homme et de la nature » selon les termes de la Convention du patrimoine mondial."⁷

Dans ce contexte de volonté de valoriser ce patrimoine minier, l'ANR a octroyé des fonds à un projet de recherche sur la conversion des traces mémorielles en médiations numériques. Ce projet a pour intitulé : Mémo-Mines. Ce programme vise à préserver la mémoire des mines du Nord en œuvrant sur le développement d'un tourisme grand public, mais aussi d'un autre type de diffusion mémorielle destinée à un public scientifique en mettant à sa dispositions des ressources. La problématique d'ensemble du projet consiste, donc, à organiser la conservation de tout ce qui relève de ce patrimoine minier avec pour objectif final de rendre disponible et accessible le corpus d'archives sous format numérique. Le projet se propose de relever les cinq défis qui suivent : ⁸

- la constitution/production de la mémoire minière
- l'explicitation de l'univers de connaissance de la mine, la définition et l'élaboration d'un métalangage
- la réalisation des archives audiovisuelles de la mémoire minière sous forme d'un portail web
- la conception et la mise en place de dispositifs techniques innovateurs pour l'exploitation – la remédiatisation
- la spécification et la mise en place d'une « plateforme de travail » intégrant les différents outils

Ces travaux de recherche ont débuté en janvier 2017 et se déroulera sur 4 années. Il est spécifiquement lié à la région Nord-Pas de Calais mais sera fortement ancré dans un contexte européen, dans la mesure où les régions voisines (Ruhr, Wallonie, Angleterre) seront concernées.

⁷ <http://www.bassinminier-patrimoine mondial.org/> [consulté en ligne le 03 juin 2017]

⁸ <http://www.agence-nationale-recherche.fr/?Projet=ANR-16-CE38-0001-> [consulté en ligne le 03 juin 2017]

1 - 3 - 2 Les partenaires du projet

Ce projet ambitieux s'inscrit dans le cadre d'un Projet de Recherche Collaborative (PRC) octroyé par l'ANR qui réunit l'INALCO, De Visu et GERiiCO. L'Agence Nationale de la Recherche est un établissement public à caractère administratif placé sous la tutelle du ministre chargé de la recherche et dirigé par Michael Matlosz, membre de l'Académie des technologies. Sa mission principale est d'organiser la mise en œuvre du financement de la recherche sur des projets scientifiques en France. L'ANR⁹ se donne pour mission de contribuer au développement des sciences et des technologies en favorisant la créativité, le décloisonnement, les partenariats, d'encourager les interactions entre disciplines...

Dans cette optique de vouloir encourager les partenariats, ce programme Mémo-Mines se réalisera en collaboration avec les laboratoires De Visu (Université de Valenciennes), GERiiCO (Université de Lille 3) et l'INALCO (Institut National des Langues et Civilisations Orientales). La coordination du projet est confiée à un membre du laboratoire De Visu (de l'Université de Valenciennes et du Hainaut Cambrésis). L'activité de recherche du laboratoire De Visu est à la fois axée sur l'information et de la communication mais aussi sur des disciplines très techniques telles que le génie mécanique, ou encore le génie informatique. Cette interdisciplinarité, qui permet au laboratoire de faire se croiser des méthodes et des méthodologies de disciplines très différentes, est un véritable enrichissement pour la recherche qui y est menée.

L'INALCO est l'acronyme de l'Institut National de Langues et Civilisations Orientales, c'est un établissement de recherche et d'enseignement de plus d'une centaine de langues. Toutefois au sein de l'INALCO, le projet est pris en charge par l'équipe de recherche PLIDAM (Pluralité des Langues et des Identités en Didactique : Acquisitions et Médiations) qui est une équipe pluridisciplinaire qui s'inscrit dans le champ de la linguistique, de la pédagogie de l'éducation, des sciences sociales, de la géopolitique, des sciences de l'information et de la communication et des technologies.

1 - 3 - 3 Le stage et ses missions

D'un point de vue linguistique, une langue devient langue morte dès qu'elle cesse d'être parlée. Ce qui sera le cas rapidement après la disparition des derniers mineurs, car puisque les mines ne fonctionnent plus, leur langue a arrêté de se transmettre, d'évoluer,

⁹ <http://www.agence-nationale-recherche.fr/missions-et-organisation/missions/> [consulté en ligne le 03 juin 2017]

d'être utilisée dans son cadre particulier d'utilisation. Or, comme beaucoup de parlers argotiques ou liés à des métiers, la langue de la mine est surtout une langue orale ; et une langue orale qui ne se parle plus n'est pas une langue morte, elle est une langue perdue pour l'humanité.

La volonté forte de conserver des patrimoines culturels souvent proches de la disparition est donc, aujourd'hui, un véritable enjeu. C'est celui du projet Mémo-mines qui tente de relever le défi en travaillant à sauvegarder toutes traces mémorielles de ce patrimoine minier.

Dans ce contexte, la mission de stage a consisté à identifier et caractériser de manière précise les ressources sémantiques du domaine de la mine (français et parlers dialectaux tels que le chti, etc... en France). Les ressources à identifier sont les lexiques, glossaires, terminologies, thésaurus et tout outil linguistique de représentation du domaine concerné. L'objectif est de préparer la conception d'une ontologie de la mine.

Durant ce stage, les missions à effectuer se sont déroulées en plusieurs étapes. Premièrement, il a fallu s'atteler à un travail de veille informative sur le web. Cette activité de veille s'est faite tout au long de la période de stage dans la mesure où des informations nouvelles et intéressantes pouvaient apparaître et, également parce qu'il pouvait arriver que des lexiques pertinents soient passés sous silence par le web à un moment donné.

L'étape suivante a consisté à collecter, à extraire l'information, puis à traiter les données collectées pour pouvoir les instancier dans un tableur. Toutefois, au préalable, il a fallu réfléchir à la mise en place et à la robustesse d'un modèle de fiche terminologique.

Le travail de ce stage, qui se situe au tout début du projet "Mémo-Mines" est d'un grand intérêt pour les chercheurs qui vont le reprendre s'il est bien mené car la modélisation proposée va permettre une utilisation aisée ou non des ressources collectées. D'où l'importance de la structure dans le travail mené.

2 - Les outils théoriques nécessaires à la modélisation

Pour mener à bien les missions confiées, il a fallu convoquer un certain nombre des connaissances acquises durant la formation universitaire clairement associée à l'organisation des connaissances et de ses systèmes. Nous allons, dans cette partie expliciter celles dont nous avons eu besoin durant le stage.

2 - 1 - La science de l'information : de la documentarisation à la redocumentarisation

Dans un contexte d'infobésité ou de surcharge informationnelle lié à la multiplication d'informations, mais également à l'essor des nouvelles technologies, la science de l'information doit faire face, depuis quelques années, à de profondes mutations qui viennent quelque peu bousculer les lignes traditionnelles de cette discipline.

2 - 1 - 1 De la naissance de la science de l'information

Toutes les questions liées à l'information et à ses opérations comme le stockage et sa transmission sont au cœur des préoccupations qui ont traversé durant des siècles nos sociétés. C'est précisément le sujet de la conservation de l'information et de sa pérennisation qui demeure la question centrale parce qu'en dépit des moyens matériels mis en œuvre, l'information ou la connaissance ne parvient pas à être conservée de manière exhaustive. En effet, de l'usage du papyrus dans la civilisation égyptienne en passant par le parchemin dès le Moyen-âge, le souci de consigner des informations démontre parfaitement cette volonté de garder des traces d'une culture, d'une civilisation, d'un savoir et une détermination à vouloir les transmettre. D'un point de vue anthropologique, et plus particulièrement de celui de Jack Goody, nous pouvons revenir sur un article¹⁰ de Mathieu Quet où il explique et décrit sous forme de note de lecture, au travers d'un ouvrage¹¹ de l'anthropologue, les enjeux réels du passage d'une culture de l'oralité à une culture de l'écrit. La pensée de Jack Goody défend l'idée que l'écriture permet de poser le savoir, de le figer en quelque sorte, ce que la tradition orale n'a pas toujours permis.

¹⁰ Mathieu Quet, « Jack Goody, *Pouvoirs et savoirs de l'écrit* », *Lectures* [En ligne], Les comptes rendus, 2007, mis en ligne le 30 août 2007, consulté le 09 juin 2017. URL : <http://lectures.revues.org/429>

¹¹ Jack Goody, *Pouvoirs et savoirs de l'écrit*, La Dispute, 2007, 269 p.

L'avènement de l'imprimerie a également participé au développement accru de la production d'informations ce qui a favorisé le développement des sciences. En effet, on ne peut nier les effets de cette révolution technique qui a organisé la mise en place progressive d'une société de l'information. Cette prolifération d'objets contenant de l'information a vu l'émergence de nouvelles disciplines telles que "la bibliothéconomie, la muséoeconomie, la documentation et le journalisme"¹² qui portent "un intérêt particulièrement grand aux supports d'information et non à l'information elle-même"¹³. En tout état de cause, l'émergence de ces disciplines est le résultat de préoccupations bien réelles : vouloir mettre de l'ordre à des fins de conservation, de pérennisation et de partage. Dans ce contexte, se sont mises en place, dès le 19ème siècle, les classifications décimales de Dewey et celle de Paul Otlet dans laquelle ce dernier a entamé une réflexion sur la notion même du document qui impliquait de toute évidence une nouvelle approche s'éloignant des techniques de la bibliothéconomie traditionnelle pour présenter, dans son ouvrage majeur¹⁴, une vision "moderne" de la documentation. Paul Otlet avait parfaitement pris conscience de l'importance de cette prolifération de documents qu'il devenait déjà nécessaire d'organiser et de classer dans un souci de vouloir partager la connaissance. Il explique, par ailleurs, dans son traité que "pour rendre accessible la quantité d'informations et d'articles donnés chaque jour dans la presse quotidienne, dans les revues, pour conserver les brochures, comptes-rendus, prospectus, les documents officiels, pour retrouver les matières éparées dans les livres, pour faire un tout homogène de ces masses incohérentes, il faut des procédés nouveaux, très distincts de ceux de l'ancienne bibliothéconomie tels qu'ils sont appliqués"¹⁵. Toutefois le concept de documentation ne s'est imposé réellement qu'après la seconde guerre mondiale qui est, par ailleurs, un tournant majeur pour cette science.

¹² Le Coadic Y-F., La Science de l'information, Paris, PUF, 1994. (Que sais-je ?)

¹³ *ibid.*

¹⁴ Otlet P., Traité de documentation : le livre sur le livre : Théorie et pratique, Brussels, éditions Mundaneum, 1934 (réédité sur l'initiative d'A. Canonne en 1989 par le Centre de lecture publique de la Communauté française de Belgique, disponible en ligne.

¹⁵ *ibid.*

2 - 1 - 2 De l'information aux technologies de l'information

On ne peut parler de science de l'information sans l'évoquer selon une approche épistémologique et sans s'intéresser à l'objet scientifique qu'est l'information en lui-même. Hubert Fondin définit¹⁶ la discipline comme “ la science qui étudie la façon dont on crée, diffuse, stocke et retrouve une information (connaissance communiquée) quelle qu'elle soit”. Il reprend également dans cet article¹⁷ une vision développée, quant à elle, par Yves-François Le Coadic, pour qui la science a pour but “d'étudier la nature et la genèse de l'information ; (d') analyser les processus de production, de communication et d'usage de ces informations ; et (de) concevoir les systèmes qui permettent leur diffusion et leur usage”. La discipline est, au fond, très associée au monde anglo-saxon mais elle se positionne dans un contexte purement technique voire scientifique où elle est plus communément appelée *Information Science*. En revanche, en France, la science de l'information s'inscrit dans le champ plus large des Sciences de l'Information et la Communication (SIC), un schéma vivement défendu par les pères fondateurs de la discipline en France.

Comme mentionné plus haut, la spécificité “française” de cette discipline est justement caractérisée par une interdisciplinarité qui “se traduit par une collaboration entre des disciplines conduisant à des interactions, c'est à dire à une certaine réciprocité dans les échanges telle qu'il y ait au total enrichissement mutuel. (...)”¹⁸

¹⁶ Fondin H., La science de l'information : posture épistémologique et spécificité disciplinaire, Documentaliste-Sciences de l'Information, 2001/2 (Vol. 38), [consulté en ligne le 09 juin 2017]

¹⁷ *ibid.*

¹⁸ Le Coadic Y-F., La Science de l'information, Paris, PUF, 1994. , p.23 (Que sais-je ?)

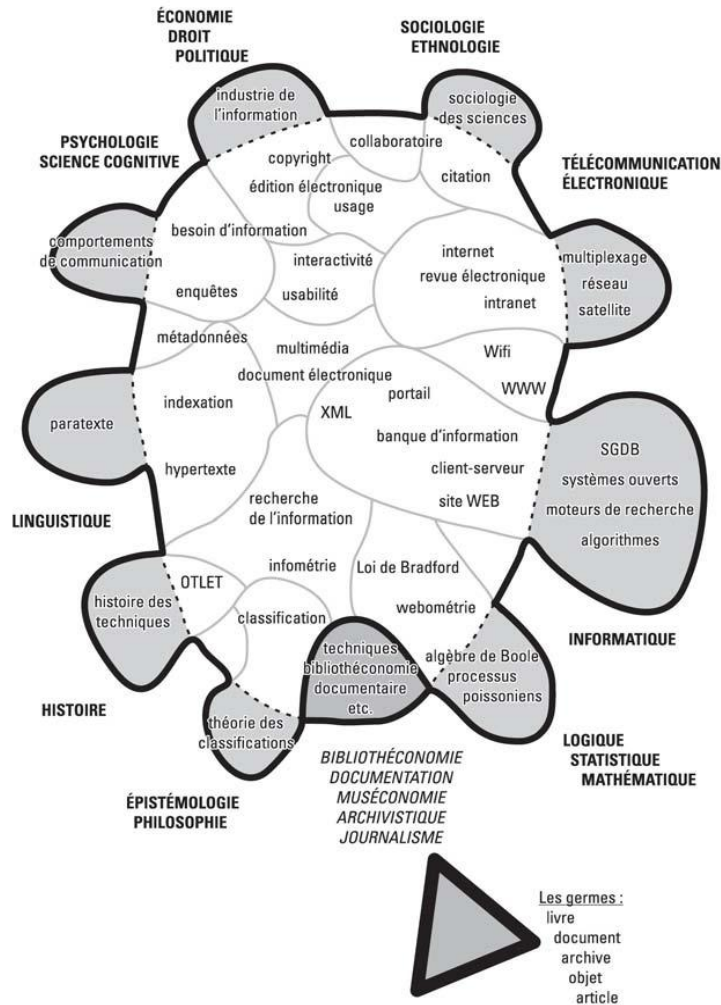


Fig.1 - La carte de la science de l'information

Cette carte, extraite de l'ouvrage d'Y-F Le Coadic,¹⁹ met en valeur distinctement la diversité des disciplines et des différentes thématiques afférentes qui apparaissent à la périphérie.

La science de l'information ou *Information Science* fortement dominée par les techniques a connu un véritable tournant dès les années 50 avec l'arrivée et l'usage de l'informatique qui a profondément bouleversé la discipline et amené de nouvelles pratiques professionnelles dans la manière de gérer le document contenant l'information. Mais c'est surtout après 1981, comme le précise Le Coadic, que la révolution s'est produite avec l'apparition du micro-ordinateur "personnel" et l'arrivée du premier Macintosh en 1984, puis avec l'apparition au début des années 90 d'internet qui a permis l'interconnexion des ordinateurs entre eux. Le développement des réseaux causé par cette révolution du numérique a donné un autre statut au document ; ce que

¹⁹ *ibid.*

Jean-Michel Salaün appelle la “redocumentarisation”. Il “s’agit (alors) de traiter à nouveau des documents traditionnels qui ont été transposés sur un support numérique en utilisant les fonctionnalités de ce dernier.”²⁰

2 - 1 - 3 L’information au cœur du stage

Dans le cadre du projet Mémo-mines, il est question d’organiser la conversion de traces mémorielles autour du patrimoine minier dans un souci de valorisation d’un héritage qui a profondément transformé les paysages à la fois culturels et industriels de l’Europe.

Ce projet s’inscrit complètement dans le processus de “redocumentarisation” qui consiste à prendre en charge des documents traditionnels en les transposant “sur un support numérique en utilisant les fonctionnalités de ce dernier”.²¹ Il intègre, d’une manière générale, que la révolution et l’ampleur du numérique sont la marque d’une nouvelle étape, voire une nouvelle vie pour le document. En effet, comme l’exprime J.P Salaün, “il s’agit alors d’apporter toutes les métadonnées indispensables à la reconstruction à la volée de documents et toute la traçabilité de son cycle.” C’est l’émergence du document numérique que l’on peut définir comme un “ensemble composé d’un contenu, d’une structure logique, d’attributs de présentation permettant sa représentation, exploitable par une machine afin de restituer une version intelligible pour l’homme. Le document numérique peut être créé à l’état natif ou obtenu par un processus de transformation d’un document physique, on parle dans ce cas de document numérisé.”²² Sa spécificité est qu’il est décrit par un ensemble de métadonnées correspondant à un ensemble structuré d’informations renseignant de manière précise une ressource documentaire. Les métadonnées à la fois descriptives, structurelles et administratives permettent, quand elles sont créées judicieusement, une indexation pertinente pour faciliter l’identification du document.

Actuellement, le document numérique, de par sa généralisation, fait l’objet de nombreux travaux de recherche parce qu’il est devenu une vraie problématique de l’Information Scientifique et Technologique des professionnels de la science de l’information et de la documentation. Dans ce cadre, il est important de citer pour exemple les travaux du

²⁰ Salaün, Jean-Michel « La redocumentarisation, un défi pour les sciences de l’information », *Études de communication* [En ligne], 30 | 2007, mis en ligne le 01 octobre 2009, consulté le 09 juin 2017. URL : <http://edc.revues.org/428> ; DOI : 10.4000/edc.428

²¹ *ibid.*

²² <http://www.proarchives-systemes.fr/infos-pratiques/lexique/document-numerique/> [consulté en ligne le 10 juin 2017]

réseau Roger T. Pédaque²³ qui définit le document numérique selon trois dimensions : la forme ou le signe, le contenu ou le texte, le médium ou la relation. C'est un travail de réflexion sur la question de l'organisation documentaire.

Il s'agit, en quelque sorte, de procéder à un travail d'organisation "d'objets-documentaires"²⁴ comme le définit J.P. Salaün ce qui permet de faciliter la recherche d'information en réduisant les bruits et silences dans un contexte de surinformation.

2 - 2 - L'Organisation des Connaissances, une discipline de la Science de l'Information

C'est donc parce que l'Organisation des Connaissances est indispensable à la Science de l'Information qu'il faut ici expliciter ce qu'elle est. Appelée en anglais le "*Knowledge Organisation*", en Sciences de l'Information et de la Communication, elle correspond à l'organisation de l'information disponible dans des dispositifs informationnels.

2 - 2 - 1 De l'organisation des connaissances

On l'a vu, pérenniser les connaissances par des écrits est une problématique ancestrale qui a traversé toutes les périodes de la réflexion humaine. Dans le même temps, l'être humain a cherché à organiser et à structurer ces connaissances. En effet, l'individu, depuis toujours, cherche, réfléchit en permanence à organiser tous les éléments constitutifs du contexte dans lequel il évolue. Il est vrai que la préoccupation de vouloir mettre de l'ordre dans un souci de conservation, de protection des éléments de connaissances est un enjeu majeur : classer permet surtout de retrouver et donc de pouvoir utiliser les connaissances à des fins de consultation ou de recherche. Joseph Tennis²⁵ décrit l'organisation des connaissances comme une discipline qui porte sur la conception, l'étude et la critique des méthodes d'organisation et de représentation des documents qu'une société juge méritant d'être conservés (Tennis, 2008)²⁶. Aussi, on ne peut traiter de l'organisation des connaissances sans évoquer son objet même qu'est la connaissance qui mérite quelques clarifications d'ordre notionnels liés à l'univers de

²³ nom collectif d'un réseau de scientifiques francophones travaillant dans les divers domaines d'expertise des sciences humaines et sociales ainsi que des sciences et techniques de l'information et de la communication.

²⁴ *ibid.*

²⁵ Enseignant spécialisé sur les théories des classifications, Information School, Université de Washington

²⁶ Joseph T. Tennis, « Le poids du langage et de l'action dans l'organisation des connaissances : Position épistémologique, action méthodologique et perspective théorique », *Études de communication* [En ligne], 39 | 2012, mis en ligne le 01 décembre 2014, consulté le 07 juin 2017. URL : <http://edc.revues.org/4010> ; DOI : 10.4000/edc.4010

l'organisation des connaissances. Ainsi, l'idée de la connaissance est définie comme correspondant "à l'appropriation et à l'interprétation des informations dans un contexte donné"²⁷, elle "désigne alors un rapport de la pensée à une réalité extérieure. Elle est ici une adéquation de l'esprit à l'objet, un mode de mise en relation de phénomènes distincts dont on suppose qu'ils ont des propriétés communes qui donnent un sens à une situation observée"²⁸. Elle peut être également perçue comme une entité où les contours ne sont pas clairement perceptibles d'où la nécessité de vouloir l'organiser pour la rendre plus compréhensible. Pour cela, l'ensemble des connaissances nécessitent d'être classifiées.

2 - 2 - 2 Les Systèmes d'Organisation des Connaissances (SOC)

La question des SOC est au cœur de la problématique de notre mémoire qui invite justement à réfléchir sur la pertinence d'un système qui rendrait compte des connaissances sur la mine pour pouvoir les partager et valoriser un domaine. C'est le principe même de la mission de ce stage qui a pour objectif d'identifier, collecter des ressources sémantiques du domaine de la mine et de les organiser.

Les SOC, dans l'essence même de ces systèmes, se mettent donc en place pour répondre précisément à des besoins classificatoires. Car, s'il n'existe pas de structure d'organisation naturelle, il n'y a pas non plus de structure idéale. En effet, à chaque problématique documentaire peut correspondre un système. Les origines des SOC sont millénaires car, déjà sous Aristote, des méthodes classificatoires étaient pratiquées. Toutefois, l'avènement de l'imprimerie a produit un réel impact sur les systèmes classificatoires en raison de la multiplication des documents. Manuel Zacklad reprend dans un article²⁹ une définition de Gail Hodge pour préciser que les SOC "englobe(nt) tous les types de schèmes permettant d'organiser des informations et de promouvoir la gestion des connaissances". M. Zacklad identifie aussi six familles de Systèmes d'Organisation des Connaissances qui sont les suivantes³⁰ :

²⁷ https://fr.wikipedia.org/wiki/Connaissance#En_gestion [cité en cours d'Organisation des Connaissances de Mme W. Mustafa El Wadi]

²⁸ Serge Agostinelli, « Connaissance : pseudo-concept partiellement opératoire », *Études de communication* [En ligne], 39 | 2012, mis en ligne le 16 décembre 2014, consulté le 07 juin 2017. URL : <http://edc.revues.org/3883> ; DOI : 10.4000/edc.3883

²⁹ Manuel Zacklad, Évaluation des systèmes d'organisation des connaissances, Les Cahiers du numérique 2010/3 (Vol. 6), [consulté en ligne le 07 juin 2017], URL : <https://www.cairn.info/revue-les-cahiers-du-numerique-2010-3-page-133.htm>

³⁰ cité en cours d'Organisation des Connaissances de Mme W. Mustafa El Wadi le 11/10/2016

- les classifications épistémiques universelles de la bibliothéconomie pour ne pas mentionner les Classification Décimale de Dewey (CDD), Classification Décimale Universelle (CDU) de Paul Otlet et les approches à facettes universelles, et la Colon Classification (CC) de Ranganathan.
- les langages documentaires et le thésaurus
- les ontologies “formelles” et le web sémantique
- les approches multidimensionnelles : les ontologies “sémiotique” et les approches à facettes
- les annuaires de ressources internet et collaboratifs et les folksonomies
- les index automatiques des moteurs de recherche

Le tableau³¹ ci-dessous reprend une typologie des SOC proposée par Hodge³² (2000) qui expose parfaitement de manière schématique les différents systèmes.

Listes de termes	Classifications et catégories	Listes de relations
Fichiers d'autorité	Vedettes-matière	Thésaurus
Glossaires	Schémes de classification*	Réseaux sémantiques
Dictionnaires	Taxinomies*	Ontologies
Répertoires géographiques	Schémes de catégorisation*	

(* Ces trois termes/expressions sont souvent utilisés l'un pour l'autre)

Il faut ajouter que les thésaurus, qui sont des vocabulaires contrôlés servant à indexer ou à décrire des ressources documentaires, restent des langages documentaires utilisés de façon universelle malgré leur caractère très spécialisé. L'ontologie, autre langage documentaire, apparaît au début des années 90 presque simultanément à Internet. Censée représenter des connaissances, l'ontologie est, selon Bachimont³³ “utilisée, depuis plusieurs années, dans l'Ingénierie des Connaissances (IC) et l'Intelligence Artificielle (IA) pour structurer les concepts d'un domaine. Les concepts sont rassemblés et ces derniers sont considérés comme des briques élémentaires permettant d'exprimer les connaissances du domaine qu'ils recouvrent”.

³¹ *ibid.*

³² Gail Hodge, *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*, The Digital Library Federation Council on Library and Information Resources 1755 Massachusetts Avenue, NW, Suite 500 Washington, DC 20036, [consulté en ligne le 07juin 2017]

³³ Bruno Bachimont, directeur scientifique à l'INA et enseignant-chercheur contractuel (HdR) à l'Université de Compiègne.

Les ontologies sont utiles pour partager des connaissances, créer un consensus, construire des systèmes à base de connaissances”. D’ailleurs, rappelons que la finalité du projet Mémo-mines prévoit, à terme, la structuration d’une ontologie de la mine dont le but est de pouvoir mettre à disposition une terminologie de la mine en ligne.

2 - 2 - 3 Outils d’organisation et modélisation de l’information utilisés lors de ce stage

L’organisation des connaissances est une discipline qui recouvre, donc, une multitude de champs scientifiques et une diversité des Systèmes d’Organisation de Connaissances qui peuvent varier à différents niveaux comme la taille du système, par exemple. Dans le cadre du stage, et tout particulièrement dans le travail de collecte de ressources sémantiques, l’usage de vocabulaires contrôlés ou pas ont été sollicités. Ces outils dont leur utilisation permet une représentation structurée de connaissances ont été l’occasion de repérer des ressources terminologiques exploitables.

On peut citer l’usage de glossaire comme un système simplifié des SOC qui est défini, selon le CNRTL, comme une “liste de mots d’une langue, d’une œuvre, accompagnée de définitions, d’explications, de références”. Le principe même du glossaire est de traiter d’un domaine spécialisé.

On peut également mentionner les vocabulaires contrôlés et structurés, de type thésaurus pour le traitement de l’information en lien avec le domaine industriel en général et minier en particulier.

Pour résumer, il a semblé utile durant ce stage de convoquer, parmi les SOC, les formes les plus simples comme des listes de termes même si, toutefois, l’usage de dictionnaires a révélé un niveau de complexité plus élevé dans la mesure où ils contiennent plus d’informations comme des définitions, des contextes, des notes... Or, là où le niveau de complexité s’intensifie, ce sont sur les ontologies et les terminologies.

2 - 3 - De la linguistique en général et de la terminologie en particulier

L'objet de cette partie est de décrire les principes de la terminologie comme une discipline traitant de concepts langagiers en tentant de comprendre les liens qu'elle peut entretenir avec d'autres disciplines. Toutefois, elle reste intrinsèquement rattachée au domaine de la linguistique même si la question de la parenté peut parfois faire débat. Cela étant, elle est une science qui relève de la linguistique appliquée traitant d'unités linguistiques.

2 - 3 - 1 La place et le rôle de la terminologie dans les Sciences de l'information et de la communication

Avant d'évoquer la question du statut de cette discipline, il serait fructueux de resituer, d'un point de vue historique, les évolutions marquantes de la terminologie. Beaucoup de spécialistes s'accordent à dire que cette science dont "l'objet est l'étude et la compilation des termes spécialisés est ancienne" puisque déjà sous Platon, le besoin de vouloir nommer les choses se ressentait fortement. Toutefois, elle commence à être reconnue qu'après la Renaissance.³⁴ Ainsi, dans l'ouvrage de Maria-Teresa Cabré, on retrouve une réflexion d'Alain Rey qui exprime totalement l'évolution de cette discipline.

“Spontanée et théorique dès sa naissance qui correspond à celle des sciences et des techniques, puis ressentie comme nécessaire à partir du XVIIIème siècle, avec les difficultés aux développements simultanés du savoir, des techniques et de la communication sociale, la terminologie ne devient un projet scientifique qu'au XXème siècle et une activité reconnue que tout récemment.”³⁵

La terminologie, avant qu'elle ne prenne définitivement cette dénomination, était au XVI ème appelée “nomenclature” au sens de “glossaire, liste de noms”³⁶, puis très rapidement “langue des arts”. Il faut remarquer, toutefois, qu'au commencement, la

³⁴ Cabré M-T, La terminologie, Terminologie : théorie, méthode et applications, Les presses de l'Université d'Ottawa, Armand Colin

³⁵ Rey A., Les fonctions de la terminologie : du social au théorique, Actes du sixième colloque OLF-STQ de terminologie. L'ère nouvelle de la terminologie, Québec, gouvernement du Québec, pp.87-108

³⁶ Rey A., La terminologie, noms et notions, col. Que sais-je ?, PUF, Paris, 1979

terminologie était largement utilisée et pratiquée par les scientifiques pour devenir, à partir du XX^{ème} siècle, une science orientée vers la technique. Elle est définie, selon le Larousse, comme une “ discipline qui a pour objet l’étude théorique des dénominations des objets ou des concepts utilisés par tel ou tel domaine du savoir, le fonctionnement dans la langue des unités terminologiques, ainsi que les problèmes de traduction, de classement et de documentation qui se posent à leur sujet.”³⁷ Cependant, la terminologie connaît un véritable essor à partir du XX^{ème} siècle compte tenu des évolutions techniques qui se produisent dans la société. Cette discipline est consacrée grâce notamment à l’école de Vienne incarnée par la personne d’Eugen Wüster, ingénieur autrichien. Il est considéré comme le père fondateur de la terminologie pour avoir mis en place le premier une théorie de la terminologie. Pour lui, la terminologie est au carrefour de plusieurs disciplines, fortement caractérisée par une multidisciplinarité. Ses principaux travaux ont consisté à désambiguïser des problèmes de langues par le traitement de termes ce qui a nécessité de faire appel aux outils de la documentation.

Comme le fait remarquer Guy Rondeau, professeur à l’Université de Laval, “terminologie et documentation sont indissociables l’un de l’autre, en ce sens qu’il est impossible de faire de la terminologie sans avoir accès, directement ou indirectement, à une abondante documentation spécialisée.”³⁸ Il est vrai que l’objet même de la terminologie est le terme, terme que l’on peut retrouver dans un corpus aussi varié que des lexiques, des dictionnaires, ou encore des glossaires. Toutefois, c’est dans les systèmes classificatoires que la terminologie prend toute sa place, et pour cela Maryvonne Holzem, maître de conférences à l’université de Rouen, parle dans son ouvrage³⁹ de “classification terminologique”. Il est vrai que le langage classificatoire tel que la Colon Classification (CC) de ce mathématicien indien, Ranganathan, procède à un découpage du langage en domaine et sous-domaines, ou encore à un classement des termes selon les relations qu’ils entretiennent entre eux.

Bien entendu, une terminologie ne peut fonctionner de manière ambiguë, et pour cela, elle se doit de veiller à travailler selon une forme, un modèle qui puisse répondre à un certain nombre de critères répondant aux rigueurs imposées par la discipline elle-même.

³⁷ <http://www.larousse.fr/dictionnaires/francais/terminologie/77407> - [consulté en ligne le 10 juin 2017]

³⁸ Rondeau G., Terminologie et documentation, La documentation, Volume 25, numéro 1, mars 1980 - [consulté en ligne le 10 juin 2017]

³⁹ Holzem M., Terminologie et documentation, pour une meilleure circulation des savoirs, ADBS Editions, 1999

2 - 3 - 2 La modélisation comme outil de la terminologie

Le principe même de toute activité terminologique, qui consiste à être en mesure de repérer les termes désignant un concept relatif à un domaine, a pour objectif premier de permettre et de faciliter une communication en la rendant dénuée de toute ambiguïté. La terminologie demeure une discipline qui demande beaucoup de rigueur en développant des outils efficaces.

Pour pouvoir assurer une bonne compréhension de données terminologiques, il faut être en mesure d'organiser leur gestion et leur structuration. Le travail du terminologue, après avoir collecté l'ensemble de connaissances nécessaires, réside dans la manière dont les données vont être gérées pour rendre homogène un système et surtout le rendre compréhensible pour l'utilisateur. Toutefois, la structuration doit pouvoir répondre à un degré de précision et essentiellement à un besoin de normes qui "s'applique en général à tout usage linguistique, à toute formation théorique, à toute pratique complexe, notamment en technologie et à toute transmission du savoir."⁴⁰ Vouloir mettre de l'ordre dans les idées, dans des techniques passe incontestablement par un processus de normalisation indispensable dès lors que les domaines se spécialisent. Cette phase de normalisation permet surtout d'éviter la multiplication des notions et de l'ambiguïté qu'elle peut provoquer. Aussi, des organismes de normalisation se mettent progressivement en place comme l'Organisation Internationale de Normalisation (ISO) créée en 1947 et se compose de représentants d'organisations nationales de normalisation tels l'AFNOR (Association Française de NORmalisation) et qui a pour but de produire des normes. Le terme de norme se définit comme suit :

Une Norme internationale fournit des règles, des lignes directrices ou des caractéristiques relatives à des activités ou à leurs résultats, dans le but de réaliser le degré optimal d'ordre dans un contexte donné. Elle peut prendre de nombreuses formes : normes de produits, mais aussi méthodes d'essai, codes de bonne pratique, lignes directrices et normes de systèmes de management⁴¹.

La question d'organisation et de structuration des données terminologiques invite à réfléchir à la mise en place d'un modèle qui demeure l'instrument essentiel pour représenter des connaissances et les accueillir dans un système ; un modèle que Jean-

⁴⁰Rey A., La terminologie, noms et notions, col. Que sais-je ?, PUF, Paris, 1979

⁴¹<https://www.iso.org/fr/deliverables-all.html>, [consulté en ligne le 10 juin 2017]

Louis Le Moigne qualifie précisément de théorie.⁴² C'est pourquoi, l'entreprise de modélisation doit réussir à être le reflet de l'organisation conceptuelle du domaine traité et c'est celui-ci qui va nous permettre de fonctionner.

2 - 3 - 3 La terminologie en pratique

Une des missions du stage a consisté à la collecte de lexiques afin de mettre en place une terminologie structurée du domaine de la mine ; l'objectif final est d'organiser une ontologie de la mine dont le but est de contribuer à la valorisation d'un patrimoine proche de la disparition.

Depuis la mise en place d'une terminologie moderne vers les années 30, Wüster a posé une problématique centrée sur la systématisation des méthodes de travail pour rendre plus efficace le traitement des données terminologiques. Mais depuis le début des années 60, la discipline connaît un véritable essor principalement provoqué par le développement de l'informatique et des techniques documentaires. En effet, l'immersion dans l'outil informatique va largement contribuer à modifier la manière dont on traite les données, et également les méthodes de travail. Aussi, l'ensemble des tâches réalisé au cours du stage a fait appel à l'outil informatique notamment dans la collecte de ressources sémantiques par un travail de veille informationnelle, puis de traitement de données.

L'étape suivante concernait le travail de modélisation d'une terminologie qui doit servir à la science, à la technique et doit faciliter la communication par l'échange d'informations. La démarche terminologique doit définir un modèle qui lui servira à représenter les données dont il faut rendre compte. Marie-Claude L'Homme, spécialiste en terminologie, décrit dans son ouvrage⁴³ trois types de modèles pour réunir les données terminologiques :

- le modèle plat qui recense les informations dans une structure unique
- le modèle relationnel qui est une forme adaptée pour accueillir des données de nature différente
- le modèle hiérarchique conçu pour représenter des relations sémantiques.

Tenir compte de cette typologie proposée nous a permis de prendre un chemin réflexif autour de la structure à donner ce qui nous a amené progressivement à l'élaboration d'une

⁴² Le Moigne, J.L, Qu'est-ce qu'un modèle ?<http://archive.mcxapc.org/docs/ateliers/lemoign2.pdf>, [consulté en ligne le 10 juin 2017]

⁴³ L'Homme M-C, La terminologie : principes et techniques, Les presses de l'Université de Montréal, 2004

fiche terminologique chargée d'ordonner les différentes rubriques d'un terme pour y recenser chacune des données. Le but est de pouvoir créer un outil performant, pertinent et surtout utilisable.

3 - La modélisation d'un SOC au cœur de la démarche méthodologique

Cette partie présente la méthodologie mise à l'œuvre durant la période de ce stage au sein du laboratoire GERiiCO, et elle a pour but également de montrer l'extrême importance de la modélisation d'un Système d'Organisation de Connaissances dans la démarche méthodologique.

3 - 1 - Collecter des ressources sémantiques

On l'a dit, le projet de ce stage qui consiste à identifier des ressources sémantiques du domaine de la mine s'inscrit dans un projet plus global de valorisation du patrimoine minier (cf. annexe 1). Ce programme d'envergure prévoit d'organiser la conversion de traces mémorielles en les rendant disponibles sous forme numérique. Dans cette perspective, il a fallu mettre en place des procédures pour faciliter la recherche d'information.

3 - 1 - 1 Présentation de cette mission de stage

Une des missions de ce stage s'est précisément portée sur l'identification de ressources sémantiques traitant de l'univers de la mine, et tout particulièrement des mines de charbon. Comme le projet n'est pas circonscrit à la région des Hauts de France uniquement, mais qu'il s'inscrit dans un cadre géographique beaucoup plus large, c'est au niveau national, voire européen que le projet doit aussi se développer. En effet, Mémo-mines porte une attention particulière aux voisins européens dont les paysages sont aussi fortement marqués par l'industrie charbonnière. C'est la raison pour laquelle, le travail de collecte s'est également intéressé à des ressources étrangères, et notamment d'origine belge puisque la communauté de langue avec nos voisins wallons est double : notre français et le leur sont très proches linguistiquement, comme le sont le patois du nord et le patois wallon. Dans cette perspective, il y a fort à parier que la langue de la mine ait, des deux côtés de la frontière, des similitudes. Rappelons que la finalité du projet mémo-mines, dans ce cadre linguistique, est de concevoir et d'élaborer un métalangage qui correspond à un "modèle

formalisé d'un langage destiné à en rendre compte.”⁴⁴ En d'autres termes, les travaux, qui n'en sont encore qu'à leur prémices, doivent aboutir à la mise en place d'une ontologie de la mine et, enfin, pour assurer la pérennité d'un langage spécialisé en voie de disparition.

La première étape a consisté à collecter et à établir un inventaire de tous les éléments de types glossaires, lexiques, thésaurus... Autrement dit, il a fallu repérer des éléments terminologiques liés à l'environnement minier en repérant formellement la source de la ressource, mais également en apportant des informations sur le format, le volume et la langue d'usage du lexique. Bien entendu, cette phase préliminaire suppose de réfléchir à la mise en place d'une méthodologie de recherche d'information efficace.

3 - 1 - 2 Méthodologie de la recherche d'information

Pour mener à bien ce projet, il a fallu, dans un premier temps, mettre en place une méthode pour identifier au mieux les besoins. Durant les premières journées de stage, un travail de repérage et de collecte de ressources sémantiques a été mis en place. Il s'est agi de recueillir, dans un premier temps, des éléments de connaissances dans le domaine de l'industrie minière, et plus particulièrement celle du charbon à partir d'un environnement précis : le web car la recherche sur internet offre un accès à une quantité importante d'informations. Nous avons donc organisé une stratégie de recherche sur le web en faisant le choix de Google comme moteur de recherche, car il est reconnu comme le moteur de recherche grand public le plus performant. Les premières requêtes ont été réalisées à partir de cet environnement en saisissant des mots-clés dans le moteur de recherche comme “lexique mine” ou “vocabulaire de la mine”, parfois pour affiner la recherche, des opérateurs booléens tels que “and” ou “or” venaient compléter la requête. Lors de la collecte de lexiques, l'évaluation de l'information reste une étape cruciale parce qu'elle permet d'obtenir des éléments de connaissances essentiels à la constitution d'un fichier comme l'identification de la ressource, l'auteur de la source, la date de création et éventuellement la date de mise à jour. Lorsque l'information n'était pas clairement visible, il a fallu se rendre au code source de la page web pour y consulter les métadonnées du document.

Tout le travail de recherche d'information a consisté à mettre en place des opérations de veille, sur la thématique de recherche. La veille informationnelle consiste à “surveiller” de manière active tous les éléments qui peuvent intéresser notre sujet d'application. Ces

⁴⁴ <http://www.cnrtl.fr/definition/m%C3%A9talangage>, [consulté en ligne le 11 juin 2017]

opérations de veille se sont déroulées quotidiennement de manière dynamique les deux premières semaines, puis progressivement, y ont été consacré une à deux heures par jour. Cependant faute de maîtriser des outils de veille précis, pour pouvoir collecter de manière efficiente les données recherchées, la méthode “pull” a été adoptée. C’est donc plutôt une recherche d’information des plus classiques sur internet. Cette pratique revêt quelques avantages comme la pertinence de l’objet recueilli. Toutefois, elle nécessite d’y passer davantage de temps et de voir revenir très souvent des informations déjà collectées. Nous avons aussi constaté qu’un certain nombre d’informations, lors de l’interrogation du moteur, pouvaient passer sous silence, ce qui est probablement dû à un problème de formulation de la requête mais peut être aussi à l’importance du bruit rencontré sur les réseaux. Par exemple, le lexique « les mots de la mine » de Béatrice Turpin n’a pu être repéré qu’au bout de plusieurs semaines.

3 - 1 - 3 Résultat d’une collecte non exhaustive

Lors de cette étape de collecte, nous avons identifié différents types de ressources qui appartiennent à des catégories différentes. D’abord, les premiers éléments correspondent à des lexiques qui ont été constitués, élaborés par des personnes totalement amateurs mais qui sont associées, de près ou de loin, à l’univers de la mine. Ils sont passionnés, quelquefois enfants d’anciens mineurs, et ils ressentent un besoin évident de transmettre des connaissances liées à ce patrimoine. Nous avons pu en collecter plus d’une vingtaine avec une volumétrie très fluctuante selon les lexiques puisque les listes de mots vont de 3 à 868 termes ce qui correspond une liste de plus de 2500 termes. Or, à l’issue du stage, 22 lexiques ont pu être traités.

Ensuite, l’autre catégorie relève d’une série de vocabulaires contrôlés, comme les thésaurus, trouvés sur les sites suivants :

- **UNESDOC**, le thésaurus de l’UNESCO qui dresse une liste de termes contrôlés et structurés pour l’analyse thématique et la recherche de documents et publications dans les domaines d’action de l’organisation. Un lexique de la mine était recensé dans la base de données documentaire mais il n’est disponible qu’en version de papier
- **Le Thésaurus du BIT** (Organisation Internationale du Travail)
- **RAMEAU** (Répertoire d’Autorité-Matière Encyclopédique et Alphabétique Unifié) est langage d’indexation utilisé, en France, par la Bibliothèque Nationale de France, les bibliothèques universitaires, de nombreuses bibliothèques de lecture publique

ou de recherche ainsi que plusieurs organismes privés où ont été collectées des notices

- **EUROVOC** est le thésaurus multilingue et multidisciplinaire couvrant la terminologie des domaines d'activité de l'Union européenne
- **JOCONDE** est un catalogue qui regroupe plus de 550000 notices de toute nature
- **THÉSAURUS de statistique Canada** renferme une liste de termes thématiques utilisés par l'organisme pour décrire ses ressources d'information.

Notons, toutefois, que ces thésaurus sont généralistes et ne traitent que partiellement de la mine de charbon ou alors d'un point de vue particulier comme celui de l'industrie ou de la géologie.

Une troisième catégorie de ressource a été répertoriée mais ce sont des informations non disponibles sur le web. Un extrait d'un document de taille a été repéré sur internet ; il s'agit d'un ouvrage spécialisé de la mine paru en 1926, *La Houilleries liégeoise, vocabulaire philologique et technologique de l'usage moderne dans le bassin de Seraing-Jemeppe-Flémalle*, de Jean Haust. Cet ouvrage est disponible et consultable auprès du Centre Liégeois d'Archives et de Documentation de l'Industrie Charbonnière. Le CLADIC "rassemble et met gratuitement à la disposition des historiens, chercheurs, enseignants ou amateurs éclairés plus de 100.000 documents de tous types, portant sur l'industrie minière en général et plus particulièrement sur l'exploitation du charbon dans le bassin liégeois"⁴⁵.

Il va sans dire que ce centre de documentation ne pourra que devenir un partenaire privilégié du projet compte tenu de l'importante documentation qu'il recèle. D'autres ressources bibliographiques ont été identifiées mais n'ont pu être consultés, et ne sont disponibles qu'en version papier comme :

- *La petite encyclopédie de l'énergie*, un ouvrage paru en 1988 et édité par le Ministère de l'industrie. Ce volume a été repéré sur la base documentaire de l'Unesco (Unesdoc).
- *Les mots de la mine* de Brigitte Benneteux paru en novembre 2010 édité par Un autre reg' art.

⁴⁵<http://www.blegnymine.be/fr/centre-liegeois-d-archives-et-de-documentation-de-l-industrie-charbonniere-cladic>, [consulté en ligne le 11 juin 2017]

3 - 2 - Constitution d'un corpus : description des choix

Le premier des choix fait découle directement de la façon de rechercher les informations : en effet, les requêtes ont toutes été faites en français, ce qui implique *a priori* des choix linguistiques et géographiques se centrant sur des zones minières francophones : France, Canada, Suisse, Belgique. Ce choix est cohérent dans la mesure où, même si le projet est européen, il semblait impossible de traiter tous les lexiques de toutes les langues du projet durant les six semaines de stage. Par ailleurs, se concentrer sur des lexiques de la langue maternelle du chercheur permet de pouvoir évaluer la richesse et l'intérêt des termes trouvés.

Ensuite, il a fallu réfléchir à la composition du lexique agrégé durant ce stage. En effet, à partir des références mises à jour sur Internet, la constitution d'un lexique s'est mise en place progressivement. Celui-ci sera bien évidemment amené à évoluer lorsque d'autres lexiques viendront à être identifiés et intégrés à l'outil, mais déjà lors de cette étape déterminante, il a fallu faire des choix de conservation ou de rejet des sources.

3 - 2 - 1 Le rejet de certaines sources et ses justifications

Cette étape d'élaboration et de construction de corpus a consisté à faire des choix qu'on va justifier pour expliquer comment des lexiques sont maintenus et d'autres pas. Après un premier bilan de la collecte réalisée, un constat a été posé à savoir celui d'avoir mis à jour des glossaires autour de la mine élaborés par des non-spécialistes de la terminologie. Dans l'ensemble de ces lexiques, il apparaît que les auteurs sont des personnes passionnées par l'univers de la mine, et particulièrement sensibles aux questions de valorisation du patrimoine minier. De plus, les listes de mots accompagnés de définitions viennent enrichir, et nourrir le lexique académiquement connu. A titre d'exemple, l'on citera le lexique d'André de Marles, pseudonyme d'André Paillart, qui contient près de 300 termes spécialisés autour du domaine minier. La richesse de ces lexiques, et le fait qu'ils ne soient pas académiquement répertoriés sont particulièrement intéressants d'un point de vue de la recherche. Aussi, notre choix s'est porté sur ces lexiques en particulier, au total ils représentent plus de 1500 termes même s'ils comportent bien évidemment des doublons. A cela, s'ajoute le lexique de Béatrice Turpin, *Les mots de la mine*, qui comporte 868 termes précisément. Ce travail, disponible en version numérique et papier sous forme d'ouvrage, est un glossaire du parler des mineurs de la région Nord-Pas de Calais qui a été constitué à partir d'un corpus d'œuvres de mineurs écrivains de la région. Un lexique bilingue français

- anglais de plus de 700 termes a été collecté et propose une terminologie technique de l'industrie minière. Ce document des cahiers de l'Office de la Langue Française du Canada n'a pu être exploité dans le cadre du stage car le fichier PDF est sous format image. Il est conservé en attendant de passer un traitement de type OCR sur le document.

En revanche, les vocabulaires de type thésaurus n'ont pas été retenus pour le moment pour deux raisons. La première est que leur langage repose déjà sur une structure, il ne semblait donc pas judicieux de refaire ce qui avait déjà été fait par d'autres spécialistes. Par ailleurs, nous avons considéré que ces listes de mots ne pouvaient pas être considérées comme prioritaires de par leur caractère trop généraliste. Par exemple, nous avons rejeté le thésaurus Joconde parce qu'il n'apportait aucune information d'ordre lexical. En revanche, sa fonction indexe des pièces de musées ce qui nous a permis de réaliser l'importance d'œuvres picturales décrivant des paysages miniers ou encore des scènes de la vie quotidienne. Cette ressource devra bien évidemment être utilisée dans le projet, mais pas dans le travail concernant les ressources sémantiques.

3 - 2 - 2 Le corpus et ses traitements

Pour pouvoir exploiter les informations de manière efficace, nous avons, dans un premier temps, mis en place une fiche descriptive pour chaque ressource identifiée (cf. annexe 2). Cette fiche apporte des informations à la fois administratives et structurelles sur le lexique. Cependant, les champs prédéfinis ne sont pas nécessairement complétés dans la mesure où les auteurs ne donnent pas systématiquement d'éléments précis ou alors ne le font que très partiellement. Ce manque d'informations apparaît très souvent dans des lexiques élaborés par des non-spécialistes et atteste, d'ailleurs, d'un réel problème de structuration de données.

C'est pourquoi, la deuxième partie de la mission a consisté à structurer les éléments d'informations en procédant à des opérations de traitements.

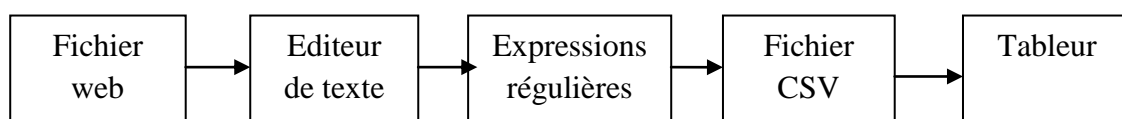


Figure 2 : schéma des différentes opérations de traitement

Le schéma ci-dessus représente les différentes étapes et processus du traitement d'un fichier texte. Il a fallu, après récupération du document depuis le web, copier le lexique vers un éditeur de texte de type Notepad++ pour redonner une structure exploitable au document. Pour bien comprendre ce dont il est question, un focus sur cette étape d'expressions régulières est nécessaire. En effet, les expressions régulières, plus communément appelées regex, sont des outils très efficaces qui permettent de représenter des modèles de chaîne de caractères. Pour les besoins du stage, nous avons fait appel à cet outil pour traiter de fichiers dont la volumétrie est importante et d'éviter des opérations manuelles chronophages (cf. annexe 3). Il va de soi que cet exercice est quelque peu fastidieux car il est relativement difficile à appréhender lorsque les connaissances ne sont pas réunies. Une fois le fichier traité en texte, il est enregistré sous un format csv (cf. annexe 4) dans lequel les valeurs sont organisées par des séparateurs ce qui va permettre de sauvegarder les données dans un format de type tableur. Chaque lexique traité est ensuite inséré dans un document sous forme de tableur où à chaque feuille correspond un lexique. L'étape suivante, qui sera probablement réalisée par les chercheurs qui prendront la suite de ce travail, sera d'envisager la fusion de tous les lexiques. Dans cette phase de traitement de fichier, nous sommes intervenus sur les questions de corrections orthographiques qui étaient nombreuses à faire, et également sur l'homogénéisation du fichier global où il arrivait parfois d'avoir des termes écrits en majuscule, et d'autres en minuscule. Une fois le corpus plus ou moins constitué, il a été instancié dans une feuille de type tableur. Rappelons, toutefois, que ce document qui reproduit l'ensemble des lexiques collectés est loin d'être exhaustif ; il est, de toute évidence, évolutif.

Cependant, l'opération de traitement la plus longue et la plus fastidieuse concerne le fichier de Béatrice Turpin contenant une volumétrie de 868 termes

3 - 2 – 3 Un cas pratique : Les mots de la mine

Ce lexique mérite une attention toute particulière pour deux raisons : la première est que son contenu est d'une grande richesse et que les informations paraissent de prime abord structurées, ordonnées ; la seconde est que son exploitation a nécessité un grand nombre d'interventions et d'opérations.

L'auteur de ce lexique est Béatrice Turpin qui est une spécialiste de la langue puisqu'elle est maître de conférences en science du langage à l'université de Cergy-Pontoise depuis

2002, après avoir été enseignante-chercheuse à l'université de Rabat au Maroc, puis à l'université de Lille 1 et à l'université d'Artois.

Pour revenir au lexique, il a été récupéré sur internet où il est proposé avec un index thématique ou alphabétique (cf. annexe 5). En cliquant sur chaque terme, une fiche apparaît et apporte des informations d'ordre terminologique. Grâce à l'aide d'un étudiant spécialisé en lexicographie, linguistique et en traitement automatique du langage, une boucle et la commande wget du terminal linux ont permis le téléchargement des 868 pages du lexiques. Cette opération a permis d'obtenir pour chaque terme un fichier en format html lui correspondant, cela revient à dire 868 fiches totalement indépendantes les unes des autres. Ensuite, Monsieur Jacquemin a été sollicité pour capter l'ensemble des informations dans un format exploitable. Après avoir généré un script sur l'ensemble des fiches, nous avons obtenu sous un format csv l'ensemble des éléments du lexique répartis sur 96 colonnes. Ce nombre de colonnes correspondent aux 96 valeurs associées de l'attribut "class" de chaque terme.

Il va de soi qu'une difficulté majeure s'est posée, celle de devoir insérer les données réparties sur 96 colonnes dans le tableau regroupant les autres lexiques. Ce constat posé, il a fallu se rendre à l'évidence : le modèle initialement conçu ne correspondait plus à la réalité et il a fallu réfléchir à la mise en place d'une modélisation robuste et stable.

3 - 3 - Vers la robustesse d'un modèle

Pour commencer, il est intéressant de s'interroger sur la notion de modèle et sur le besoin de formaliser une structure pour la rendre intelligible et surtout performante. C'est tout l'intérêt porté durant le stage d'aller vers une modélisation qui pourrait stabiliser un système d'organisation d'ordre terminologique.

3 - 3 - 1 De l'importance de modéliser

D'emblée, la question de la modélisation a clairement été posée : quel outil utiliser et comment représenter de manière opératoire les informations ? Les premières données collectées ne nous ont pas permis d'établir un modèle fiable et robuste en raison de rubriques linguistiques faiblement renseignées sur les premiers lexiques collectés. Nous avons, néanmoins, esquissé une ébauche de modèle pour représenter les renseignements concernant les termes qui se décomposent selon les champs suivants :

- le nom de la ressource correspondant à l'auteur de la ressource
- l'entrée correspondant au terme présenté dans le lexique
- la définition du terme qui permet la compréhension du concept d'un domaine
- la variante d'un terme
- le renvoi d'un terme vers un autre

Chacune de ces rubriques a été consignée dans un tableur pour y insérer les informations. On peut, dans ce cadre, considérer cette opération comme un début de processus de modélisation. En effet, dès lors que des informations sont enregistrées dans une structure, le modèle est en cours de création. Sauf que cette première structure n'était pas en mesure d'accueillir les données terminologiques de tous les lexiques récoltés, et tout particulièrement celui de Béatrice Turpin, *Les mots de la mine*. Or, un modèle doit servir d'instrument explicatif et de compréhension de l'objet qu'il décrit. C'est pourquoi, les premières maquettes manquaient de robustesse parce qu'elles ne contenaient justement pas l'ensemble des champs nécessaires au recueil des informations terminologiques. C'est pourquoi, le processus de validation s'est mis en place de manière progressive parce qu'en terminologie, les connaissances sont loin d'être figées et sont en constante évolution. En tout état de cause, cette partie de la mission a été l'occasion de réaliser toute la difficulté de modéliser quand il y a une pauvreté informationnelle, et *a contrario* d'un besoin impératif de description des données quand elles deviennent conséquentes.

Aussi, nous avons voulu, dans cette entreprise de constitution de lexique, analyser de manière quantitative l'ensemble des termes du lexique.

3 - 3 - 2 Une analyse lexicale

Pour continuer à avancer dans ce processus de modélisation, nous avons entamé, mais pas réellement abouti, une analyse quantitative de la couverture informationnelle des mots contenus dans le lexique. Il est vrai que dans ce contexte d'appropriation d'un lexique spécialisé sous la forme de tableur, il est difficile d'analyser et d'avoir une vue d'ensemble du lexique sans avoir recours à des outils permettant d'explorer du texte. Aussi, visualiser l'ensemble du lexique peut offrir une cartographie des termes qui permettrait d'envisager une meilleure gestion des connaissances.

Dans cette tâche, nous avons fait appel à un outil utilisé au cours de cette année universitaire, dans le cadre de l'enseignement en visualisation de l'information, il s'agit de R studio. Cet outil de *text mining* ou de fouille de textes est une solution de récupérations

de données qui recouvre plusieurs fonctionnalités comme le traitement de données et l'analyse statistique. C'est pourquoi, nous y avons eu recours pour obtenir une analyse lexicale, et plus spécifiquement une représentation de la fréquence des mots (cf. annexe 6). Pour ce faire, et après avoir exporté l'ensemble des entrées du lexique sous R, nous avons procédé à une série de prétraitements ce qui nous a permis d'obtenir l'histogramme ci-dessous.

Grâce à ce logiciel, R studio offre une description quantifiée des entrées du lexique dans sa globalité. Ici, l'objectif de l'exercice était de mesurer la fréquence des termes, autrement dit le nombre d'occurrences de chaque unité linguistique à travers tout le lexique. En d'autres termes, les mots qui reviennent le plus fréquemment représentent, dans ce cas de figure, les notions les plus courantes du lexique. Aussi, cette analyse révèle que le mot le plus fréquent "porion", qui désigne le "maître mineur ou encore le contremaître", apparaît 25 fois. Le terme suivant est la "taille" désignant le chantier où l'on abat le charbon avec une apparition de 22 fréquences.

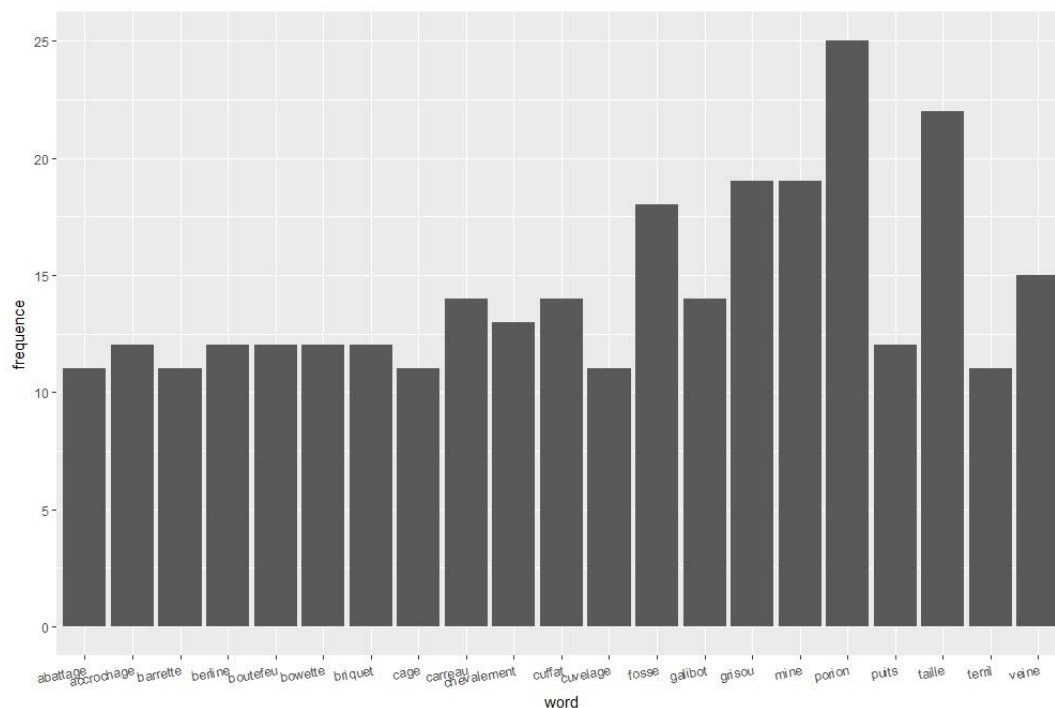


Figure 3 : histogramme des occurrences des termes du lexique

Cette analyse pourrait être un point de départ pour mettre en place un système de classification, et pour établir une hiérarchie selon les spécificités des termes que l'on pourrait associer à des sous-domaines de la mine. Cependant, cette observation n'a été que partiellement menée mais elle aurait, dans l'idéal, dû servir à contribuer plus largement à la démarche réflexive de modélisation du champ lexical.

Cette opération d'analyse de la fréquence des termes constitue une étape intermédiaire dans le processus de modélisation qui nous a amené à réfléchir à la définition du modèle.

3 - 3 - 3 De la modélisation au modèle : la fiche terminologique

La modélisation est définie, selon le CNRTL, comme une «opération par laquelle on établit le modèle d'un système complexe, afin d'étudier plus commodément et de mesurer les effets sur ce système des variations de tel ou tel de ses éléments composants.⁴⁶

C'est la raison pour laquelle, la question de la modélisation d'un système a dominé tout le stage ; c'est aussi pourquoi au cœur de cette démarche d'élaboration d'un modèle on trouve la fiche terminologique.

« La fiche terminologique est un outil de synthèse et de systématisation des données. Les principaux critères de la rédaction d'une fiche étant la validité, la concision, l'actualité et la complémentarité des données, le terminologue sélectionne à partir de son dossier terminologique la définition ou le contexte qui décrit le mieux le concept, et qui met le mieux en évidence le crochet terminologique. Il évite autant que possible les informations répétitives dans les justifications textuelles et vise à les rendre plutôt complémentaires, de manière à faciliter la construction d'une image du concept en son entier par l'utilisateur de la fiche. (Pavel et Nolet 2002 : 48) »⁴⁷

Ce modèle, par la formalisation d'un lexique, deviendrait un instrument de compréhension, d'explication et de communication. Comme déjà mentionné plus haut, les différentes opérations de modélisation menées pour la création et l'organisation de la fiche terminologique se sont mises en place progressivement, et surtout celle-ci a évolué en fonction des lexiques collectés. Par ailleurs, notre méthodologie a consisté à travailler uniquement sur le corpus en veillant à ce qu'aucune information ne soit complétée ou supprimée. Toutefois, l'intérêt et l'enjeu majeur de cet outil est de permettre une bonne compréhension d'un domaine pour contribuer aux efforts de communication et de valorisation d'un patrimoine culturel et industriel qui a profondément marqué nos sociétés. Pour résumer, la fiche doit devenir l'outil nécessaire à rendre compréhensible le concept et pour en faciliter l'usage.

⁴⁶ <http://www.cnrtl.fr/definition/mod%C3%A9lisation>, [consulté en ligne le 11 juin 2017]

⁴⁷ https://www.lli.ulaval.ca/fileadmin/llt/fichiers/recherche/revue_LL/vol35/A.Francoeur.pdf, [consulté en ligne le 11 juin 2017]

Chacune des étapes de travail nous a amené à réfléchir à la robustesse du modèle en veillant à répondre aux principes de normalisation. La partie consacrée à la recherche de normes terminologiques a été longue en raison de la difficulté à retrouver des éléments d'informations sur les normes terminologiques. Cela est, en partie, dû à une mauvaise qualité de nos recherches mais également à un manque de compréhension des principes de la terminologie. Cette étape a été l'occasion de nous documenter sur les principes généraux de la terminologie pour pouvoir mesurer et comprendre toute l'importance de cette discipline qui exige beaucoup de rigueur d'où l'importance de la normalisation.

Nous avons, finalement, pu accéder mais de manière partielle aux normes ISO comme la **Norme 1087-1 : 2000** qui offre une description des concepts d'un domaine donné. La structure est établie conformément à la **Norme ISO 10241**. Les éléments de chaque article sont affichés dans l'ordre suivant :

- le numéro d'article
- le(s) terme(s) privilégié(s)
- le(s) terme(s) toléré(s)
- le(s) terme(s) déconseillé(s)
- une définition
- un (des) exemple(s)
- une (des) note(s)

Cette partie qui a été appliquée à l'élaboration de la fiche terminologique se réfère principalement à un des plus grands spécialistes de la terminologie en France, Daniel Gouadec. Son ouvrage majeur, *Terminologie, constitution de données*, publié par l'AFNOR a été une référence de base pour le travail de conception de la fiche terminologique (cf. annexe 7). Nous avons fait le choix de classer les éléments du lexique selon trois rubriques :

- la rubrique linguistique qui concerne tous les informations linguistiques du terme comme le synonyme, la variante, le dérivé...
- la rubrique encyclopédique qui apporte des éclaircissements sur le terme comme la définition, le(s) contexte(s)...
- la rubrique administrative correspondant à des informations sur l'auteur de la fiche, ou encore de la ressource.

La fiche, dans son état actuel, devra être testée pour pouvoir évaluer de son efficacité. Le temps a manqué pour pouvoir effectuer ce test durant la période de stage, mais il est prévu

dans les semaines à venir de tester la robustesse de ce modèle de fiche en la complétant afin de voir si toutes les informations collectées dans les corpus y trouvent leur place. Si tel est le cas, l'outil créé sera utile à d'autres chercheurs qui pourront aisément réutiliser le modèle et retrouver les informations concernant les termes du lexique dans les différentes entrées de la fiche.

Conclusion

Le point de départ de notre travail avait pour objectif d'identifier et de collecter des ressources sémantiques liées au domaine de la mine. L'étape suivante a abouti à une réflexion sur les moyens et procédures à mettre en place pour l'organisation et la structuration d'un lexique spécialisé dont l'objectif final du projet serait de mettre en place une ontologie de la mine.

La problématique centrale de notre mission de stage résidait dans le fait de devoir penser une modélisation dont la démarche mènerait à un modèle efficace, robuste, stable pour pouvoir y formaliser un lexique préalablement constitué. Ce modèle doit répondre à un certain nombre de besoins pour le rendre efficient ce qu'Alain Rey ⁴⁸ décrit clairement comme des besoins de description de l'ensemble des termes, de transmission et de diffusion des connaissances d'un domaine précis, et de normalisation de tout usage linguistique. Autrement dit, le modèle doit remplir plusieurs finalités comme sauvegarder des éléments de connaissances pour ensuite, pouvoir assurer leur communication ce qui s'inscrit formellement dans le projet global de mémo-mines : mettre à disposition sous forme de corpus d'archives numériques accessibles à tous dans un souci de conservation et de valorisation du patrimoine minier.

Il a fallu travailler à la conception et à l'élaboration du modèle qui puisse accueillir les données lexicales collectées et surtout de les rendre compréhensibles. C'est la fiche terminologique qui deviendra l'outil de base pour représenter, et insérer des termes miniers. Toutefois, ce travail de conception de modèle de fiche terminologique ne s'est pas facilement constitué dans la mesure où il a fallu tenir compte des spécificités du lexique et du domaine. C'est pourquoi, les modèles de fiches standard proposés ne sont pas systématiquement applicables à un domaine précis, et par conséquent, il a fallu justifier des choix de rubriques qui permettraient de représenter au mieux l'ensemble des concepts. Pour cela, le travail de conception de fiche terminologique revêt un degré de complexité qui nécessite, de toute évidence, de faire preuve de beaucoup de rigueur. De plus, cette entreprise de modélisation qui était le coeur de notre travail a surtout engendré des questionnements sur la nécessité d'élaborer un modèle de fiche qui

⁴⁸ Rey A., La terminologie, noms et notions, col. Que sais-je ? PUF, Paris, 1979

répondent à la fois à des règles de normalisations et à des besoins spécifiques au domaine.

Ce projet d'élaboration d'un système d'organisation de connaissance qui rendrait compte de la langue de la mine et des mineurs n'est qu'au commencement du projet. Les premiers résultats de ce travail offrent les premières bases indispensables à la construction d'une terminologie de la mine : un lexique qui regroupe 2500 termes comprenant des doublons. Cette liste de termes a permis de modéliser une fiche terminologique en tenant compte des spécificités du lexique, la perspective étant de mettre en place un métalangage en créant un thésaurus puis une ontologie qui participeraient à expliciter l'univers de la mine et des mineurs.

Bibliographie

CABRE M-T, *La terminologie : théorie, méthode et applications*, Les presses de l'Université d'Ottawa, Armand Colin

DEPECKER L., *Entre signe et concept, éléments de terminologie générale*, Presses Sorbonne Nouvelle, 2002

GOODY J., *Pouvoirs et savoirs de l'écrit*, La Dispute, 2007, 269 p.

GOUADEC D., *Terminologie, constitution de données*, AFNOR, 1990

HOLZEM M., *Terminologie et documentation, pour une meilleure circulation des savoirs*, ADBS Editions, 1999

LE COADIC Y-F., *La Science de l'information*, Paris, PUF, 1994. (Que sais-je ?)

L'HOMME M-C, *La terminologie : principes et techniques*, Les presses de l'Université de Montréal, 2004

MANIEZ J., *Actualités des langages documentaires, fondements théoriques de la recherche d'information*, ADBS Editions, 2002

OTLET P., *Traité de documentation : le livre sur le livre : Théorie et pratique*, Brussels, éditions Mundaneum, 1934 (réédité sur l'initiative d'A. Canonne en 1989 par le Centre de lecture publique de la Communauté française de Belgique, disponible en ligne

REY A., *La terminologie, noms et notions*, col. Que sais-je ?, PUF, Paris, 1979

REY A., *Les fonctions de la terminologie : du social au théorique*, Actes du sixième colloque OLF-STQ de terminologie. L'ère nouvelle de la terminologie, Québec, gouvernement du Québec, pp.87-108

Webographie

AGOSTINELLINI S., « Connaissance : pseudo-concept partiellement opératoire », *Études de communication* [En ligne], 39 | 2012, mis en ligne le 16 décembre 2014, consulté le 07 juin 2017. URL : <http://edc.revues.org/3883> ; DOI : 10.4000/edc.3883

CHAUDET B., HUET R., TENIER S., MALLEDANT N., THOA R. et SAUVAIRE D., « La revue *Études de communication* », *Revue française des sciences de l'information et de la communication* [En ligne], 3 | 2013, mis en ligne le 04 novembre 2014, consulté le 03 juin 2017. URL : <http://rfsic.revues.org/655> ; DOI : 10.4000/rfsic.655

FONDIN H., La science de l'information : posture épistémologique et spécificité disciplinaire, Documentaliste-Sciences de l'Information, 2001/2 (Vol. 38), [consulté en ligne le 09 juin 2017]

FRANCOEUR A., *la fiche terminologique, entre théorie et pratique*, langues et linguistique, n°35, 2015 : p. 24-39

HODGE G., Systems of Knowledge Organization for Digital Libraries : Beyond Traditional Authority Files, The Digital Library Federation Council on Library and Information Resources 1755 Massachusetts Avenue, NW, Suite 500 Washington, DC 20036, [consulté en ligne le 07 juin 2017]

JEANNERET Y. (coord.) et OLLIVIER B. (coord.) (2004). Les Sciences de l'Information et de la Communication : savoirs et pouvoirs. Hermès, mai, n° 38, 256 p. [consulté en ligne le 02 juin 2017]

LE MOIGNE, J.L, Qu'est-ce qu'un modèle ? <http://archive.mcxapc.org/docs/ateliers/lemoign2.pdf>, [consulté en ligne le 10 juin 2017]

LUSSO B., « Patrimonialisation et greffes culturelles sur des friches issues de l'industrie minière », *EchoGéo* [En ligne], 26 | 2013, mis en ligne le 19 décembre 2013, consulté le 14 juin 2017. URL : <http://echogeo.revues.org/13645> ; DOI : 10.4000/echogeo.13645

QUET M., « Jack Goody, *Pouvoirs et savoirs de l'écrit* », *Lectures* [En ligne], Les comptes rendus, 2007, mis en ligne le 30 août 2007, consulté le 09 juin 2017. URL : <http://lectures.revues.org/429>

RONDEAU G., Terminologie et documentation, La documentation, Volume 25, numéro 1, mars 1980 - [consulté en ligne le 10 juin 2017]

SALAÜN, J - M « La redocumentarisation, un défi pour les sciences de l'information », *Études de communication* [En ligne], 30 | 2007, mis en ligne le 01 octobre 2009, consulté le 09 juin 2017. URL : <http://edc.revues.org/428> ; DOI : 10.4000/edc.428

TENNIS J.T, « Le poids du langage et de l'action dans l'organisation des connaissances : Position épistémologique, action méthodologique et perspective théorique », *Études de communication* [En ligne], 39 | 2012, mis en ligne le 01 décembre 2014, consulté le 07 juin 2017. URL : <http://edc.revues.org/4010> ; DOI : 10.4000/edc.4010

ZACKLAD M., Évaluation des systèmes d'organisation des connaissance, Les Cahiers du numérique 2010/3 (Vol. 6), [consulté en ligne le 07 juin 2017], URL : <https://www.cairn.info/revue-les-cahiers-du-numerique-2010-3-page-133.htm>

ANNEXES

- Annexe 1 : Présentation du projet Mémo-mine
- Annexe 2 : Modèle de fiche descriptive de la ressource
- Annexe 3 : Expressions régulières
- Annexe 4 : Présentation d'un fichier CSV
- Annexe 5 : Les mots de la mine de B. Turpin
- Annexe 6 : Liste de mots de l'analyse lexicale
- Annexe 7 : Modèle de fiche terminologique



Projet ANR

Explorer les résumés
des projets financés >

Portail thématique >

Financer votre projet >

Documents >

Page précédente >

(DS0807) 2016

Projet **MEMO-MINES**

Conversion des traces mémorielles en médiations numériques : le cas de la mémoire minière

La problématique générale du projet MEMO-MINES porte sur la conservation de faits ou d'événements constituant des patrimoines culturels souvent proches de la disparition (endangered cultural heritage) : elle traite en particulier de la sauvegarde des mémoires individuelles par leur conversion en traces mémorielles, de leur mise à disposition sous forme de corpus d'archives numériques accessibles à tous, de leur remédiation pour des usages sociaux circonscrits, et de leur mise en scène/installation (trans-médiatique) dans le cadre de lieux de mémoire dédiés.

Le projet MEMO-MINES comprend la recherche fondamentale, la conception et la réalisation de dispositifs innovants et leur évaluation en situation écologique d'usage. L'ensemble de la problématique, traitée et analysée dans une perspective interdisciplinaire (sciences de l'information et de la communication, sciences du langage, muséologie, informatique), est appliquée dans le cadre spécifique de la mémoire de la mine.

Le projet MEMO-MINE se propose les 5 défis suivants :

1. la constitution/production de la mémoire minière sous forme 1) d'un choix raisonné de traces mémorielles à numériser ; 2) de la captation d'entretiens préalablement scénarisés avec d'anciens professionnels de la mine (mineurs, ingénieurs, géomètres, techniciens, personnels de santé...) ; 3) de la restitution du cadre de vie social, de travail et personnel de ces acteurs ;
2. l'explicitation de l'univers de connaissance de la mine, la définition et l'élaboration d'un métalangage (d'une ontologie au sens informatique du terme et d'un thesaurus) pour enregistrer les traces mémorielles préalablement captées, pour les analyser (décirer et indexer), pour les publier et diffuser en ligne et, enfin, pour assurer leur pérennité ;
3. la réalisation des archives audiovisuelles de la mémoire minière sous forme d'un portail web dont les accès au fonds préalablement collecté et analysé tiennent compte de l'univers de connaissance spécifique au monde de vie des mineurs ;
4. la conception et la mise en place de dispositifs techniques innovateurs pour l'exploitation – la remédiation – des traces mémorielles numérisées et archivées sous forme de projets cross- ou transmédiés à caractère notamment culturel, ludique et pédagogique, sur site en musée et en consultation numérique distante ;
5. la spécification et la mise en place d'une « plateforme de travail » intégrant les différents outils, procédures et méthodologies pour mener à bien la collecte et l'archivage des traces mémorielles de la mine, leur diffusion via un portail web et leur réutilisation/remédiation.

Partenaires

INALCO/PLIDAM Institut National des Langues et Civilisations Orientales/Pluralité des Langues et des Identités: Didactique, Acquisition et Médiations

UVHC - DeVisu Université de Valenciennes et du Hainaut-Cambrésis - Laboratoire DeVisu

Lille 3 - GERIICO Université Sciences Humaines et Sociales Lille 3

Aide de l'ANR 748 759 euros

Début et durée du projet scientifique janvier 2017 - 48 mois

Annexe 2 : Modèle de fiche descriptive de la ressource

FICHE DESCRIPTIVE

Ressource Sémantique

°1 Lexique minier - Gaillette

Organisme titulaire de la ressource :

Adresse : inconnue

Tel :

Courriel :

Site web :

Type de la ressource	
Titre	
Langue(s)	
Description	
Source	
Zone Géographique	
Structure	
Concepteur	J
Volumétrie	
Format	

