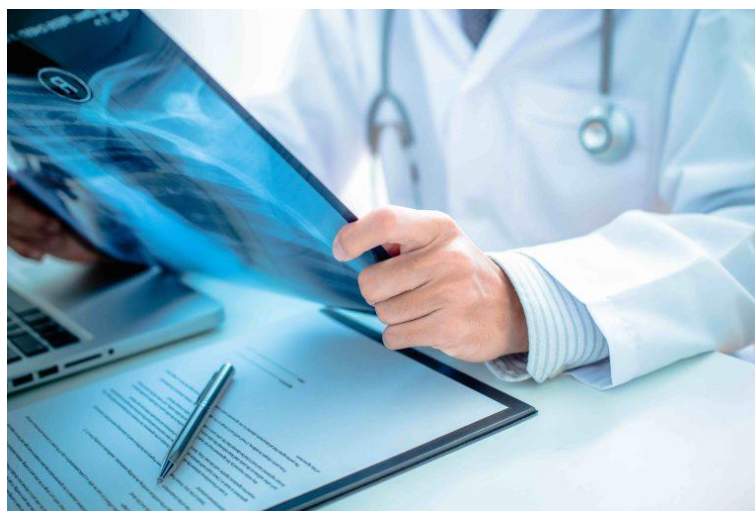


Travail de Bachelor 2018

Automatic classification of 3R radiology reports using text analysis



HES-SO Valais
Filière Informatique de Gestion
www.hevs.ch

Étudiante : Sandy Millius
Professeur : Dr Henning Müller
Déposé le : 30 juillet 2018

Sources des illustrations de la page de titre : <http://www.centre-medical-stmichel.com/radiologie-imagerie/>

TABLE DES MATIÈRES

LISTE DES FIGURES	1
RÉSUMÉ	2
AVANT-PROPOS	3
REMERCIEMENTS	4
LISTE DES ABREVIATIONS	5
GLOSSAIRE	7
1. INTRODUCTION	8
1.1. PROBLÉMATIQUE ET CONTEXTE	8
1.2. OBJECTIFS	8
1.3. ETAPES	8
2. ANALYSE	10
2.1. OUTILS ET LICENCE	10
2.1.1. <i>MetaMap</i>	10
2.1.2. <i>MetaMap Java API</i>	10
2.1.3. <i>MetaMap Lite</i>	10
2.1.4. <i>cTAKES</i>	11
2.1.5. <i>HeTOP</i>	13
2.1.6. <i>HeTOP API</i>	15
2.1.7. <i>ECMT</i>	17
2.1.8. <i>ECMTE</i>	20
2.1.9. <i>Licence UTS</i>	22
2.2. TERMINOLOGIES/ONTOLOGIES	23
2.2.1. <i>Définitions</i>	23
2.2.2. <i>MeSH</i>	23
2.2.3. <i>RadLex</i>	24
2.2.4. <i>SNOMED</i>	24
2.2.5. <i>UMLS</i>	24
2.3. TECHNOLOGIES	25
2.3.1. <i>JAVA</i>	25
2.3.2. <i>C# et .NET</i>	25
2.3.3. <i>Protocoles SOAP et REST</i>	26
2.3.4. <i>XML</i>	26

3. RÉSULTATS DE RECHERCHES	28
3.1. LICENCE UTS.....	28
3.2. METAMAP	29
3.2.1. Aspects négatifs.....	29
3.3. cTAKES.....	30
3.3.1. Tests effectués	30
3.3.2. Avis/Retours	32
3.4. HeTOP ET ECMT	32
3.4.1. Tests effectués sur HeTOP	33
3.4.2. Tests effectués sur ECMT	33
3.4.3. Avis/Retours	34
4. CHOIX	36
4.1. TECHNIQUE	36
4.2. MÉTHODOLOGIE ET PLANNING	36
5. DÉVELOPPEMENT	38
5.1. PARTICULARITÉS TECHNIQUES ET ARCHITECTURE	38
5.2. PHASES DE DÉVELOPPEMENT	39
5.2.1. Récupération de fichier et extraction de texte.....	39
5.2.2. Transformation de texte compatible pour URL	41
5.2.3. Encodage d'autorisations	42
5.2.4. Passage dans API ECMT.....	43
5.2.5. Récupération des résultats depuis ECMT.....	43
5.2.6. Transformation du résultat String en XML	44
5.2.7. Tri et stockage des résultats propres de ECMT	44
5.2.8. Passage des IDs dans API HeTOP	47
5.2.9. Récupération des résultats XML de HeTOP	47
5.2.10. Stockage des résultats propres de HeTOP	49
5.2.11. Comparaison.....	49
6. CONCLUSION	51
6.1. BILAN TECHNIQUE	51
6.2. PROBLÈMES RENCONTRÉS	51
6.3. RESPECT DU CAHIER DES CHARGES	52
6.4. AMÉLIORATIONS FUTURES ET RECOMMANDATIONS	53
6.5. CONCLUSION PERSONNELLE	53

RÉFÉRENCES.....	55
ANNEXE I : FORMULAIRE DE DEMANDE DE LICENCE UTS	56
ANNEXE II : CAHIER DES CHARGES	57
ANNEXE III : JOURNAL DE BORD (1).....	58
ANNEXE IV : JOURNAL DE BORD (2)	59
ANNEXE V : JOURNAL DE BORD (3)	60
ANNEXE VI : JOURNAL DE BORD (4)	61
ANNEXE VII : JOURNAL DES RENDEZ-VOUS	62
ANNEXE VIII : PV RÉUNION CLIENT DU 07.05.2018.....	63
ANNEXE IX : PV RÉUNION CLIENT DU 22.05.2018	64
ANNEXE X : PV RÉUNION CLIENT DU 04.06.2018.....	65
ANNEXE XI : PV RÉUNION CLIENT DU 18.06.2018	66
ANNEXE XII : PV RÉUNION CLIENT DU 09.07.2018.....	67
ANNEXE XIII : PV RÉUNION CLIENT DU 19.07.2018.....	68
ANNEXE XIV : PV RÉUNION CLIENT DU 26.07.2018.....	69
ANNEXE XV : RÉSULTAT XML POUR ECMT GÉNÉRAL SUR INDICATION	70
ANNEXE XVI : RÉSULTAT XML POUR ECMT PARTIEL ET DÉTAILLÉ SUR INDICATION	71
ANNEXE XVII : RÉSULTAT XML POUR HETOP POUR LE TERME « DOS » (1)	72
ANNEXE XVIII : RÉSULTAT XML POUR HETOP POUR LE TERME « DOS »(2)	73
ANNEXE XIX : EXEMPLE DE RAPPORT ANONYMISÉ (1)	74
ANNEXE XX : EXEMPLE DE RAPPORT ANONYMISÉ (2)	75
ANNEXE XXI : DÉTAILS TESTS CTAKES (1).....	76
ANNEXE XXII : DÉTAILS TEST CTAKES (2).....	77
ANNEXE XXIII : DÉTAILS TESTS CTAKES (3).....	78
DÉCLARATION DE L'AUTEUR	79

LISTE DES FIGURES

Figure 1 : cTAKES - Outil de matching	12
Figure 2 : cTAKES - Fonctionnalité de ligne temporelle	13
Figure 3 : HeTOP - Réseau T/O.....	14
Figure 4 : HeTOP	15
Figure 5 : HeTOP API - Options	16
Figure 6 : HeTOP API - Balises de réponse	17
Figure 7 : ECMT - Réponses	18
Figure 8 : ECMT - Recherche.....	19
Figure 9 : ECMT - Résultat	20
Figure 10 : ECMTE - Recherche.....	21
Figure 11 : ECMTE - Résultat mise en évidence du texte.....	21
Figure 12 : ECMTE - Résultat matching de mots selon leur concept.....	22
Figure 13 : XML - Exemple structure	27
Figure 14 : Problème Licence UTS - Email de confirmation de demande.....	28
Figure 15 : Problème License UTS - Email de refus.....	29

RÉSUMÉ

L'objectif principal de ce travail de Bachelor est de fournir un prototype permettant d'effectuer une analyse de l'exactitude des rapports en radiologie. Ce prototype se doit d'être basique car le réel but est de définir la faisabilité d'un tel outil d'amélioration continue. Afin de pouvoir analyser la qualité du rapport, l'indication et la conclusion de celui-ci doivent être comparées. Si les principaux termes de la conclusion correspondent aux principaux termes de l'indication, le taux de probabilité d'exactitude du rapport est très élevé ; il s'agit donc d'un rapport de bonne qualité. A contrario, si ce taux se retrouve être fréquemment bas, une solution de formation interne devra être mise en place par les clients afin d'améliorer la qualité d'écriture des rapports par les radiologues.

Ce projet, mené en collaboration avec l'Institut de Radiologie de Sion, est principalement constitué de recherches et de tests réalisés sur différents outils existants permettant d'effectuer une partie du travail nécessaire. En effet, afin de pouvoir comparer deux termes, il faut être capable de les trouver dans une certaine terminologie/ontologie. Ainsi, il est possible de comparer leurs arborescences et de voir s'ils sont suffisamment proches pour être jugés semblables.

Le travail est organisé en deux parties. Dans un premier temps, une analyse de l'état de l'art, des recherches et des tests sont effectués et retranscrits. Dans un deuxième temps, suite aux résultats obtenus, le développement d'un prototype est réalisé. Le fonctionnement de ce prototype se base sur l'utilisation de deux APIs complémentaires et nécessaires à la récupération des termes et concepts dans une certaine terminologie.

Mots-clés : radiologie, rapports, terminologie/ontologie, analyse de texte.

AVANT-PROPOS

Le thème de ce travail de Bachelor a été proposé par mon professeur responsable, Monsieur Dr. Henning Müller. Nous avons étroitement travaillé avec le cabinet privé de radiologie 3R-Réseau Radiologie Romand, ce qui a impliqué d'analyser de vrais rapports. Afin de ne pas perturber le secret médical, tous les rapports qui m'ont été transmis par le cabinet ont bien évidemment été anonymisés.

Ce travail de Bachelor répond à une problématique concrète dans le domaine de la radiologie, mais qui peut s'appliquer à tous les domaines médicaux : l'analyse de l'exactitude des rapports entrés par les radiologues à l'aide de la comparaison entre l'indication et la conclusion. À ce stade, il ne constitue que la réalisation d'un simple prototype. Des améliorations futures pourront bien évidemment être apportées afin d'améliorer l'outil et de l'adapter en fonction du secteur demandé.

REMERCIEMENTS

Je tenais tout particulièrement à remercier les personnes suivantes qui m'ont aidé et ont rendu la réalisation de ce projet possible :

Dr Henning Müller pour avoir été mon professeur référent durant ce projet. Très impliqué durant ces trois mois, il a fait preuve d'une grande disponibilité et m'a fourni de nombreux conseils.

Dr Hughes Brat et M. Cyril Thouly de l'Institut de Radiologie de Sion pour avoir été les clients de ce projet, pour la confiance témoignée et pour m'avoir fourni des rapports radiologiques anonymisés.

Dr Stefan Darmoni et M. Julien GrosJean de l'équipe CISMeF du CHU-Rouen pour m'avoir aiguillé dans la compréhension de leurs outils (HeTOP, ECMT et ECMTE) et pour l'intérêt qu'ils portent à la traduction de RadLex.

Mme Julie Préperier pour son apport en connaissances techniques de développement Java et pour son soutien durant ce projet.

M. Luca Centofanti pour la relecture du document ainsi que pour son soutien durant ce projet.

LISTE DES ABREVIATIONS

NLM = National Library of Medicine

UMLS = Unified Medical Language System

NLP = Natural-language processing

MTI = Medical Text Indexer

API = Application Programming Interface

NIH = National Institutes of Health

UTS = UMLS Terminology Services

HeTOP = Health Terminology/Ontology Portal

T/O = Terminologies et ontologies

CISMeF = Catalogue et Index des Sites Médicaux de langue Française

URI = Uniform Resource Identifier

XML = Extensible Markup Language

ECMT = Extracteur de Concepts Multi-Terminologique

ECMTE = ECMT Evaluator

MeSH = Medical Subject Headings

RadLex = Radiology Lexicon

RSNA = Radiological Society of North America

SNOMED = Systematized Nomenclature of Human and Veterinary Medicine

SNOMED CT = Systematized Nomenclature of Human and Veterinary Medicine - Clinical Terms

UMLS = Unified Medical Language System

SOAP = Simple Object Access Protocol

REST = Representational State Transfer

POO = Programmation Orientée Objet

HTTP = Hypertext Transfer Protocol

CSV = Comma-Separated Values

JSON = JavaScript Object Notation

GLOSSAIRE

Output : résultat de données en sortie d'un système informatique (programme, API, etc.).

Matching / Mapping : lorsqu'un terme concorde avec un autre ou, dans notre cas, avec des termes d'une terminologie.

Plugin : il s'agit d'un petit logiciel qu'on greffe à un autre programme afin qu'il lui apporte des fonctionnalités supplémentaires.

Base64 : conçu dans le but d'être compatible avec la plupart des systèmes, base64 est en réalité un codage de 64 caractères. Son utilisation la plus connue concerne la transmission de courrier électronique.

Daily scrum : appliqué à la méthodologie Scrum, il s'agit d'une réunion journalière de toute l'équipe afin d'avoir une discussion sur l'avancement d'un sprint de projet et des éventuels problèmes rencontrés.

1. Introduction

1.1. Problématique et contexte

Dans un cabinet de radiologie, de nombreux rapports sont entrés par jour par divers radiologues. En effet, dans le cas de notre client, il s'agit de plus de 400 rapports à travers tout leur réseau romand. Bien qu'une certaine rigueur dans la rédaction et la structure des rapports existe, il est fréquent que des petites fautes apparaissent. En effet, après quelques recherches manuelles en interne, il s'est avéré que le problème le plus fréquemment rencontré était le suivant : l'indication présente au début du rapport ne trouvait pas de conclusion correspondante précise à la fin de celui-ci. Ce genre de problème peut être dû à de nombreux facteurs. En effet, le radiologue peut simplement n'avoir fourni qu'une réponse peu précise ou avoir posé une question peu claire. Il arrive également, et ceci constitue le pire cas, qu'un rapport qui n'a pas été écrit en une fois comporte un mix de patients : l'indication pour l'un et la conclusion de l'autre. Bien que rares, ces cas peuvent poser un grave problème. Ces fautes sont bien évidemment dues au facteur humain et l'entreprise est convaincue qu'une formation pourrait améliorer ces résultats. Elle a donc besoin d'un outil permettant d'analyser des rapports en comparant leurs indications et leurs conclusions, et qui retourne un résultat du degré d'exactitude de ces derniers. Avec un tel outil et une formation interne mise en place, il sera possible par la suite de voir si ces mesures ont eu un réel effet d'amélioration dans la rédaction des rapports ; une amélioration continue étant donc leur but principal.

1.2. Objectifs

Afin de mener à bien le but de notre client, ce projet regroupe de nombreux objectifs. Tout d'abord, il est essentiel de pouvoir effectuer du mapping de texte, dans notre cas en français, sur différentes terminologies. Le but étant bien évidemment de pouvoir ressortir les termes essentiels à l'analyse et à la comparaison. La négation de ces termes doit être gérée. Il faut donc pouvoir ressortir les termes de l'indication et de la conclusion d'un rapport afin de pouvoir les comparer. Le résultat de la comparaison devra être stocké afin de pouvoir en tirer des conclusions. Finalement, une automatisation du processus d'extraction/comparaison peut s'avérer nécessaire pour la gestion d'un grand nombre de documents à la fois.

1.3. Etapes

Plusieurs étapes sont nécessaires à la réalisation de ce travail. Dans un premier temps, une analyse complète est réalisée afin de connaître l'existant et de permettre d'effectuer des choix qui soient les plus pertinents possible pour le développement du prototype. Des recherches et des tests sont bien évidemment menés afin d'affiner ces choix. S'en suit une phase de développement du prototype qui est départagée en plusieurs parties : l'extraction de texte, l'utilisation d'APIs

existantes afin d'obtenir des résultats de mapping sur des terminologies, la récupération d'outputs et le traitement de ces derniers, la comparaison des résultats et pour finir la possibilité d'automatiser le processus.

2. Analyse

2.1. Outils et licence

2.1.1. MetaMap

MetaMap est outil de mapping de texte hautement configurable et open source développé par le Dr Alan Aronson à la National Library of Medicine (NLM). Ce programme a pour but d'analyser un texte donné en entrée (d'origine médicale), de trouver les termes importants et de les faire correspondre avec des termes contenus dans UMLS Metathesaurus, qui est une immense bibliothèque de dictionnaires numériques du domaine médical. Il se base principalement sur le traitement du langage naturel (NLP) afin de pouvoir distinguer les différentes parties d'un texte, d'une phrase ou même d'un mot.¹

MetaMap constitue un fondement de l'indexeur de textes médicaux (MTI) car il permet une indexation semi-automatique/automatique de la littérature rattachée à NLM. C'est un logiciel qui utilise le langage Java et qui nécessite d'avoir soit un système d'exploitation Linux, soit un système d'exploitation Windows XP ou 7. Les différentes étapes d'installation et d'utilisation se passent directement en lignes de commandes, ce qui nécessite d'être à l'aise avec ces dernières. MetaMap constitue la base mais il existe plusieurs dérivés, créés toujours par la même personne, ayant une utilisation légèrement différente. Ces derniers seront expliqués ci-dessous.¹

2.1.2. MetaMap Java API

MetaMap Java API constitue la version API de MetaMap et permet donc un accès à son moteur de cartographie via des programmes codés en Java.²

L'installation de cette version requière quelques étapes préliminaires : avoir téléchargé la version complète de de MetaMap, disposer d'un environnement d'exécution Java 1.8 ou postérieur et bien évidemment d'avoir souscrit à une licence UTS. La dernière version à jour de l'API (2016v2) n'est utilisable que sur Linux ou Mac OS/X. Il est cependant possible d'utiliser une version antérieure (2014) sur Windows XP ou 7.²

2.1.3. MetaMap Lite

MetaMap Lite a été conçu afin de proposer le même type de services que MetaMap tout en étant un outil considérablement plus rapide. Une telle différence est possible dû au manque de rigueur de

¹ <https://metamap.nlm.nih.gov/> consulté le 10.05.2018

² <https://metamap.nlm.nih.gov/JavaApi.shtml> consulté le 10.05.2018

MetaMap Lite par rapport à la version standard et à la personnalisation des diverses options.³ Dans les options qui sont facultatives, on retrouve :

- Restriction par type sémantique et par source UMLS ³
- Marquage de parties de discours ³
- Détection de négation ³

Ce dernier point est géré selon deux possibilités, au choix : le contexte de Wendy Chapman ou un algorithme basé sur le NegEx de Wendy Chapman (à savoir que la seconde solution est plus rapide, mais moins efficace). Il est donc évident que MetaMap Lite apparaît comme un outil tout à fait modulable selon les besoins et spécificités liés à chaque projet. Les développeurs du NIH travaille activement sur cet outil et les mises à jour sont plus fréquentes que pour MetaMap et MetaMap Java API. ³

Il est important de mentionner que les prérequis à l'installation de MetaMap Lite peuvent être plus conséquents que ceux des autres outils. En effet, en plus des habituelles prérequis, il est nécessaire de prévoir au minimum 16GB d'espace disque ainsi que 2GB de Ram minimum (4GB conseillés). ³

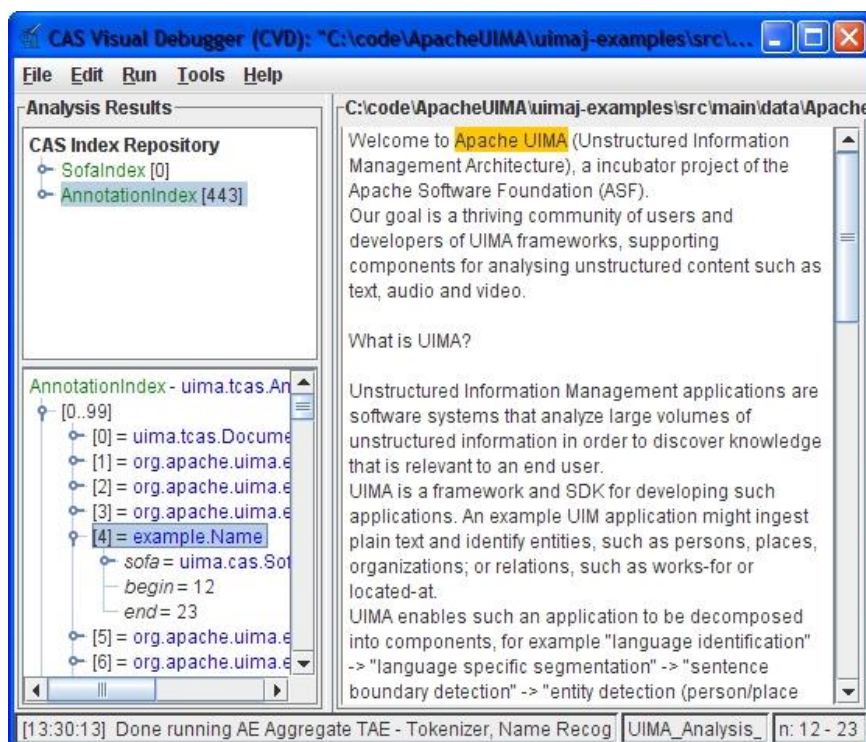
2.1.4. cTAKES

cTAKES est un outil doté d'une interface graphique servant au traitement du langage naturel dans un contexte médical. En effet, son but est d'extraire des informations à partir de rapports médicaux contenant du texte libre. A partir de ces extractions, il est possible d'analyser les résultats selon ses besoins. cTAKES a initialement été conçu par la Clinique Mayo, mais se rendant compte de l'importance d'un tel outil dans le domaine médical, les développeurs ont décidé d'en faire un outil open source et gratuit afin de permettre à différentes institutions de divers pays de l'utiliser. Il comporte de nombreux autres avantages tels que : sa puissance, sa rapidité, son évolutivité, sa modularité et sa portabilité. ⁴

³ <https://metamap.nlm.nih.gov/MetaMapLite.shtml> consulté le 10.05.2018

⁴ <http://ctakes.apache.org/> consulté le 22.05.2018

Figure 1 : cTAKES - Outil de matching

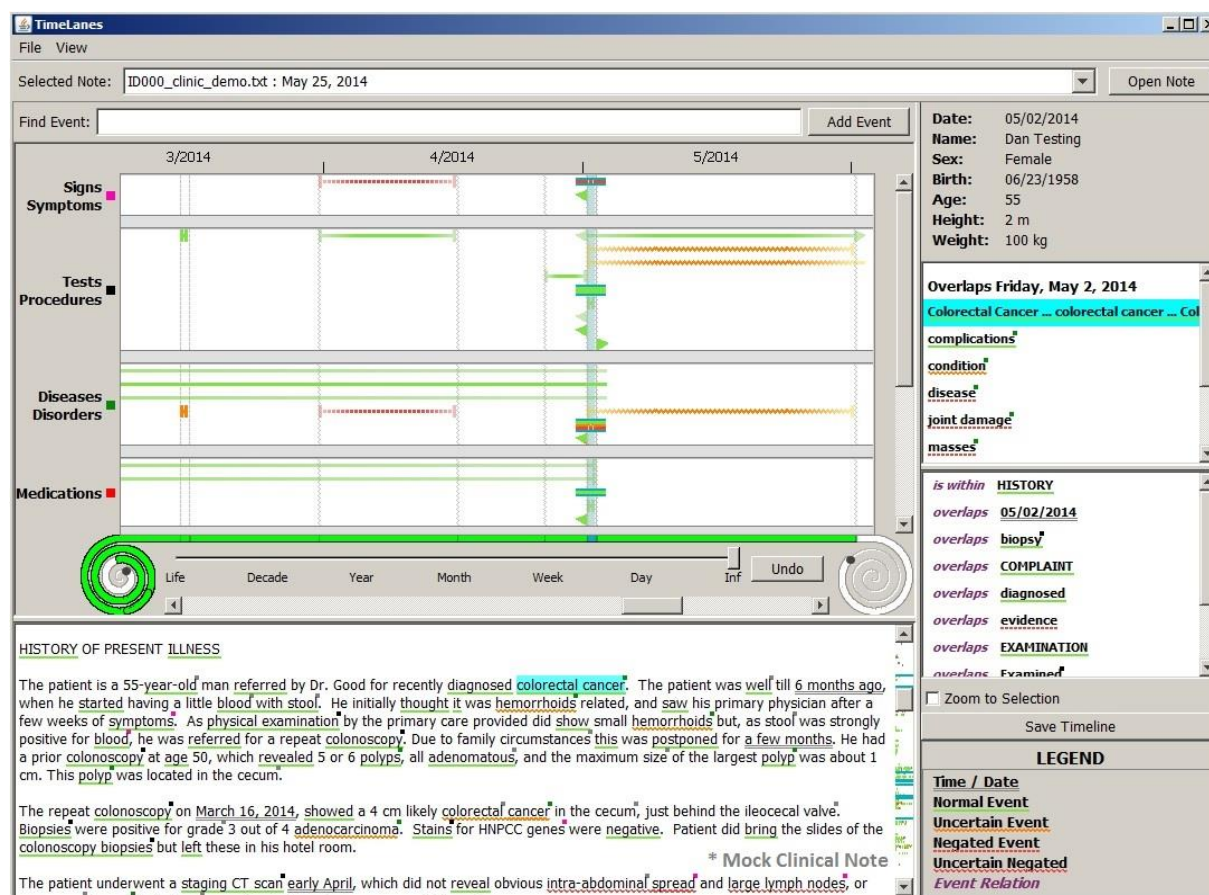


Sources : CvdMain1 [554x472, JPG] Récupéré sur <http://ctakes.apache.org/examples.html>

cTAKES comporte de nombreuses fonctionnalités. En plus de simplement effectuer un matching sur des termes d'un texte avec leurs concepts dans des terminologies, il est également possible de personnaliser les dictionnaires souhaités. En effet, l'utilisateur peut, s'il le souhaite et au besoin, créer son propre dictionnaire avec les terminologies et ontologies souhaitées et de l'importer dans le programme afin d'effectuer un matching personnalisé. cTAKES permet également de faire un matching sur les termes liés au temps et aux dates. Ainsi, il peut recréer une ligne temporelle et placer sur cette dernière, les différents événements liés et détectés grâce aux termes. ⁵

⁵ <http://ctakes.apache.org/examples.html> consulté le 22.05.2018

Figure 2 : cTAKES - Fonctionnalité de ligne temporelle



Sources : TimeLanesBig [1087x793, JPG] Récupéré sur <http://ctakes.apache.org/examples.html>

2.1.5. HeTOP

HeTOP est un portail conçu pour effectuer des recherches terminologiques mis à disposition gratuitement aux utilisateurs tant que son utilisation reste dans un cadre de recherches ou d'activités académiques. Il a été « développé [sic] par l'équipe CISMef (Département d'Informatique et d'Information Médicales), CHU hôpitaux de Rouen, France. »⁶ Très complet, il permet l'accès à plus de 70 terminologies et ontologies (T/O) concernant le domaine de la Santé. Ces dernières sont généralement spécifiques à un domaine précis de la Santé (maladies, anatomie, etc.).⁶ Les différentes particularités des ontologies seront traitées plus précisément au chapitre 2.2.* de ce document.

« HeTOP repose sur un méta-modèle multi-terminologique pour intégrer toutes les T/O dans sa base de données. Il est interlingue car les T/O sont souvent disponibles en plusieurs langues. De

⁶ https://www.hetop.eu/hetop/documentation/about_fr.html/hetop/documentation/about_fr.html consulté le 25.06.2018

plus, les T/O sont liées entre elles via des alignements et constituent ainsi un véritable réseau permettant l'interopérabilité sémantique. » ⁷

Figure 3 : HeTOP - Réseau T/O



Sources : Screenshot_HeTOP_Reseau [716x784, PNG] Illustration réalisée par l'auteur sur
https://www.hetop.eu/hetop/documentation/sem_net.htm

Ce portail possède une interface graphique le rendant simple d'utilisation. Il suffit de taper le terme souhaité dans la barre de recherche et de valider. Une série de résultats classés apparaît (cf. figure 4). Il est donc possible d'accéder à de nombreuses informations telles que l'ontologie, l'identifiant, les traductions ainsi que différents synonymes et bien d'autres. Il est également possible de s'intéresser à la hiérarchie et aux relations du terme recherché. Par défaut, les

⁷ https://www.hetop.eu/hetop/documentation/about_fr.html/hetop/documentation/about_fr.html consulté le 25.06.2018

recherches s'effectuent dans la terminologie CISMef ainsi qu'au sein de MeSH. ⁸ Il reste cependant bien évidemment possible de personnaliser notre recherche selon les T/O souhaité à l'aide du menu déroulant sur la gauche.

Figure 4 : HeTOP

The screenshot displays the HeTOP web interface. At the top, there is a search bar with the text 'fractur' and a language dropdown set to 'français'. Below the search bar, there are checkboxes for 'Recherche sans troncature (sans wildcard)' and 'Sélection terminologies' (checked). The main content area is divided into two columns. The left column, titled 'Vos recherches', shows '119 ressource(s) trouvée(s) en 0,02 s' and a list of 'Meilleurs candidats' including 'fractures osseuses [Descripteur MeSH]', 'cryofracture [Descripteur MeSH]', 'fractures de la mâchoire [Descripteur MeSH]', 'fractures de côte [Descripteur MeSH]', 'fractures ouvertes [Descripteur MeSH]', 'fractures de la hanche [Descripteur MeSH]', 'fractures de l'ulna [Descripteur MeSH]', 'risque fracturaire [Stratégie de Recherche CISMef]', 'Lésions d'écrasement [Descripteur MeSH]', and 'fractures du fémur [Descripteur MeSH]'. The right column, titled 'Fractures osseuses (Descripteur MeSH)', shows tabs for 'Description', 'Hiérarchies', 'Relations', and 'PubMed / DocCISMef'. It includes logos for NLM, Inserm, BioPortal, and RDF/XML. Below these, there is a language selection dropdown set to 'Off'. The 'Libellé préféré' section shows 'fractures osseuses' (French) and 'fractures, bone' (English). The 'Identifiant d'origine' is 'D050723'. The 'Définition du MeSH' is 'Breaks in bones.' The 'Synonyme CISMef' is 'fractures'. The 'Synonyme DeCS' and 'Synonyme MeSH' sections list various synonyms in French, English, and other languages, including 'Fractures de l'os', 'Fractures des os', 'Os fracturés', 'bone fracture', 'bone fractures', 'bone, broken', 'bones, broken', 'broken bone', 'broken bones', 'fracture, bone', 'Fracture, Spiral', 'Fracture, Torsion', 'Fractures, Spiral', 'Fractures, Torsion', 'Spiral Fracture', 'Torsion Fracture', and 'Torsion Fractures'. The 'CUI UMLS' is 'C0016658'.

Sources : Screenshot_HeTOP [1277x856, PNG] Illustration réalisée par l'auteur sur
https://www.hetop.eu/hetop/#rr=MSH_D_050723&q=fracture

2.1.6. HeTOP API

Les fonctionnalités proposées par HeTOP sont également accessibles sous forme de Service Web. Il est pour cela possible d'utiliser le protocole REST ou le protocole SOAP, selon les besoins et connaissances de l'utilisateur. Une authentification reste cependant nécessaire afin de pouvoir utiliser ces services. Il est donc important de faire une demande afin de bénéficier d'un identifiant et d'un mot de passe. Il est possible de tester ces services directement via un navigateur en utilisant un plugin en fonction de ce dernier (RESTClient pour Mozilla Firefox et Client Insomnia REST pour Google Chrome). L'URI à tester est la suivante :

⁸ https://www.hetop.eu/hetop/documentation/about_fr.html/hetop/documentation/about_fr.html consulté le 25.06.2018

<http://www.hetop.fr/CISMeFhetopservice/REST/searchConcepts/asthme/fr/>. Les résultats retournés par le Service Web sont au format XML. Ce service peut bien évidemment être utilisé comme API à ajouter dans un programme. L'authentification reste bien évidemment toujours nécessaire, c'est pourquoi un encodage de l'identifiant et du mot de passe en base64 peut s'avérer nécessaire.⁹

Diverses options peuvent être ajoutées à la requête afin de rendre le résultat le plus correct et précis possible. Elles se rajoutent à l'URL sous forme d'une chaîne de caractères telle que suit : `a=true&d=false&sn=true&rw=true&at=true`. L'ordre de ces dernières n'a pas d'influence sur le résultat. Les différentes arguments ajoutables et leurs explications apparaissent dans la figure 5 ci-dessous.⁹

Figure 5 : HeTOP API - Options

Argument	Forme développée	Type	Description
c	<i>categorization</i> (catégorisation)	booléen	À <code>true</code> , cela retournera les métatermes (super-concepts) et les Types Sémantiques de l'UMLS des concepts trouvés. La valeur par défaut est <code>true</code> .
sn	<i>semantic network</i> (réseau sémantique)	booléen	À <code>true</code> , les concepts alignés (exactement) aux concepts trouvés seront également retournés. Les relations d'alignements concernées sont les "Alignements exacts supervisés CISMeF" et les "Alignements manuels CISMeF". La valeur par défaut est <code>true</code> .
e	exclusions	identifiants de concepts ou types de concepts	Les identifiants donnés seront exclus de la réponse. Exemple : <code>e=CIS_MT_8,UML_ST_T060,MSH_D_C,T_DESC_PHARMA_RACINE</code> ne retournera pas les concepts liés à chirurgie (concept d'id CIS_MT_8), à procédure de diagnostic (concept d'id UML_ST_T060), les descendants de maladies (concept d'id MSH_D_C) et toutes les Racines Pharmaceutiques (type de concepts T_DESC_PHARMA_RACINE). La valeur par défaut est vide.
f	<i>filters</i> (filtres)	types de concepts	Seuls les concepts appartenant aux types spécifiés seront retournés. Exemple : <code>f=T_DESC_MESH_DESCRIPTEUR,T_DESC_SNOMED_NOTION</code> ne retournera que les descripteurs MeSH (T_DESC_MESH_DESCRIPTEUR) et les notions SNOMED int. (T_DESC_SNOMED_NOTION). La valeur par défaut est vide.
a	<i>ancestors</i> (ancêtres)	booléen	À <code>true</code> , cela retournera les ancêtres (à partir de leur hiérarchies : parents, grand-parents, etc.) des concepts trouvés. La valeur par défaut est <code>true</code> .
d	descendants	booléen	À <code>true</code> , cela retournera les descendants (à partir de leur hiérarchies : enfants, petits-enfants, etc.) des concepts trouvés. La valeur par défaut est <code>false</code> .
at	<i>alternative terms</i> (termes alternatifs)	booléen	À <code>true</code> , les termes alternatifs (~synonymes) des concepts trouvés seront également retournés. La valeur par défaut est <code>true</code> .

Sources : Screenshot_HeTOPAPI_Options [1326x648, PNG] Illustration réalisée par l'auteur sur

https://www.hetop.eu/hetop/documentation/ws_fr.html

Afin d'améliorer la compression du fichier XML sortant, une liste des différentes balises possibles a été établie par les créateurs de l'API (cf. figure 6).

⁹ https://www.hetop.eu/hetop/documentation/ws_fr.html consulté le 25.06.2018

Figure 6 : HeTOP API - Balises de réponse

Balise	Description
cis:dboResp	Liste de concepts
cis:dbo	Concept
cis:id	Identifiant de concept
cis:ti	Identifiant de métadonnée (concept, propriété ou relation)
cis:lb	Libellé de concept ou de métadonnée
cis:an	Annotation de concept ou de métadonnée
cis:dps	Liste d'attributs (<i>data property</i>)
cis:dp	Attribut (<i>data property</i>)
cis:hies	Liste de relations hiérarchiques
cis:hie	Relation hiérarchique
cis:bt	Concept "large" (<i>broader</i>)
cis:nt	Concept "étroit" (<i>narrower</i>)
cis:path	Chemin hiérarchique
cis:ops	Liste de relations
cis:op	Relation (entre concepts)
cis:tar	Identifiant de concept cible

Sources : Screenshot_HeTOPAPI_Reponse [1071x711, PNG] Illustration réalisée par l'auteur sur
https://www.hetop.eu/hetop/documentation/ws_fr.html

2.1.7. ECMT

« L'Extracteur de Concepts Multi-Terminologique (ECMT) permet d'annoter automatiquement des textes à l'aide des concepts des principales terminologies de santé disponibles en Français (c'est-à-dire incluses dans HeTOP). »¹⁰ Cet outil a également été développé par l'équipe CISMef et a pour but de compléter HeTOP. En effet, ECMT ne se concentre pas uniquement sur la recherche d'un seul terme, mais sur l'annotation de termes dans un texte donné par l'utilisateur. Les arguments applicables à la recherche dans le Service Web de HeTOP s'appliquent également pour ECMT. Trois arguments supplémentaires ont néanmoins été introduits :

- « "r" : refined. Si "r=true" l'algorithme s'arrête lorsqu'il trouve un concept couvrant un maximum de termes (valeur par défaut : "true"). Exemple : pour "cardiopathie hypertensive", si "r=true" on renvoie[*sic*] "hypertension artérielle" uniquement, si "r=false", retourne "hypertension artérielle" et "maladie cardiaque" (ce dernier du fait de "cardiopathie"). »¹¹

¹⁰ <https://ecmt.chu-rouen.fr/> consulté le 28.06.2018

¹¹ <https://ecmt.chu-rouen.fr/doc/howto-fr.html> consulté le 28.06.2018

- « "p" : priorization. Si "p=true", seul le concept issu de la terminologie la plus pertinente est retourné, en fonction de son type sémantique. » ¹²
- « "amin" : acronyms minimal size. Ne renvoie que les acronymes de taille >= amin. » ¹²

De même que pour les arguments, dans ECMT, la liste des réponses possibles est bien plus complète que celle de HeTOP (cf. figure 7).

Figure 7 : ECMT - Réponses

- **cis-sentences** : liste des phrases
- **timemillis** : temps de traitement en ms
- **cis-sentence** : une phrase
- **idsentence** : numéro de la phrase
- **position** : position du début de la phrase dans le texte
- **idpatient** : déprécié
- **iddoc** : déprécié
- **cis:indexation** : une indexation donnée
- **start** : position de début de l'indexation
- **end** : position de fin de l'indexation
- **idterm** : identifiant du concept dans sa terminologie d'origine
- **idcismef** : identifiant du concept dans le système d'information CISMef
- **offset** : liste des positions des termes composant le concept
- **ter** : acronyme de la terminologie du concept
- **umlsui** : identifiant UMLS du concept
- **matchterms** : liste des libellés ayant permis de retrouver le concept
- **typeid** : type_id CISMef identifiant la nature du concept
- **condition** : élément modifiant la portée de l'information
- **context** : information pouvant être utilement rattachée au concept trouvé
- **cis:term** : libellé préféré du concept
- **cis:label** : chaîne de caractère
- **lang** : langue de la chaîne de caractère
- **cis:altterms** : liste des libellés alternatifs du concept
- **cis:altterm** : libellé alternatif du concept (synonyme)
- **cis:categorization** : liste des métatermes ou types sémantiques
- **cis:category** : un métaterme ou type sémantique
- **idcategory** : identifiant de la catégorie dans sa terminologie d'origine
- **idcismef** : identifiant de la catégorie dans le système d'information CISMef
- **origin** : origine de la catégorie : cismef, umls, ou autre
- **cis:descendants** : liste des concepts en descendance hiérarchique avec le concept (enfants, petits-enfants, ...)
- **cis:descendant** : un concept en descendance hiérarchique avec le concept (enfant, petit-enfant, ...)
- **cis:ancestors** : liste des concepts ancêtres hiérarchiques du concept (parents, grands-parents, ...)
- **cis:ancestor** : un concept ancêtre hiérarchique du concept (parent, grands-parent, ...)
- **cis:relateds** : liste des concepts en relation sémantique avec le concept (parmi des natures de relations fiables choisies)
- **cis:related** : un concept en relation sémantique avec le concept (parmi des natures de relations fiables choisies)
- **relationTypeId** : identifiant interne CISMef de la nature de la relation
- **relationLabel** : libellé de la nature de la relation

Sources : Screenshot_ECMT_Reponse [867x771, PNG] Illustration réalisée par l'auteur sur <https://ecmt.chu-rouen.fr/doc/howto-fr.html>

ECMT comporte l'avantage d'être un Web Service et d'être lui aussi callable en REST ou SOAP. Ainsi, il est utilisable telle une API en faisant appel vers l'URI <https://ecmt.chu-rouen.fr/CISMefecmtservice/REST/getAutomaticIndexingWithOptions/at=true&a=false&d=false&r=t>

¹² <https://ecmt.chu-rouen.fr/doc/howto-fr.html> consulté le 28.06.2018

[rue=&c=true&sn=true&e=&f=/](#) suivie du texte de recherche souhaité. Cette adresse mène sur une petite interface graphique afin de faciliter son utilisation (cf. figure 8).

Figure 8 : ECMT - Recherche

Extracteur de Concepts Multi-Terminologique (ECMT v3)

[How-to](#) - [Contact](#) - © 2017 CHU de Rouen - CISMef.

Taper ou coller votre texte ici

options : at ☒ a ☐ d ☐ r ☐ c ☐ sn ☐ p ☐ e f

Effacer Annoter

Sources : Screenshot_ECMT_Recherche [1062x575, PNG] Illustration réalisée par l'auteur sur <https://ecmt.chu-rouen.fr/CISMeFecmtservice/REST/getAutomaticIndexingWithOptions/at=true&a=false&d=false&r=true&c=true&sn=true&p=false&e=&f=/>

Une fois le texte voulu inséré et la recherche exécutée, un résultat s'affiche sur l'interface graphique. Le format de ce résultat est, comme pour HeTOP, du XML. ECMT va donc annoter les termes qui ont une correspondance dans différentes T/O. Ces termes sont visibles directement dans le texte entré par l'utilisateur car ils apparaissent en vert, et sont également repris dans une liste comportant leur terminologie et leur code (cf. figure 9).

Figure 9 : ECMT - Résultat

Extracteur de Concepts Multi-Terminologique (ECMT v3)

[How-to](#) - [Contact](#) - © 2017 CHU de Rouen - CISMeF.

Douleurs basi-thoraciques gauches et au niveau de l'hypochondre gauche, avec irradiation vers le dos.

Effacer

1 phrases annotées en 3 ms. 21 codes distincts identifiés.

Codes identifiés

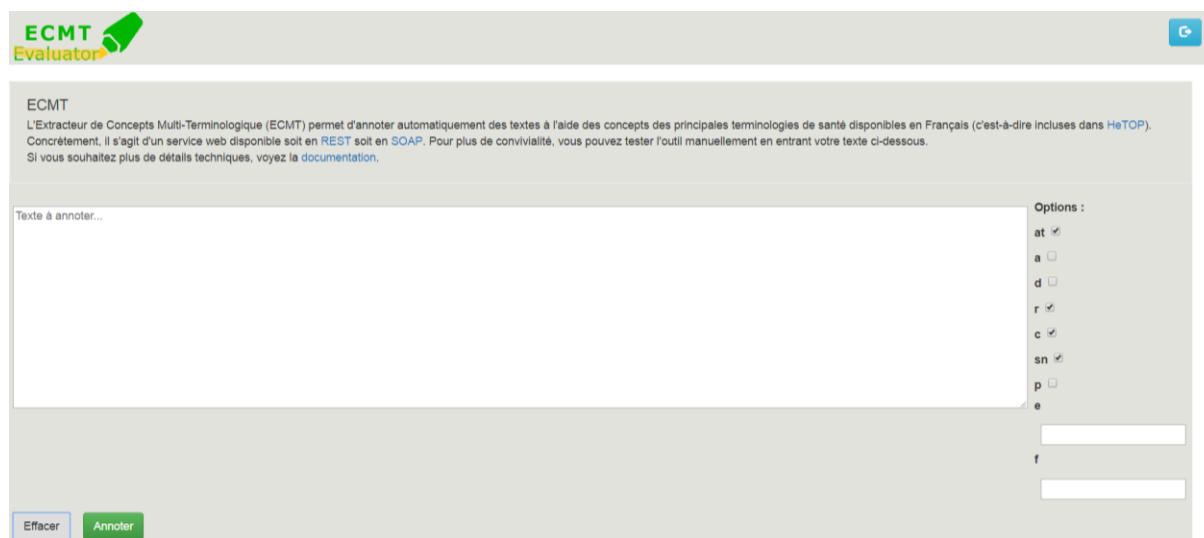
Terme	Ter. Code	CUI	Cond. Ctxt.
Dos	FMA 14181		
Dos	MSH M0002114		
Dos	NCI C13062		
Dos	RAD RID30781		
Dos	TSP 003965		
dos	MSH D001415		
Douleur	MSH M0015742		
Douleur	RAD RID28803		
Douleur	TSP 003994		
douleur	CIS MT154		
douleur	MED T351		
douleur	MSH D010146		
douleur	NCI C3303		
douleur	SNO F-A2600		
Exposition aux radiations	MSH D000069079		
Hypochondre gauche	FMA 20392		
Hypochondre gauche	RAD RID30002		
Irradiation	TSP 006679		
niveau	NCI C25554	C0441889	
Régions du dos	FMA 71938		
sans douleur	NCI C25272		

Sources : Screenshot_ECMT_Resultat [1542x687, PNG] Illustration réalisée par l'auteur sur <https://ecmt.chu-rouen.fr/CISMeFecmts-service/REST/getAutomaticIndexingWithOptions/at=true&a=false&d=false&r=true&c=true&sn=true&p=false&e=&f=/Douleurs%20basi-thoraciques%20gauches%20et%20au%20niveau%20de%20l'hypochondre%20gauche%2C%20avec%20irradiation%20vers%20le%20dos>

2.1.8. ECMTE

ECMTE (ECMT Evaluator) constitue la vraie version graphique du service ECMT. Son principe est donc le même avec un visuel qui diffère légèrement. Deux différences principales sont à noter : 1) ce service n'est pas callable sous forme d'API avec REST ou SOAP, 2) les résultats donnés pour la liste de termes annotés permettent d'accéder directement à la page web HeTOP correspondante au terme dans un concept donné.

Figure 10 : ECMTE - Recherche



ECMT Evaluator

ECMT
L'Extracteur de Concepts Multi-Terminologique (ECMT) permet d'annoter automatiquement des textes à l'aide des concepts des principales terminologies de santé disponibles en Français (c'est-à-dire incluses dans HeTOP). Concrètement, il s'agit d'un service web disponible soit en [REST](#) soit en [SOAP](#). Pour plus de convivialité, vous pouvez tester l'outil manuellement en entrant votre texte ci-dessous. Si vous souhaitez plus de détails techniques, voyez la [documentation](#).

Texte à annoter...

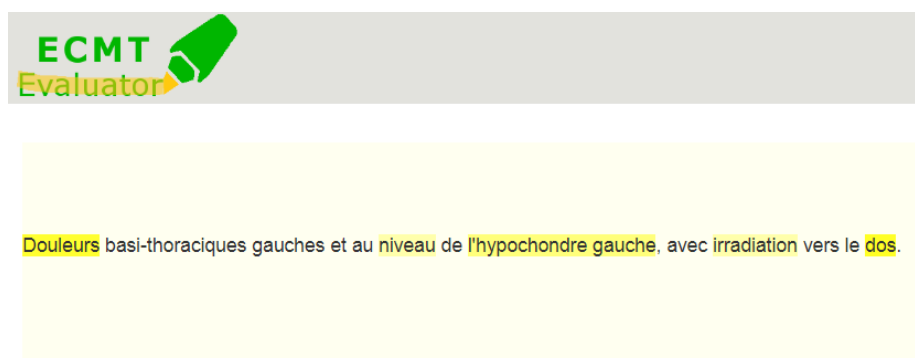
Options :

- at ☒
- a ☐
- d ☐
- r ☒
- c ☒
- sn ☒
- p ☐
- e ☐
- f ☐

Effacer Annoter

Sources : Screenshot_ECMTE_Recherche [1861x881, PNG] Illustration réalisée par l'auteur sur <https://ecmt.chu-rouen.fr/ECMTE/>

Figure 11 : ECMTE - Résultat mise en évidence du texte



ECMT Evaluator

Douleurs basi-thoraciques gauches et au niveau de l'hypochondre gauche, avec irradiation vers le dos.

Sources : Screenshot_ECMTE_Recherche [836x321, PNG] Illustration réalisée par l'auteur sur <https://cispro.chu-rouen.fr/ECMTE/?>

Figure 12 : ECMTE - Résultat matching de mots selon leur concept

Afficher éléments Filtre :

#	Concept
1	Douleur
2	douleur
3	sans douleur
4	douleur
5	Douleur
6	niveau
7	Hypochondre gauche
8	Hypochondre gauche
9	Exposition aux radiations
10	Dos
11	Régions du dos
12	Dos
13	dos
14	Dos
15	Dos

Affichage de l'élément 1 à 15 sur 15 éléments

Précédent **1** Suivant

Sources : Screenshot_ECMTE_Recherche [782x633, PNG] Illustration réalisée par l'auteur sur <https://cispro.chu-rouen.fr/ECMTE/?>

2.1.9. Licence UTS

Afin de pouvoir utiliser les différents outils cités ci-dessus, une licence UTS est nécessaire. Cette licence permet l'accès à différents dictionnaires médicaux tels que RadLex, SNOMED, etc. Il est donc évident qu'elle est indispensable pour l'utilisation d'outils d'analyse de textes afin de pouvoir faire correspondre différents termes avec des dictionnaires médicaux spécifiques.

L'accès à une telle licence se fait sur demande. Pour ce faire, il faut accéder au site officiel <https://uts.nlm.nih.gov//license.html>, accepter les différentes conditions inhérentes à l'utilisation de ces dictionnaires et remplir un formulaire (cf. annexe I). Il est important d'être extrêmement précis et complet au niveau des informations données dans le formulaire et quant à l'utilisation qui sera faite de cette licence. En effet, la demande est analysée dans les trois jours ouvrables et si leur service juge que des informations sont manquantes, incomplètes ou non-satisfaisantes, il refuse l'accès à la licence en citant les raisons du refus. Il est cependant possible d'effectuer une nouvelle demande en améliorant les points mentionnés. Lorsque la demande est acceptée, l'utilisateur bénéficie alors d'un nom d'utilisateur et d'un mot de passe afin d'utiliser ladite licence. Cette utilisation est gratuite mais comprend cependant une contrainte. Il est en effet nécessaire d'écrire un rapport annuel expliquant ce que l'utilisateur en a fait, et pourquoi elle a été nécessaire. Ce rapport doit se rédiger en anglais.

2.2. Terminologies/Ontologies

2.2.1. Définitions

Bien qu'à la base ils soient fondamentalement opposés, les termes terminologie et ontologie s'utilisent de façon associée de nos jours, particulièrement dans le domaine de l'informatique et de l'ingénierie. Ils sont donc devenus complémentaires. Afin de comprendre leurs différences, voici tout d'abord une définition de la terminologie :

« Plus une pratique qu'une théorie - on peut se poser la question de la considérer comme une pratique théorisée - une terminologie peut être vue comme une langue de spécialité, c'est-à-dire une langue liée à une pratique technique ou scientifique propre à un domaine. Les termes employés sont spécifiques à un métier et renvoient à une réalité extralinguistique partagée : objets (artefacts), pratiques, méthodes, processus, etc. » (Roche, 2005)

Et ci-dessous, une définition de l'ontologie au sens de l'ingénierie des connaissances :

« Définie pour un objectif donné et un domaine particulier, une ontologie est pour l'ingénierie des connaissances une représentation d'une modélisation d'un domaine² partagée par une communauté d'acteurs. Objet informatique défini à l'aide d'un formalisme de représentation, elle se compose principalement d'un ensemble de concepts définis en compréhension, de relations et de propriétés logiques ». (Roche, 2005)

Les différents points détaillés dans les chapitres 2.2.* se basent donc sur un mélange de ces deux termes et peuvent porter l'appellation de « concepts ». Ils constituent une base nécessaire afin de pouvoir utiliser les différents outils et réaliser du mapping de termes. Bien évidemment, tous ont attiré au domaine médical afin de répondre aux besoins du projet.

2.2.2. MeSH

Construit par la NLM en 1954 « MeSH (Medical Subject Headings) est le thésaurus de référence dans le domaine biomédical »¹³. Il est constitué de seize catégories thématiques dont l'Anatomie, les Organismes, les Maladies, etc. Chaque catégorie est répertoriée par une lettre afin de plus facilement les retrouver lors de recherches sur un terme. Pour ne citer que deux exemples, Anatomie est répertoriée par [A] et Organismes par [B]. Toutes les catégories sont organisées en arborescence de descripteurs qui peuvent eux-mêmes se retrouver dans plusieurs catégories, et donc avoir des localisations différentes. « Chaque localisation est représentée par un code

¹³ <http://mesh.inserm.fr/FrenchMesh/presentation.htm> consulté le 01.07.2018

d'arborescence [Tree Number], qui comprend une lettre indiquant la catégorie et une séquence numérique précisant le niveau dans la hiérarchie. » ¹⁴

Le mapping de texte se basant sur du français au sein du projet, il a été choisi de se concentrer sur le MeSH bilingue développé par l'Inserm (Institut national de la santé et de la recherche médicale).

2.2.3. RadLex

Créé par la RSNA (Radiological Society of North America), RadLex est une terminologie médicale principalement axée sur le domaine de la radiologie. Plus spécifique, elle permet aux radiologues de pouvoir pleinement bénéficier d'un langage unifié afin de pouvoir analyser leurs rapports d'imagerie. De plus, RadLex est très complet car il est constitué de plus de 75'000 termes et synonymes en anglais. Des débuts de traduction sont mis en place depuis quelques temps maintenant notamment en allemand et en portugais. ¹⁵ Une traduction française de cette terminologie a été débutée par l'équipe CISMeF de Rouen. De nos jours, selon un échange d'email daté du 24 mai 2018 avec M. Julien Grosjean de CISMeF, 10'260 termes sont déjà traduits par leur équipe et peuvent donc être mappés dans leurs outils.

2.2.4. SNOMED

SNOMED (ou SNOMED CT) « est le produit terminologique de santé clinique le plus complet et le plus précis au monde, détenu et distribué dans le monde entier par SNOMED International. » (ma traduction) ¹⁶ Cette terminologie a donc été développée à large échelle dans le but de pouvoir répondre au mieux aux attentes et besoins des cliniciens. ¹⁶ Complète, elle permet la gestion et compréhension de faits médicaux complexes autant pour la signification professionnelle que pour l'aspect sémantique. ¹⁷ Se destinant à un public international, SNOMED CT a été créée en anglais mais est traduite dans plusieurs langues, dont le français.

2.2.5. UMLS

« L'UMLS, ou Unified Medical Language System, est un ensemble de fichiers et de logiciels qui regroupe de nombreux vocabulaires et normes de santé et biomédicaux pour permettre

¹⁴ <http://mesh.inserm.fr/FrenchMesh/presentation.htm> consulté le 01.07.2018

¹⁵ <https://www.rsna.org/RadLex.aspx> consulté le 01.07.2018

¹⁶ <https://www.snomed.org/snomed-ct> consulté le 01.07.2018

¹⁷ <https://www.e-health-suisse.ch/fr/technique-semantique/interoperabilite-semantique/snomed-ct.html> consulté le 01.07.2018

l'interopérabilité entre les systèmes informatiques. » (ma traduction) ¹⁸ UMLS est partagé en trois outils : métathésaurus, réseau sémantique ainsi que lexique et outils lexicaux. Le premier de ces outils est le plus intéressant car c'est celui qui contient les termes et codes des vocabulaires expliqués ci-dessus : MeSH, RadLex et SNOMED. ¹⁸ Apparaissant régulièrement comme option dans les différents outils expliqués aux points 2.1.*, il paraît donc pertinent de prendre basiquement connaissance d'UMLS afin de mieux comprendre ce qui le compose réellement.

2.3. Technologies

2.3.1. JAVA

« Java est un langage de programmation et une plate-forme informatique qui ont été créés par Sun Microsystems en 1995. » ¹⁹ Java est utilisé pour faire fonctionner de nombreux sites Web, ainsi que de nombreuses applications. Il est notamment devenu extrêmement populaire pour cela. Régulièrement mis à jour, les performances, la stabilité et la sécurité constituent ses points clés. ¹⁹

La partie intéressante de Java qui concerne ce projet est celle du langage de programmation, et plus spécifiquement, celle du la POO (programmation orientée objet). Très largement utilisé dans le monde entier, les sources de code en Java sont immenses et permettent ainsi à son utilisateur de trouver des réponses à la plupart des questions qui peuvent se poser durant un développement. Il est donc un langage de référence.

2.3.2. C# et .NET

Il convient de présenter C# et brièvement Microsoft .NET conjointement car ils sont indissociables. En effet, « C # est un langage orienté objet élégant et sécurisé qui permet aux développeurs de créer une variété d'applications sécurisées et robustes qui s'exécutent sur .NET Framework. » (ma traduction) ²⁰ Il est utile pour créer des services Web XML, des applications client Windows, des applications client-serveur et bien d'autres. C# reste plutôt abordable pour des personnes ayant des connaissances en Java car de nombreuses similarités se retrouvent dans l'écriture de leurs codes et dans le fait que tous deux permettent la POO. Néanmoins, des différences persistent, notamment pour ce qui est des fonctionnalités proposées. En effet, C# propose certaines fonctionnalités non-inclues dans Java dont l'accès direct à la mémoire. De plus, C# nécessite une plateforme .NET afin de pouvoir y développer des programmes. ²⁰

¹⁸ <https://www.nlm.nih.gov/research/umls/quickstart.html> consulté le 01.07.2018

¹⁹ https://www.java.com/fr/download/faq/whatis_java.xml consulté le 10.07.2018

²⁰ <https://docs.microsoft.com/en-us/dotnet/csharp/getting-started/introduction-to-the-csharp-language-and-the-net-framework> consulté le 10.07.2018

2.3.3. Protocoles SOAP et REST

SOAP et REST sont deux protocoles permettant l'accès à un service Web. SOAP étant plus ancien, son bon fonctionnement a été largement prouvé. Or, il reste néanmoins un protocole plutôt complexe qui comporte de nombreuses règles à connaître pour pouvoir le faire fonctionner correctement. Contrairement à REST, « SOAP repose exclusivement sur XML pour fournir des services de messagerie. »²¹

« REST offre une alternative plus simple. Au lieu d'utiliser XML pour faire une demande, REST repose sur une simple URL. »²¹ Il constitue donc un protocole bien plus simple à l'utilisation car sa compréhension est plus basique. L'utilisateur doit simplement entrer une adresse URL afin de pouvoir accéder facilement à un service Web. Afin d'effectuer des tâches, quatre méthodes sont possibles avec HTTP : GET, POST, PUT et DELETE. GET permet de récupérer des données, POST permet d'envoyer des données, PUT les mets à jour et DELETE les supprime. Une différence majeure réside entre REST et SOAP : le format de sortie des réponses. Étant limitées à du XML en SOAP, les réponses REST peuvent être données dans des formats tels que CSV (Comma-Separated Value) ou JSON (Javascript Object Notation). Selon les besoins de l'utilisateur, ce dernier point peut donc influencer son choix entre SOAP et REST.²¹

2.3.4. XML

XML est un langage de stockage et transport de données. Il contient des balises (non-prédéfinies) et une structure afin de pouvoir être interprété. Ces balises doivent être définies par l'utilisateur selon ses besoins spécifiques. Très répandu grâce à la simplicité qui en découle, « XML stocke les données au format texte brut. Cela permet de stocker, de transporter et de partager des données indépendamment du logiciel et du matériel. »²²

²¹ <https://www.supinfo.com/articles/single/2441-comprendre-soap-rest-leurs-differences> consulté le 10.07.2018

²² https://www.w3schools.com/xml/xml_what.asp consulté le 10.07.2018

Afin de mieux comprendre la structure d'une sortie XML, en voici un simple exemple :

Figure 13 : XML - Exemple structure

```
<?xml version="1.0" encoding="UTF-8"?>
<etudiants>
  <etudiant classe="C1">
    <nom>Bonvin</nom>
    <prenoms>
      <prenom>Cécile</prenom>
      <prenom>Marie</prenom>
    </prenoms>
  </etudiant>
  <etudiant classe="C2">
    <nom>Doe</nom>
    <prenoms>
      <prenom>John</prenom>
    </prenoms>
  </etudiant>
</etudiants>
```

Sources : Screenshot_XML_Exemple [385x321, PNG] Illustration réalisée par l'auteur depuis un exemple personnel

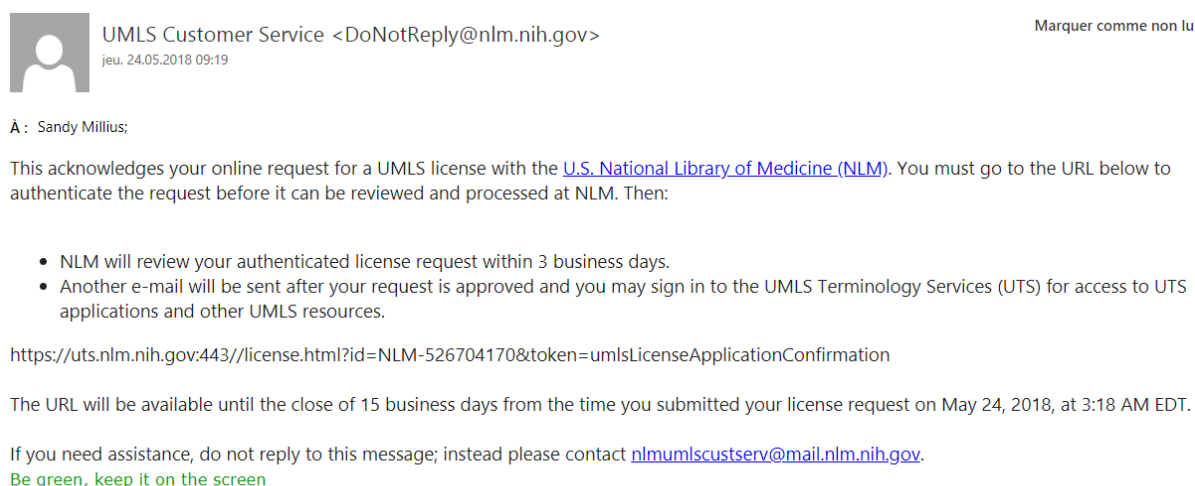
3. Résultats de recherches

A la suite de l'analyse des différents outils existants, de nombreuses recherches et tests ont été effectués afin de voir l'utilisation réelle desdits outils et leur pertinence dans ce projet. Cette phase est essentielle afin que le choix final soit le plus optimal possible. Elle rassemble toutes les étapes réalisées ainsi que les problèmes rencontrés qui ont été déterminants dans cette prise de décision et dans la mise en place d'une structure pour le développement. Tous les tests ont été menés à l'aide de réels rapports anonymisés fournis par le client et choisis au hasard.

3.1. Licence UTS

La demande de création d'un utilisateur pour bénéficier de l'accès à la licence UTS est une étape essentielle car sans elle, la plupart des outils de mapping complets n'auraient pu être utilisés (l'accès aux ontologies étant bien évidemment nécessaires).

Figure 14 : Problème Licence UTS - Email de confirmation de demande

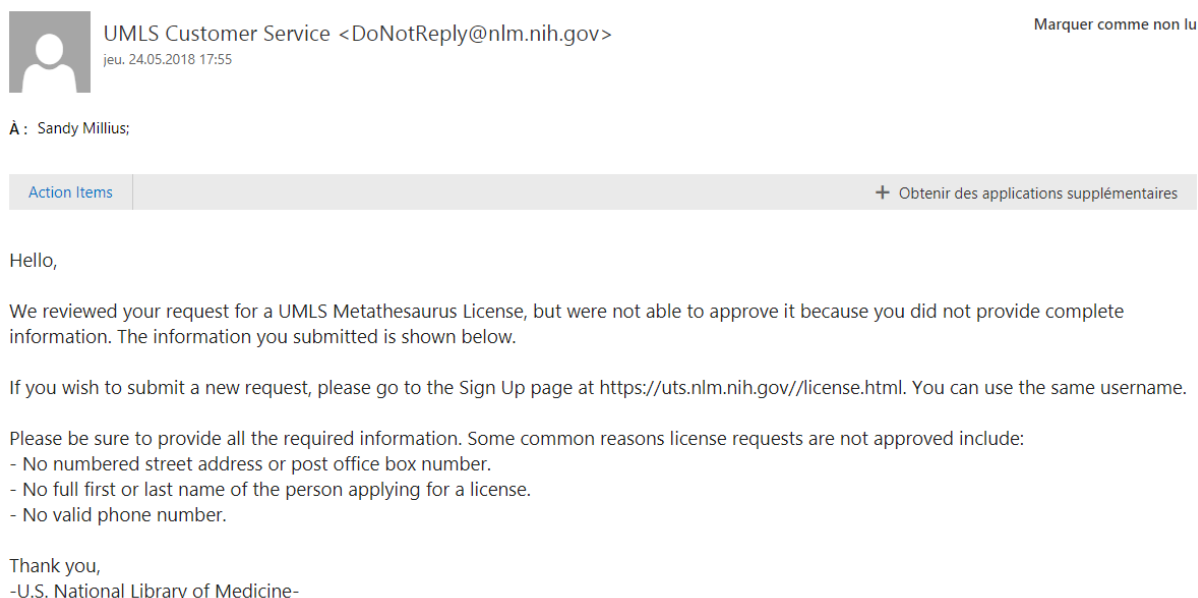


Sources : Screenshot_Probleme_Licence_MailAuto [1220x531, PNG] Illustration réalisée par l'auteur sur sa boîte mail personnelle <https://outlook.hevs.ch/owa/#path=/mail>

Or, il se trouve que le temps de traitement initialement transmis de trois jours, a en réalité pris beaucoup plus de temps et a donc engendré du retard dans la mise en place des tests des outils. De plus, malgré des données personnelles complètes et exactes fournies pour la création d'un utilisateur, la première demande effectuée s'est vue refusée sans une justification qui soit réellement pertinente. Cette dernière était qu'une case postale était selon eux manquante dans l'adresse car il n'y avait aucune indication « P.O. Box ». En réalité, une case postale avait été mentionnée mais en tant que « Case postale ». De plus, il a été mentionné que mon nom était

incomplet et mon numéro de téléphone invalide. Or, il s'avère que mon nom était bien complet et que mon numéro de téléphone était correct et rédigé avec l'indicatif international.

Figure 15 : Problème License UTS - Email de refus



Sources : Screenshot_Probleme_Licence_MailRefus [1187x604, PNG] Illustration réalisée par l'auteur sur sa boîte mail personnelle <https://outlook.hevs.ch/owa/#path=/mail>

Une seconde demande a donc été effectuée en modifiant les données selon leur demande. Or, il se trouve qu'à cette demande, aucune réponse n'a été donnée.

Une solution alternative a donc été trouvée par l'utilisation d'un identifiant disponible et non-utilisé par un assistant au sein de l'Institut de recherche pour la durée de ce travail.

3.2. MetaMap

Cette section regroupe les divers tests effectués et problèmes rencontrés sur les trois versions de MetaMap (MetaMap, MetaMap Java API et MetaMap Lite) détaillées aux chapitres 2.1.1, 2.1.3 et 2.1.4. Les versions de MetaMap et MetaMap Java API étant peu mises à jour et plutôt délaissées, il a été décidé de se concentrer principalement sur leur outil le plus récent : MetaMap Lite.

3.2.1. Aspects négatifs

MetaMap Lite a été conçu afin d'être une version plus légère des outils MetaMap. Bien que ceci aurait pu présenter un avantage certain, il s'avère qu'à trop vouloir alléger cette version, elle se retrouve malheureusement trop incomplète pour les besoins du développement. Voici les différences avec les autres versions de l'outil :

- « Pas de détection d'entités disjointes » (ma traduction) ²³
- « Aucune variante de dérivation » (ma traduction) ²³
- « Pas de désambiguïsation du sens des mots (à ajouter plus tard) » (ma traduction) ²³
- « Pas de surdimensionnement » (ma traduction) ²³
- « Pas de traitement des termes » (ma traduction) ²³
- « Pas de génération de variantes dynamiques » (ma traduction) ²³

Malgré tous ces allègements conséquents, il se trouve que l'outil nécessite malgré tout 16GB d'espace disque et un minimum de 2GB de Ram pour pouvoir le faire fonctionner. ²³ Bien que, de nos jours, ces configurations se retrouvent sommes-toutes plutôt standards, cela représente malgré tout des exigences un peu trop gourmandes pour aussi peu de fonctionnalités.

3.3. cTAKES

cTAKES avait initialement été considéré comme potentiel outil de base pour le développement de ce projet. En effet, grâce à ses nombreuses fonctionnalités, il paraissait être un outil parfaitement complet pour les besoins du client. Or, il s'est avéré qu'après avoir effectué quelques tests sur des rapports anonymisés fournis par le client, il n'est pas aussi adapté que ce qu'il faudrait.

3.3.1. Tests effectués

Tout d'abord, pour effectuer des tests, il a bien évidemment fallu l'installer. Pour ce faire, une installation préalable de Maven est nécessaire. En revanche, l'installation de cTAKES via Maven étant plutôt mal documentée, elle s'est avérée plutôt laborieuse et a nécessité plus de temps que prévu. Une explication plus détaillée des diverses étapes et commandes à effectuer en ligne de commandes de la part des développeurs aurait été grandement utile. Néanmoins, l'outil a finalement pu être installé correctement et est fonctionnel.

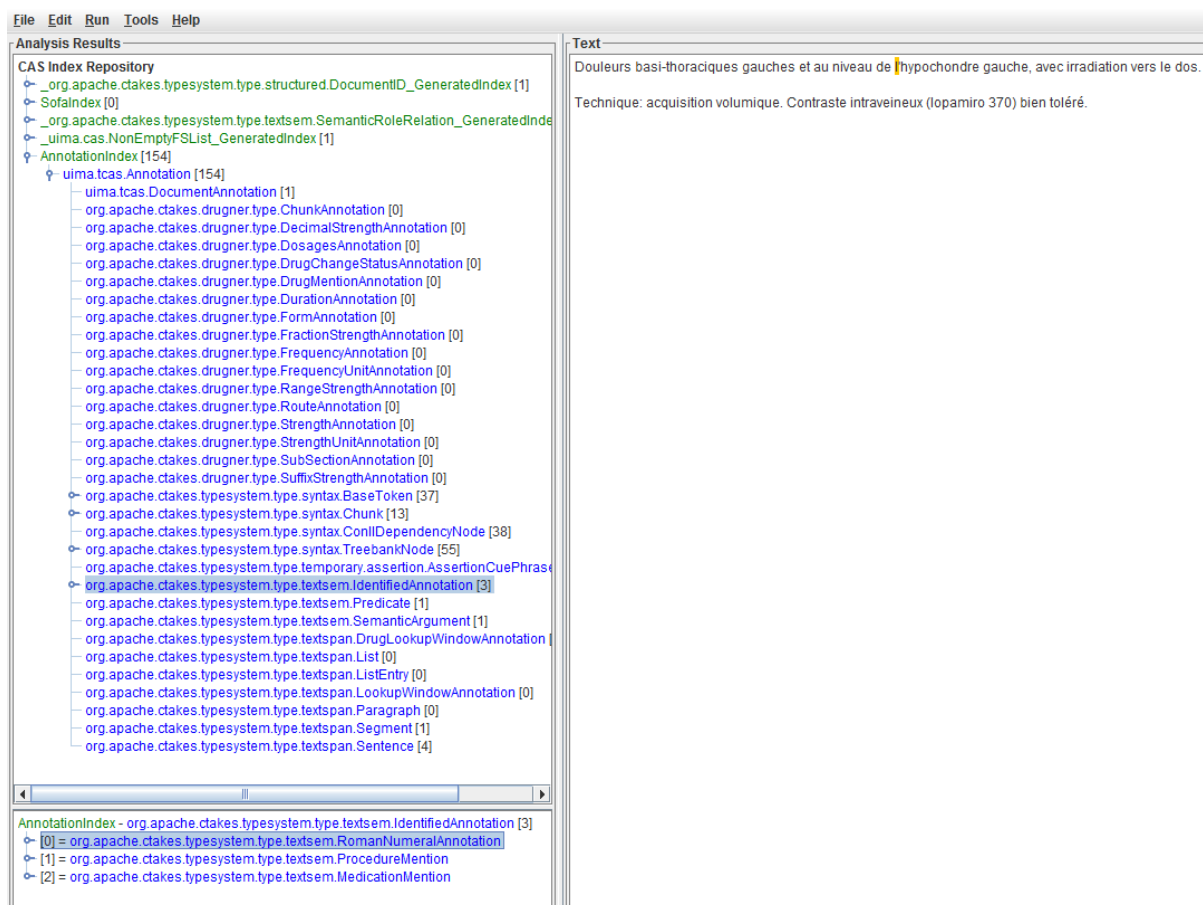
Le lancement de l'outil s'effectue en ligne de commande. Les tests ayant été effectués sur l'outil de matching de base, voici la ligne de commande nécessaire au lancement (à noter qu'il est nécessaire de se trouver dans le dossier contenant cTAKES soit « apache-ctakes-4.0.0 ») :

```
bin\runtakesCVD.bat desc\ctakes-clinical-  
pipeline\desc\analysis_engine\AggregatePlainTextFastUMLSPProcessor.xml
```

²³ https://metamap.nlm.nih.gov/Docs/README_MetaMapLite_3.6.1.html consulté le 01.07.2018

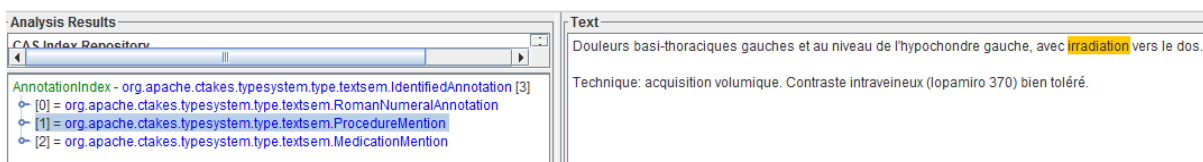
Voici un aperçu de l'interface utilisateur avec en test, l'indication d'un rapport anonymisé en français et les résultats qui en ressortent :

Figure 16 : Résultats de recherche - cTAKES avec indication en français (1)



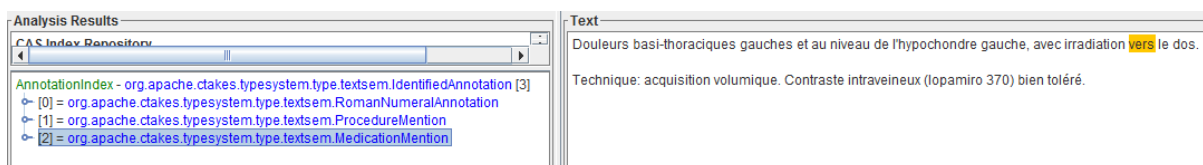
Sources : Screenshot_cTAKES_IndicationFrancais [1085x810, PNG] Illustration réalisée par l'auteur depuis cTAKES

Figure 17 : Résultats de recherche - cTAKES avec indication en français (2)



Sources : Screenshot_cTAKES_IndicationFrancais_mot2 [1081x140, PNG] Illustration réalisée par l'auteur depuis cTAKES

Figure 18 : Résultats de recherche - cTAKES avec indication en français (3)



Sources : Screenshot_cTAKES_IndicationFrancais_mot3 [1085x144, PNG] Illustration réalisée par l'auteur depuis cTAKES

Afin de comprendre au mieux l'efficacité de l'outil, des tests supplémentaires ont été effectués sur une sélection aléatoire de trois rapports qui ont été préalablement traduits en anglais car les résultats en français étaient trop peu concluants. Les résultats de ces tests ont été reportés sur un document (cf. annexes XXI, XXII et XXIII).

3.3.2. Avis/Retours

Selon les tests effectués sur des textes traduits en anglais, cTAKES semble être un bon outil de mapping car les résultats ressortant semblent tout à fait corrects. Son utilisation est plutôt simple car l'interface graphique est extrêmement basique.

Malheureusement, bien que prometteur, cet outil présente de nombreux défauts. Tout d'abord, comme dit précédemment, l'installation obligatoire via Maven s'avère être très complexe et trop peu/mal documentée, ce qui génère une grande perte de temps pour l'installation et l'utilisation de l'outil. Il est donc clair que cTAKES ne s'adresse pas aux personnes n'ayant pas de connaissances avérées en informatique. Ensuite, le lancement de l'outil s'effectue uniquement via des lignes de commande ce qui, à nouveau, nécessite que l'utilisateur bénéficie de certaines connaissances basiques en informatique et en lignes de commande. Finalement, l'interface de cTAKES étant trop basique et non-intuitive, elle complexifie la compréhension de son utilisation et de l'interprétation des résultats.

Bien que rendant l'utilisation de cTAKES fastidieuse, ces trois points ne constituent pas un élément éliminatoire majeur car l'outil est installé par des informaticiens et son utilisation, via un prototype, pourrait être facilitée pour l'utilisateur final. Ce qui rend malheureusement le choix de cet outil impossible est sa mauvaise adaptation à la langue française. Le mapping est quasiment inexistant sur des termes français ou de très mauvaise qualité et non-pertinent (comme on peut le voir dans les figures 16, 17 et 18). Le terme « irradiation » est correctement mappé car il est le même en anglais. Les rapports radiologiques des clients étant uniquement en français, l'ajout d'une étape de traduction serait nécessaire afin de pouvoir recevoir des résultats pertinents de l'outil. Les outils de traduction n'étant pas toujours parfaits, des erreurs pourraient être générées ce qui diminuerait la qualité du résultat du prototype.

3.4. HeTOP et ECMT

Les derniers outils qu'il reste à tester sont HeTOP et ECMT. Il a été décidé de présenter les résultats des tests et recherches effectués conjointement car les deux outils se complètent parfaitement et partagent leurs points forts/faibles. Les résultats ci-dessous sont représentés d'une manière graphique afin d'améliorer la compréhension de ceux-ci, mais l'utilisation finale des outils se fait via leurs versions Web (APIs).

Ces deux outils étant développés de base en français, ils permettent de pallier les problèmes rencontrés par cTAKES et constituent le choix le plus probable.

3.4.1. Tests effectués sur HeTOP

Des tests ont été effectués sur une dizaine de mots de différents types (maladie, anatomie) en français. Les résultats sont très concluants car l'outil retourne de très nombreuses informations (comme expliqué au chapitre 2.1.6). Les possibilités selon les besoins d'utilisations sont donc infinies et il appartient à l'utilisateur de récupérer celles qu'il juge nécessaires.

La possibilité de filtrer les résultats de recherche est également très intéressante car si on effectue un simple « search » pour un terme dans l'API, le fichier retourné sera très complet et donc très lourd. C'est pourquoi il est intéressant de pouvoir récupérer un plus petit fichier XML en sortie avec uniquement les informations nécessaires et utiles.

Des tests ont également été effectués sur HeTOP en entrant des phrases complètes. Malheureusement, cet outil ne gère que la recherche d'un mot, voire d'un petit groupe de mot, mais pas d'une phrase complète (cf. figure 19). C'est pourquoi il devra être complété par un outil de mapping.

Figure 19 : Résultats de recherche - HeTOP avec phrase

The screenshot shows the HeTOP web interface. At the top, there is a search bar with the text 'Douleurs basi-thoraciques gauches et au niveau de l'hypochondre gauche.' To the left of the search bar is a language dropdown menu set to 'français'. Below the search bar, there are two checkboxes: 'Recherche sans troncature (sans wildcard)' which is checked, and 'Sélection terminologies' which is unchecked. Below these checkboxes, there is a button labeled 'Vos recherches'. Below the button, it says '0 ressource(s) trouvée(s) en 0,05 s'. To the right of this, there is a message: 'La recherche ne fournit pas de résultat. Vous pouvez relancer la requête en cochant toutes les terminologies'.

Sources : Screenshot_HeTOP_TestPhrase [1071x275, PNG] Illustration réalisée par l'auteur sur <https://www.hetop.eu/hetop/#q=Douleurs+basi-thoraciques+gauches+et+au+niveau+de+l%27hypochondre+gauche.>

3.4.2. Tests effectués sur ECMT

Les tests ECMT ont été effectués sur dix rapports sélectionnés aléatoirement, afin de pouvoir effectuer une évaluation du mapping. En moyenne, environ 8 mots sont mappés sur l'indication et 11 sur la conclusion. Bien que pour la majorité, les textes étaient de longueur relativement similaire, il convient de souligner que ce chiffre peut varier selon la taille de l'indication et de la conclusion.

Les résultats sont donc très concluants et sont majoritairement corrects et pertinents. Il s'agit donc d'un très bon outil de mapping qui pourrait parfaitement compléter HeTOP. Voici les résultats sortant pour le même texte utilisé sur cTAKES :

Figure 20 : Résultats de recherche - ECMT avec indication complète en français

Extracteur de Concepts Multi-Terminologique (ECMT v3)

[How-to](#) - [Contact](#) - © 2017 CHU de Rouen - CISMef.

Douleurs basi-thoraciques gauches et au niveau de l'hypochondre gauche, avec irradiation vers le dos. Technique acquisition volumique. Contraste intraveineux Iopamiro 370 bien toléré.

Effacer 5 phrases annotées en 2 ms. 35 codes distincts identifiés.

Sources : Screenshot_ECMT_ComparaisonResultats [1075x255, PNG] Illustration réalisée par l'auteur sur

[https://ecmt.chu-rouen.fr/CISMeFecmtservice/REST/getAutomaticIndexingWithOptions/at=false&a=false&d=false&r=false&c=false&sn=false&p=false&e=&f=/Douleurs%20basi-thoraciques%20gauches%20et%20au%20niveau%20de%20l'hypochondre%20gauche%2C%20avec%20irradiation%20vers%20le%20dos.%0A%0ATEchnique%3A%20acquisition%20volumique.%20Contraste%20intraveineux%20\(Iopamiro%20370\)%20bien%20tol%C3%A9%C3%A9.%0A](https://ecmt.chu-rouen.fr/CISMeFecmtservice/REST/getAutomaticIndexingWithOptions/at=false&a=false&d=false&r=false&c=false&sn=false&p=false&e=&f=/Douleurs%20basi-thoraciques%20gauches%20et%20au%20niveau%20de%20l'hypochondre%20gauche%2C%20avec%20irradiation%20vers%20le%20dos.%0A%0ATEchnique%3A%20acquisition%20volumique.%20Contraste%20intraveineux%20(Iopamiro%20370)%20bien%20tol%C3%A9%C3%A9.%0A)

En vert apparaissent les mots reconnus comme étant des termes à mapper par l'outil et que 35 codes distincts leur correspondant ont été trouvés. Les résultats sont donc bien plus grands que sur cTAKES par exemple.

3.4.3. Avis/Retours

D'après les résultats des tests effectués, les avis concernant HeTOP et ECMT sont les mêmes.

Premièrement, les outputs sont très complets et modifiables au gré de nos envies. En effet, il est possible d'affiner sa recherche à l'aide de paramètres ou de méthodes plus précises, ce qui amène à des outputs XML plus adaptés à nos besoins et également plus légers. Les possibilités de ces outils sont vraiment infinies, ce qui donne libre court à tout type de recherches qui trouvent la plupart du temps une réponse.

Ensuite, les appels aux APIs se font en REST à l'aide d'une simple URL. Même si cela peut paraître initialement un peu complexe car il existe beaucoup de méthodes et de paramètres à connaître, l'appel est d'une grande simplicité une fois toutes ces options connues de l'utilisateur. De plus, les appels aux APIs ne nécessitent aucune installation de programmes en local, ce qui rend l'utilisation de ces outils encore plus simple et légère.

Le support apporté par l'équipe de développement de CISMeF est de très bonne qualité. Toute l'équipe est disponible et facilement joignable. Ils prennent réellement le temps d'expliquer le fonctionnement de leurs outils et d'aider les utilisateurs en cas de souci.

Dans le cadre du projet, ni ECMT ni HeTOP ne pourrait être utilisé seuls car ils n'effectuent qu'une partie du travail. C'est pourquoi ils doivent obligatoirement être utilisés ensemble ; leur complémentarité étant une volonté des développeurs.

Globalement, ces outils sont très complets et de très bonne qualité. Ils répondent parfaitement aux besoins du projet.

4. Choix

4.1. Technique

Le choix technique s'est effectué en se basant sur deux critères : les résultats des recherches préalablement développés et les connaissances acquises antérieurement au projet.

Le choix de l'ontologie/terminologie à utiliser s'est porté sur MeSH car elle est très complète et possède une très grande traduction de termes en français. RadLex aurait pu constituer un choix plus judicieux pour ce qui est du domaine de la radiologie, mais malheureusement ses lacunes en traductions françaises ont été rédhibitoires pour ce projet. De plus, le choix n'a pu se porter que sur une seule ontologie car afin de pouvoir effectuer une comparaison, il est nécessaire de remonter dans la hiérarchie et l'arborescence de cette-dernière. Il paraît donc évident que la comparaison est impossible si elle s'effectue entre des arborescences complètement différentes et non-liées.

Java s'est imposé comme le choix de langage le plus logique, car des connaissances préalables facilitaient son utilisation. De plus, il intègre parfaitement l'utilisation d'APIs (nécessaires au projet) et peut facilement être repris pour un développement futur complémentaire si les clients le souhaitent.

Pour ce qui est du choix du protocole, REST est apparu comme étant le meilleur car il est le plus récent et le plus simple d'utilisation. En effet, ses appels s'effectuent dans une simple URL qui laisse transparaître les méthodes utilisées et facilitent donc sa compréhension.

Le choix du XML s'est également imposé de lui-même car c'est dans ce langage que les outputs des outils sélectionnés sont écrits.

En ce qui concerne le choix des outils, il a été décidé de travailler uniquement avec les APIs de ceux-ci, afin de faciliter leur utilisation au travers du code Java. ECMT et HeTOP ont été les choix finaux car il s'agit des outils les plus récents et les plus complets ; ECMT étant sélectionné pour le mapping de texte et HeTOP pour la recherche plus approfondie des termes ressortis dans leurs hiérarchies. Régulièrement mis à jour, un réel désir d'amélioration continue de la part de ses créateurs s'en dégage. Ainsi, il paraît donc évident que les outils continueront de fonctionner pour une longue période et surtout, qu'ils deviendront de plus en plus précis et complets ; ceci rejoint complètement l'esprit de ce projet. De plus, le contact avec les créateurs, l'équipe CISMeF, facilite grandement la compréhension des outils et s'avère être un réel avantage.

4.2. Méthodologie et planning

Le choix de la méthodologie a été libre, car le but étant qu'elle convienne au mieux selon le temps imparti et selon les ressources. Etant donné qu'il s'agit d'un travail réalisé seule, une

méthodologie Agile, comme vu en cours, n'aurait pas pu s'appliquer telle quelle. Il a donc été décidé d'en utiliser une version allégée et simplifiée comme suit.

Un journal de bord a été mis en place et rempli de manière journalière afin d'assurer un suivi de ce qui a été fait, ce qui va être fait et les problèmes rencontrés. D'une certaine manière, ce schéma reprend les principes d'un daily scrum dans la méthodologie agile et permet un suivi régulier. A ces points, ont été ajoutés la date du jour et le nombre d'heures. Ce journal de bord se retrouve aux annexes III, IV, V et VI de ce document.

Un cahier des charges a été complété et signé par les deux parties au début du travail, afin de s'assurer qu'elles s'entendent sur les tâches à réaliser et les buts à atteindre (cf. annexe II).

Afin d'assurer au mieux le suivi et le respect de ce cahier des charges, des séances hebdomadaires ont été planifiées avec le Dr Henning Müller. Ces séances ont pour buts de se tenir informés de l'avancement du travail, ainsi que des éventuels problèmes rencontrés afin de ne pas rester bloquée dans le processus. De plus, des séances si régulières permettent de toujours rester dans la bonne direction, et donc, de ne pas perdre de temps inutilement.

A ces réunions hebdomadaires avec le professeur, s'ajoutent des réunions chaque deux semaines avec les clients afin de les tenir informés de l'avancement du travail. Ces discussions régulières ont permis de soulever des points intéressants au développement et de pouvoir réellement cerner les besoins des clients. Malheureusement, leurs attentes ont eu tendance à rapidement s'accroître au fur et à mesure des réunions dû à leur enthousiasme ; il a donc fallu leur rappeler le temps et les ressources disponibles. Les réunions ont également pu permettre de demander aux clients, les documents nécessaires aux tests des outils.

Ces divers rendez-vous ont été répertoriés dans un agenda qui se trouve sur le même fichier Excel que le journal de bord et le cahier des charges (cf. annexe VII). Le but de cet agenda est d'avoir un suivi des diverses réunions avec le professeur et avec les clients, ainsi que les principaux points abordés à ces derniers.

Finalement, afin d'avoir un aperçu plus complet des discussions découlant des diverses séances, des procès-verbaux allégés sous forme de documents Word courts et concis ont été rédigés à la fin de chaque séance (cf. annexes VIII, IX, X, XI, XII, XIII et XIV). Ainsi, il est en tout temps possible d'avoir un suivi détaillé des points abordés et des décisions prises pendant le processus. Ces documents ont servi de points de repères en cas de doutes.

5. Développement

5.1. Particularités techniques et architecture

Le développement s'effectuant en Java, la plateforme de développement Eclipse a été choisie car elle est simple et connue. La version utilisée est la dernière en date : Oxygen. Il a été nécessaire d'ajouter des bibliothèques au projet Eclipse afin de pouvoir utiliser certaines fonctionnalités. En voici la liste :

- XML beans récupérable sur :

<http://www.java2s.com/Code/Jar/x/Downloadxmlbeansjar.htm>

- Commons collection récupérable sur :

https://commons.apache.org/proper/commons-collections/download_collections.cgi

- Woodstox-core récupérable sur :

<http://www.java2s.com/Code/Jar/w/Downloadwoodstoxcoreasl420jar.htm>

- POI récupérable sur :

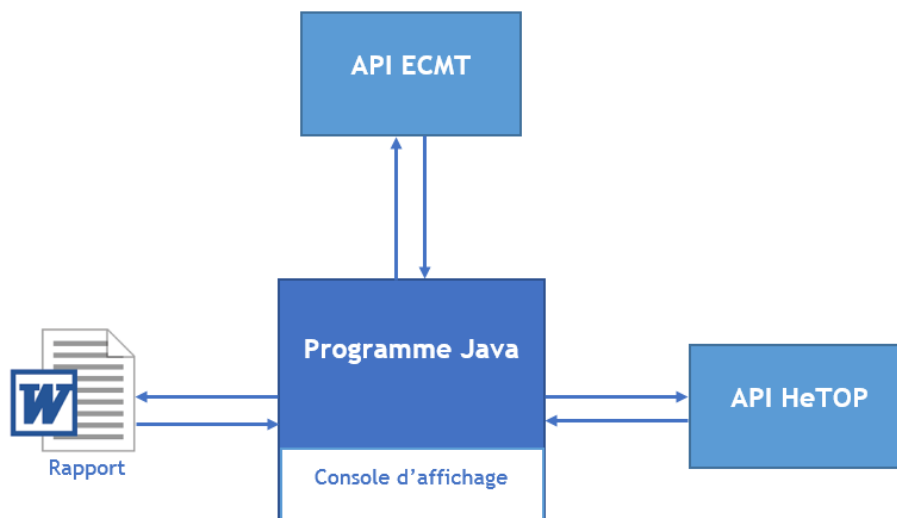
<https://www.apache.org/dyn/closer.lua/poi/release/bin/poi-bin-3.17-20170915.tar.gz>

- JDOM récupérable sur :

<http://www.jdom.org/downloads/index.html>

Afin de mieux comprendre la structure du prototype, voici son architecture :

Figure 21 : Développement - Architecture

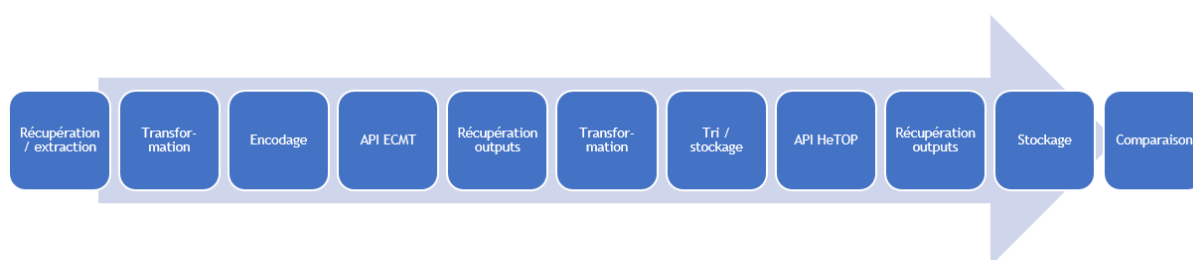


Sources : Architecture [835x459, PNG] Illustration réalisée par l'auteur depuis Word, image doc-file-icon-57c766ef3df78c71b65394d4 [640x640, PNG] récupérée sur <https://mon-set-up-gaming.fr/transformer-pdf-word-doc/>

5.2. Phases de développement

Le développement et le code du prototype peuvent être découpés en plusieurs phases. Ainsi, il est plus simple de comprendre son fonctionnement. Voici un schéma qui illustre ces phases :

Figure 22 - Développement - Phases



Sources : Processus [1219x282, PNG] Illustration réalisée par l'auteur depuis Word

Des exemples de code complètent les explications des différentes phases.

5.2.1. Récupération de fichier et extraction de texte

La première phase de développement commence bien évidemment par la récupération du fichier en entrée. Pour ce faire, une simple ligne suffit (la première de l'illustration ci-dessous). Il est important de noter que le type peut varier en fonction du genre de fichier (.doc ou .docx). Les

clients disposent de rapports en format .doc, c'est pourquoi « HWPFDocument » est utilisé. S'il s'agissait de .docx, « XWPFDocument » devrait être utilisé à la place.

Le « WordExtractor » permet d'extraire le contenu du fichier .doc donné dans le « FileInputStream ». Finalement, un tableau est créé afin de stocker les différents paragraphes qui vont être récupérés depuis le programme. En effet, comme le but est de comparer l'indication et la conclusion, il sera nécessaire de pouvoir les stocker.

Figure 23 : Développement - Récupération de fichier

```
//HWPFD s'utilise pour .doc, pour un .docx utiliser XWPF
HWPFDocument doc = new HWPFDocument(new FileInputStream("D:/TB/Test.doc"));
WordExtractor we = new WordExtractor(doc);

String[] paragraphs = we.getParagraphText();
```

Sources : Dev_1_Extracteur [765x115, PNG] Illustration réalisée par l'auteur depuis Eclipse Oxygen

Il est maintenant nécessaire d'extraire les paragraphes correspondant à l'indication et à la conclusion. Les rapports des clients étant toujours construits de la même façon (cf. annexes XIX et XX), le programme récupère ce qui se trouve entre le terme « Indications » et le terme « Description » pour ce qui est de l'indication. Pour la conclusion, il s'agit des paragraphes entre « Conclusion » et « Avec nos ». Les paragraphes récupérés sont stockés dans deux tableaux distincts : « indications » et « conclusions », afin de pouvoir les utiliser par la suite.

Figure 24 : Développement - Extraction de l'indication et de la conclusion

```
//Boucle pour récupérer tous les paragraphes du document
for(int i = 0; i<paragraphs.length;i++) {
    if(paragraphs[i].contains("Indications")) {
        //System.out.print(paragraphs[i]);
        i++;
        i++;
        while(paragraphs[i].contains("Description")==false) {
            String string1 = paragraphs[i];
            indications.add(string1);
            i++;
        }
        System.out.println(indications.toString());
    }
    if(paragraphs[i].contains("Conclusion")) {
        //System.out.print(paragraphs[i]);
        i++;
        i++;
        while(paragraphs[i].contains("Avec nos")==false) {
            String string2 = paragraphs[i];
            conclusions.add(string2);
            i++;
        }
        System.out.println(conclusions.toString());
    }
}
```

Sources : Dev _2_Récupération [630x511, PNG] Illustration réalisée par l'auteur depuis Eclipse Oxygen

5.2.2. Transformation de texte compatible pour URL

Une fois les paragraphes d'indication et de conclusion récupérés, il est nécessaire de les transformer en String afin de pouvoir les manipuler plus facilement. Cette transformation s'effectue dans les deux premières lignes ci-dessous grâce à la méthode « toString() ».

Afin de pouvoir effectuer une recherche dans l'API de ECMT, il est nécessaire de passer le texte à rechercher dans l'adresse URL. Pour ce faire, nous allons utiliser nos Strings « indicationString » et « conclusionString ». Or, des problèmes peuvent se poser pour la recherche via une URL si ces Strings contiennent des espaces ou des crochets, car ces caractères ne sont pas reconnus. Il se trouve que malheureusement, dans tous les rapports des clients, des crochets et des espaces s'y retrouvent. Par conséquent, il a été nécessaire de remplacer les espaces par des « %20 » et de supprimer les crochets qui n'apportaient aucune valeur aux paragraphes. Ces remplacements doivent s'effectuer pour l'indication et la conclusion à l'aide de la méthode « replaceAll() » existant dans la librairie Java.

Figure 25 : Développement - Transformation de String pour URL

```
indicationString = indications.toString();
conclusionString = conclusions.toString();

indicationsURLOk = indicationString.replaceAll(" ", "%20");
indicationsURLOk = indicationsURLOk.replaceAll("[\\r\\n]+", "");

conclusionsURLOk = conclusionString.replaceAll(" ", "%20");
conclusionsURLOk = conclusionsURLOk.replaceAll("[\\r\\n]+", "");
```

Sources : Dev_3_StringURL [640x174, PNG] Illustration réalisée par l'auteur depuis Eclipse Oxygen

5.2.3. Encodage d'autorisations

Afin de pouvoir accéder aux APIs de HeTOP et de ECMT, un nom d'utilisateur ainsi qu'un mot de passe sont nécessaires. Un compte a été créé dans le cadre de ce projet par l'équipe de CISMeF. Ces deux éléments doivent être transmis au moment de la connexion à l'URL et pour ce faire, il est nécessaire que les données soient encodées. La méthode « encode » permet d'encoder rapidement les accès introduits dans le format : « username:password ». Un affichage de l'encodage, mais également du décodage, ont été créés afin de vérifier l'exactitude des données et le bon fonctionnement de la méthode.

D'une fois que les accès sont encodés, il est nécessaire de les passer dans un String avec en début « Basic » puis notre « encodedBytes » afin que les APIs puissent comprendre la requête puisqu'il s'agit d'un protocole REST.

Figure 26 : Développement - Encodage autorisations

```
/*
 * Encodage et autorisation pour APIs
 */

//Encodage du username + password pour authentication
byte[] encodedBytes = Base64.getEncoder().encode("hevs:5ui98ae".getBytes());
System.out.println("encodedBytes " + new String(encodedBytes));
byte[] decodedBytes = Base64.getDecoder().decode(encodedBytes);
System.out.println("decodedBytes " + new String(decodedBytes));

//Création champ authorization
String auth = "Basic " + new String(encodedBytes);
System.out.println(auth);
```

Sources : Dev_4_Encodage [783x279, PNG] Illustration réalisée par l'auteur depuis Eclipse Oxygen

5.2.4. Passage dans API ECMT

L'appel à l'API s'effectue au moyen de l'URL à laquelle est ajoutée l'indication transformée dans un format correct pour URL. La connexion est ensuite ouverte et l'autorisation passée afin pouvoir se connecter à l'API.

Figure 27 : Développement - Connexion à l'API ECMT

```
/*
 * CALL API ECMT
 */

//Partie Indication -> Connexion à l'API
String urlecmt = "https://ecmt.chu-rouen.fr/CISMeFecmts-service/REST/getAutomaticIndexingWithOptions/"
    + "at=true&a=false&d=false&r=true&c=true&sn=true&e=&f=/" + indicationsURLOk;

URL urlIndic = new URL(urlecmt);
URLConnection urlConnIndic = urlIndic.openConnection();

urlConnIndic.setRequestProperty("Authorization", auth);
urlConnIndic.setDoOutput(true);
urlConnIndic.setAllowUserInteraction(false);
```

Sources : Dev_5_CallECMTIndication_Connexion [1015x292, PNG] Illustration réalisée par l'auteur depuis Eclipse Oxygen

Ce processus est exactement identique pour la conclusion. Seules les variables sont différentes afin de récupérer les bonnes valeurs.

5.2.5. Récupération des résultats depuis ECMT

Il faut désormais récupérer les résultats en sortie de l'API ECMT. Ces résultats sont repris grâce à un « `BufferedReader` ». Le programme va lire tous les résultats et les stocker dans un tableau de `Strings`. Une boucle « `while` » permet au programme de tourner et de remplir ce tableau, tant qu'il trouve des résultats ; il ne s'arrête que lorsqu'il a lu tout le fichier. Des « `System.out.println()` » sont utilisés afin de permettre l'affichage dans la console d'Eclipse.

Figure 28 : Développement - Récupération des outputs ECMT

```
//Récupération des outputs XML au format d'un seul String
BufferedReader brIndic = new BufferedReader(new InputStreamReader(urlConnIndic.getInputStream()));
String outputEcmtInd = null;
ArrayList<String> outputEcmtIndArray = new ArrayList<String>();
while ((outputEcmtInd=brIndic.readLine())!=null) {
    outputEcmtIndArray.add(outputEcmtInd);
    System.out.println("Output indication ECMT : "+outputEcmtInd);
    System.out.println();
}
```

Sources : Dev_6_CallECMTIndication_RecuperationOutput [997x198, PNG] Illustration réalisée par l'auteur depuis Eclipse Oxygen

Ce processus s'applique également à la conclusion avec les variables adaptées.

5.2.6. Transformation du résultat String en XML

Comme le résultat est récupéré en String et qu'il est nécessaire d'avoir un fichier en format XML afin de pouvoir naviguer dans les éléments, il faut procéder tout d'abord à une transformation du résultat. Pour ce faire, la solution apportée a été d'enregistrer le résultat dans un fichier XML vide préalablement créé.

Pour que le document XML soit valide, il faut en premier lieu, faire un nettoyage des crochets dans le fichier grâce aux méthodes « replace() ». En effet, comme le résultat était stocké dans un tableau, il a automatiquement hérité de crochets en début et fin de document. Et ce sont ceux-ci précisément qui posent problèmes.

Lorsque le nouveau résultat est stocké dans la variable « outputEcmtIndDoc », cette dernière est ajoutée dans un fichier XML vide, ici « TestXMLIndication.xml ». Ainsi, il sera possible de récupérer la valeur de résultat dans un réel format XML.

Figure 29 : Développement - Transformation du résultat String en XML

```
/*
 * Transformation de String en XML
 */

//Transformation en String moins les [ ] qui posent problème pour la transformation en Document
String outputEcmtIndDoc = outputEcmtIndArray.toString();
outputEcmtIndDoc = outputEcmtIndDoc.replace("[", "").replace("]", "");
System.out.println("Test : "+outputEcmtIndDoc);

//Enregistrement du String dans un fichier .xml
File txtFile = new File("D:/TB/TestXMLIndication.xml");
if (!txtFile.exists())
    throw new FileNotFoundException("Le fichier n'existe pas");
else{
    PrintStream l_out = new PrintStream(new FileOutputStream("D:/TB/TestXMLIndication.xml", true));

    l_out.print(outputEcmtIndDoc);

    l_out.flush();
    l_out.close();
    l_out=null;
}
```

Sources : Dev_7_TransformationStringToXML [1007x457, PNG] Illustration réalisée par l'auteur depuis Eclipse Oxygen

Cette étape est la même pour la conclusion ; seuls les variables et le fichier de sortie diffèrent.

5.2.7. Tri et stockage des résultats propres de ECMT

Maintenant que le résultat est en String, il va être possible de retrouver les bons éléments. Ce qui est nécessaire de récupérer dans la construction du XML de sortie sont l'attribut « matchterms » et l'attribut « idcismef » afin de pouvoir récupérer le nom et l'identifiant du terme dans une terminologie précise (dans notre cas MeSH).

Ci-dessous, des « ArrayList » sont créées afin de pouvoir stocker le résultat retourné par la méthode. L'appel à la méthode effectuant tout le travail est « getArrayOutputs() », avec en paramètre de méthode, le document XML à parcourir. Les deux dernières lignes de code de la figure 30 permettent d'ajouter les « matchterms » et les « idcismef » qui auront été préalablement stockés dans nos « ArrayList ».

Figure 30 : Développement - Tri et stockage des résultats finaux ECMT

```
/*
 * Récupération des matchterms et stockage dans array
 */
ArrayList<ArrayList<String>> arrayDarrayIndic;
ArrayList<String> arrayOutputsMatchtermsEcmtIndic;
ArrayList<String> arrayOutputIdcismefEcmtIndic;
arrayDarrayIndic = getArrayOutputs("D:/TB/TestXMLIndication.xml");
arrayOutputIdcismefEcmtIndic = arrayDarrayIndic.get(0);
arrayOutputsMatchtermsEcmtIndic = arrayDarrayIndic.get(1);
```

Sources : Dev_8_TriStockageECMT1 [681x196, PNG] Illustration réalisée par l'auteur depuis Eclipse Oxygen

Dans un souci de facilité de compréhension, la méthode ci-dessous a été séparée en deux parties. La première montre principalement la construction du document nécessaire pour la récupération de l'élément racine de notre XML. Le « namespace » doit être déclaré car il est utilisé à l'intérieur des éléments. La racine constitue le tout premier élément de notre document XML. C'est de là que va découler l'ensemble de la recherche qui suit.

De plus, toutes les variables, « ArrayList » et éléments nécessaires sont déclarés afin de pouvoir être utilisés dans la deuxième partie de la méthode.

Figure 31 : Développement - Méthode getArrayOutputs() pour tri des résultats ECMT (1)

```
public static ArrayList<ArrayList<String>> getArrayOutputs(String fichier){
    SAXBuilder sxb = new SAXBuilder();

    try {
        document=sxb.build(new File(fichier));
        System.out.println("Doc : "+document);
    }
    catch(Exception e) { }

    Namespace ns = Namespace.getNamespace("http://ws.ecmt.cismef.org/");
    System.out.println("namespace : "+ns.getURI());

    racine = document.getRootElement();

    String [] tabElements = {"sentence","indexations", "indexation"};
    String attributeMatchterm;
    String attributeIdcismef;
    ArrayList<String> arrayOutputsMatchtermsEcmt = new ArrayList<String>();
    ArrayList<String> arrayOutputsIdcismefEcmt = new ArrayList<String>();
    ArrayList<ArrayList<String>> arrayOutputs = new ArrayList<>();

    Element elementBase = racine;
    Element elementOK;
    Element elementTest;
```

Sources : Dev_8_TriStockageECMT2 [813x491, PNG] Illustration réalisée par l'auteur depuis Eclipse Oxygen

La deuxième partie de la méthode est la plus intéressante car c'est ici que la recherche et la récupération des attributs vont être effectives. Pour ce faire, notre programme va passer dans une boucle « for » qui va tourner sur les éléments nécessaires à la récupération des attributs voulus. A savoir que toutes les structures de documents sont les mêmes et que, par conséquent, les éléments sont toujours identiques. Dans un premier temps, nous allons récupérer l'enfant du premier élément se trouvant en dessous de la racine. A l'aide du « if », une comparaison est effectuée afin de savoir si l'élément est bien égal à « indexations » (nécessaire dans notre cas pour pouvoir aller plus loin). Si c'est le cas, les enfants de l'élément « indexations » vont être récupérés et ajoutés à une liste d'éléments. Une boucle « for » permet ensuite de tourner dans cette liste et de récupérer directement les attributs « matchterms » et « idcismef » voulus. Afin d'affiner la qualité du résultat, un nouveau « if » avec deux conditions est ajouté. La première permet d'éviter les doublons dans le tableau de résultat. La deuxième condition, quant à elle, filtre les éléments afin de ne prendre que ceux ayant l'« idcismef » commençant par « MSHD_D_ ». Cet ID correspond aux descripteurs MeSH. Il est important de préciser que ce sont les descripteurs qui sont à récupérer et non les concepts car ces derniers ne contiennent pas de hiérarchie et ne nous serviront donc pas plus tard.

Une fois les bonnes valeurs enregistrées dans notre tableau, celui-ci nous est retourné.

Figure 32 : Développement - Méthode getArrayOutputs() pour tri des résultats ECMT (2)

```
for(int k = 0; k<tabElements.length;k++) {
    elementOK = elementBase.getChild(tabElements[k], ns);
    System.out.println("For ElementOK : "+elementOK);
    if(tabElements[k]=="indexations") {
        List<Element> indexationList = elementOK.getChildren();
        System.out.println("Mes children de indexations : "+elementOK.getChildren());
        for(int l=0;l<indexationList.size();l++) {
            elementTest = indexationList.get(l);
            System.out.println("Attribute matchterms : "+elementTest.getAttributeValue("matchterms"));
            System.out.println("Attribute idcismef : "+elementTest.getAttributeValue("idcismef"));
            attributeMatchterm = elementTest.getAttributeValue("matchterms");
            attributeIdcismef = elementTest.getAttributeValue("idcismef");
            if(!arrayOutputsMatchtermsEcmt.contains(attributeMatchterm) && attributeIdcismef.startsWith("MSHD_D_")) {
                arrayOutputsMatchtermsEcmt.add(attributeMatchterm);
                arrayOutputsIdcismefEcmt.add(attributeIdcismef);
            }
        }
        elementBase = elementOK;
    }
    System.out.println(arrayOutputsMatchtermsEcmt.toString());
    System.out.println(arrayOutputsIdcismefEcmt.toString());

    arrayOutputs.add(arrayOutputsIdcismefEcmt);
    arrayOutputs.add(arrayOutputsMatchtermsEcmt);
    return arrayOutputs;
}
```

Sources : Dev_8_TriStockageECMT3 [1248x542, PNG] Illustration réalisée par l'auteur depuis Eclipse Oxygen

Toutes les étapes indiquées dans ce sous-chapitre sont, comme toujours, semblables pour la conclusion.

5.2.8. Passage des IDs dans API HeTOP

Lorsque les IDs sont correctement récupérés et stockés dans le tableau « arrayOutputIdcismefEcmtconc », il faut les faire passer dans l'API de HeTOP pour pouvoir remonter dans la hiérarchie de l'output XML. Pour ce faire, le tableau récupéré est passé en paramètre de la méthode « getPath() ». Cette méthode nécessite une authentification pour établir une connexion avec l'API. L'ArrayList « paths » stocke les résultats de la méthode « getPath() ». Ceux-ci représentent les différents chemins récupérés pour permettre une comparaison en remontant dans l'arborescence de l'ontologie MeSH.

Figure 33 : Développement - Passage IDs dans API HeTOP

```
ArrayList<ArrayList<String>> paths = new ArrayList<ArrayList<String>>();

for (int q=0; q<arrayOutputIdcismefEcmtConc.size();q++) {
    paths.add(getPath(arrayOutputIdcismefEcmtConc.get(0).toString(), auth));
}
```

Sources : Dev_9_PassageIDHeTOP [782x117, PNG] Illustration réalisée par l'auteur depuis Eclipse Oxygen

La connexion à l'API HeTOP se fait à l'aide de l'URL de base de recherche de hiérarchie dans laquelle est ajoutée l'ID se trouvant en paramètre de la méthode « getPath() ».

Figure 34 : Développement - Connexion à l'API HeTOP

```
public static ArrayList<String> getPath(String id, String auth)throws IOException{
    ArrayList<String> path = new ArrayList<String>();

    String urlPath = "https://www.hetop.eu/CISMeFhetopservice/REST/getConceptHierarchiesAscendants/"+id+"/fr";
    URL url = new URL(urlPath);
    URLConnection urlConn = url.openConnection();

    urlConn.setRequestProperty("Authorization", auth);
    urlConn.setDoOutput(true);
    urlConn.setAllowUserInteraction(false);
```

Sources : Dev_9_ConnexionHeTOP [1117x212, PNG] Illustration réalisée par l'auteur depuis Eclipse Oxygen

5.2.9. Récupération des résultats XML de HeTOP

La première partie de récupération des résultats à l'aide d'un « BufferedReader » s'exécute de la même façon que pour celle de ECMT au chapitre 5.2.5 et n'est donc pas développé à nouveau ici ; de même pour le nettoyage du String sans les crochets du chapitre 5.2.6.

Une spécificité s'impose néanmoins dans le cas de la création du fichier XML. En effet, il est fort probable que, dans la majorité des cas, la première API retourne plusieurs résultats, et donc plusieurs IDs. C'est pourquoi, le nom du fichier XML créé doit différer et s'adapter en fonction du nom de l'ID.

Figure 35 : Développement - Enregistrement du fichier XML HeTOP

```
//Enregistrement du String dans un fichier .xml
File file = new File("D:/TB/"+id+".xml");
file.createNewFile();

PrintStream l_out = new PrintStream(new FileOutputStream("D:/TB/"+id+".xml", true));

l_out.print(outputDoc);

l_out.flush();
l_out.close();
l_out=null;
```

Sources : Dev_10_EnregistrementFichier [851x232, PNG] Illustration réalisée par l'auteur depuis Eclipse Oxygen

Le principe de la récupération du « path » s'applique sur la même base que pour la récupération des « matchterms » et « idcismef » du chapitre 5.2.7. Seules les valeurs des éléments recherchés dans le « tabElements » diffèrent et le fait que les valeurs finales à récupérer soient stockées dans un élément, et non dans un attribut.

Figure 36 : Développement - Récupération de la hiérarchie dans outputs XML de l'API HeTOP

```
Namespace ns = Namespace.getNamespace("http://ws.cismef.org/");
System.out.println("namespace : "+ns.getURI());

racine = document.getRootElement();

String [] tabElements = {"dbo", "hies", "hie", "path"};

Element elementBase = racine;
Element elementOK;
Element elementTest;
Element elementPath;

for(int n = 0; n<tabElements.length;n++) {
    elementOK = elementBase.getChild(tabElements[n], ns);
    System.out.println("elementOK"+elementOK.toString());
    if(tabElements[n] == "hies") {
        List<Element> hieList = elementOK.getChildren();
        System.out.println("Mes children de hies : "+elementOK.getChildren());
        for(int o=0;o<hieList.size();o++) {
            elementTest = hieList.get(o);
            elementOK = elementTest.getChild(elementTest.toString(), ns);
            if(hieList.get(o).toString() == "hie"){
                elementPath = elementOK.getChild(hieList.get(o).toString(), ns);
                if(elementPath.toString() == "path") {
                    path.add(elementPath.toString());
                }
            }
        }
    }
}
```

Sources : Dev_10_RecuperationHierarchie [814x603, PNG] Illustration réalisée par l'auteur depuis Eclipse Oxygen

5.2.10. Stockage des résultats propres de HeTOP

Le stockage des résultats nécessaires de HeTOP se fait dans l'ArrayList « paths » comme montré à la figure 31. L'affichage de son contenu sur la console s'effectue par une simple boucle for.

Figure 37 : Développement - Affichage des paths en résultats de HeTOP

```
for(int p=0; p<paths.size();p++) {  
    System.out.println();  
    System.out.println("Paths : "+paths.get(p).toString());  
}
```

Sources : Dev_10_RecuperationHierarchie [814x603, PNG] Illustration réalisée par l'auteur depuis Eclipse Oxygen

5.2.11. Comparaison

Afin de pouvoir effectuer une comparaison, il a été décidé d'utiliser les chemins de l'arborescence ressortant dans les fichiers XML pour un terme recherché, comme expliqué ci-dessus. En effet, ces chemins, ou « paths », ont une certaine structure commune. Si l'on recherche par exemple les chemins pour l'ID « MSH_D_005490 », voici les différents résultats :

MSH_D_ARBO/MSH_D_C/MSH_D_001423/MSH_D_007239

MSH_D_ARBO/MSH_D_C/MSH_D_001423

MSH_D_ARBO/MSH_D_C

MSH_D_ARBO/MSH_D_C/MSH_D_001423/MSH_D_007239/MSH_D_005490

Ces quatre chemins constituent la base pour notre recherche. En effet, l'idée étant de stocker ces quatre « paths » dans une ArrayList. Celle-ci contiendrait, pour chaque « path », une autre ArrayList de « paths » découpés. En prenant l'exemple du premier « path » retourné, son chemin découpé contiendrait :

MSH_D_ARBO

MSH_D_C

MSH_D_001423

MSH_D_007239

Cette nouvelle liste de morceaux de « path » seraient donc comparés avec une autre, afin de pouvoir calculer le nombre d'index communs. Dans l'exemple ci-dessus, il est simple de se rendre compte visuellement que les trois premiers morceaux sont communs : MSH_D_ARBO, MSH_D_C et

MSH_D_001423. Les morceaux de « path » qui ne sont pas communs, ici MSH_D_007239, représentent la distance de ces termes dans l'arborescence MeSH. En fonction de cette distance, il est possible de déterminer si deux termes peuvent être jugés équivalents ou non. Dans notre exemple, comme il n'y a qu'un seul morceau de chemin supplémentaire, il est simple de déduire que la similitude des termes est dans ce cas correcte. L'écart des termes dans les branches est donc le moyen de comparaison choisi. Cette comparaison de « paths » doit donc s'effectuer pour chaque terme retourné par l'API ECMT.

6. Conclusion

6.1. Bilan technique

Actuellement (en l'état), le prototype ne permet malheureusement pas une comparaison correcte car une exception est renvoyée par Eclipse. Par manque de temps, celle-ci n'a pas pu être analysée et identifiée correctement. Cette exception se produisant dans la phase de tri et de récupération des outputs HeTOP, elle empêche d'effectuer de réels tests sur la suite du développement. N'étant pas en mesure de certifier la justesse du code de comparaison, il a été décidé de n'expliquer que son principe qui semble avoir convaincu le professeur et les clients. C'est pour cela que le point 5.2.11 ne comporte pas de code.

Cependant, il convient de souligner que toute la première partie du code qui concerne l'API ECMT et ses résultats est parfaitement fonctionnelle et approuvée par les clients. Grâce à leurs connaissances métiers, les « matchterms » retournés par l'API leur permettent de se faire directement une idée du type de rapport dont il peut s'agir.

6.2. Problèmes rencontrés

Durant les trois mois de réalisation de ce travail, plusieurs problèmes ont été rencontrés impactant plus ou moins le projet. Tout d'abord, un problème fréquent dans la réalisation de travaux avec des clients réels est le changement d'avis quant au type de prototype voulu. En effet, au fur et à mesure des réunions avec les clients, des idées et envies nouvelles apparaissaient et il a fallu mettre certaines priorités quant aux contraintes de temps et de moyens à disposition afin de ne pas se retrouver avec un projet trop ambitieux. L'impact de ce problème a été moindre car grâce à la très bonne communication avec les clients et à l'appui du professeur, la masse de travail est restée correcte.

Ensuite, la gestion des outputs retournés par les différentes APIs était plutôt complexe. Effectivement, cela peut prendre du temps de comprendre la structure du document XML afin de savoir quel élément ou attribut est nécessaire à notre projet ; d'autant plus que certaines méthodes REST retournent d'énormes fichiers. D'une fois que ces outputs ont été compris, la gestion de ceux-ci et leur formatage au niveau du développement a également été plutôt compliquée. Avec du temps et des explications de la part de l'équipe de CISMeF, tous ces outputs ont finalement été maîtrisés.

Puis, un problème de gestion de négation est apparu lorsque des tests ont été effectués sur des rapports contenant des mots de négation. Lors d'un échange avec l'équipe CISMeF au moment du choix des outils, il était apparu qu'une gestion des négations étaient possible. Or, après avoir effectué des tests, il se trouve que la négation n'est reconnue que pour un petit nombre de termes

très spécifiques ; les outils ne trouvent donc pas toujours un négatif. Ensuite, le réel problème s'est posé au niveau de la récupération des outputs XML où la négation n'a pas pu être gérée, et donc stockée. Cette option n'a donc pas pu être intégrée dans ce prototype.

Finalement, le problème majeur de ce projet a été le manque de temps dû aux nombreuses difficultés rencontrées en cours de réalisation. En effet, des problèmes tels que les refus de licence ou encore le temps passé sur des tests d'outils s'avérant finalement trop incomplets, ont constitué une réelle perte de temps. Cette perte a bien évidemment eu un effet direct sur la qualité et les finitions du prototype. Comme dit précédemment, à ce jour, le prototype contient une erreur qui a été bloquante dans la réalisation de la comparaison. Le concept choisi pour cette comparaison a d'ailleurs été très complexe à trouver, car il est dur de savoir si les résultats retournés seraient réellement précis et pertinents. En effet, s'agissant du domaine médical, les possibilités et liens entre certains termes sont extrêmement vastes. C'est pourquoi, des connaissances métiers seraient nécessaires pour l'évaluation de la qualité des résultats. C'est ce dernier point qui a constitué le dernier problème rencontré.

6.3. Respect du cahier des charges

Le cahier des charges initial, joint en annexe II, a été partiellement respecté. En effet, cela s'explique notamment aux nombreux changements de direction intervenus durant le projet. Les points remplis sont les suivants : 1, 2, 3, 4, 6, 8 et 11. Le point 8 concernant la gestion de la littéralité est géré automatiquement par l'outil. Pour ce qui est du point 11 qui concerne l'intégration dans le workflow du client, il est tout à fait possible étant donné que le prototype est un programme Java. Une explication plus poussée de ce dernier peut d'ailleurs être faite en août si les clients le souhaitent.

Les points 5 et 10 ont partiellement été remplis. Comme expliqué précédemment, la négation n'a pas pu être gérée complètement à cause de l'outil. Le point 10 quant à lui concernait l'implémentation d'un prototype de comparaison qui n'est pas totalement fonctionnel à ce jour.

Le point 9 concernant l'interface d'accès à la base de données des clients, de base obligatoire, a été jugée comme finalement non-nécessaire car des rapports anonymisés ont été remis par les clients afin d'éviter la nécessité d'un accès à leur base de données ; le secret médical étant ainsi préservé.

La gestion de la traduction pour l'utilisation de RadLex au point 7 n'étant pas obligatoire, il a finalement été décidé de se concentrer sur MeSH plutôt que sur une traduction de RadLex. Or, ce point a néanmoins été évoqué durant les réunions et une solution de traduction en collaboration avec les clients et l'équipe du CHU-Rouen est discutée.

6.4. Améliorations futures et recommandations

Le prototype réalisé étant très basique, de nombreuses améliorations pourraient être apportées afin de le rendre meilleur et plus complet.

Une comparaison fonctionnelle constitue le point d'amélioration principal et le plus urgent. La source de l'erreur retournée devrait être trouvée et le problème corrigé afin que l'idée de comparaison puisse réellement être testée et, au besoin, améliorée. Afin d'affiner les résultats, une amélioration continue pourrait être apportée à l'aide de professionnels disposant de connaissances médicales et qui pourraient donc juger de la qualité des résultats.

La qualité du code pourrait être améliorée en utilisant plus de classes et de méthodes séparées. Ceci rendrait le code plus facilement réutilisable et éviterait les doublons qui sont certes fonctionnels, mais peu attractifs en termes d'optimisation du code. Cette amélioration passerait notamment par un choix de dossiers de destination flexible.

L'ajout d'une interface graphique pourrait être intéressant afin de faciliter l'utilisation du programme pour un public non-initié. Cette interface pourrait, par exemple, contenir un bouton permettant de chercher le rapport et une zone d'affichage des résultats de comparaison. Une alternative plus simple serait d'afficher les résultats dans un fichier .txt. Ainsi, les utilisateurs pourraient avoir une sorte de « compte rendu » des résultats facile à la compréhension.

Une adaptation de l'outil pour une utilisation sur des rapports de différentes langues avec, de ce fait, un lien vers des ontologies adaptées à ces langues s'avérerait très utile, surtout en Suisse où les rapports en allemand, en italien et en anglais sont très nombreux.

L'automatisation du processus via la création d'un script permettrait d'effectuer des comparaisons sur un grand nombre de rapports en très peu de temps. Ceci pourrait être très intéressant car si l'on prend l'exemple de nos clients, le nombre de rapports entrés journalièrement dépasse les 400. Il est donc évident que si tous les rapports entrés sont à tester, une automatisation s'avère obligatoire pour ne pas les passer un à un dans le programme. Sur cette idée, il serait donc également envisageable de rajouter un paramètre pour l'envoi automatique de tous les rapports journaliers à une certaine heure de la journée ; par exemple, tous les soirs à 20h00.

Finalement, une dernière amélioration à apporter, qui serait obligatoire afin d'avoir une comparaison parfaite, est la gestion de la négation.

6.5. Conclusion personnelle

La réalisation de ce travail de Bachelor a été très enrichissante. En effet, le choix d'un thème ayant attrait au domaine médical a été un vrai challenge personnel car il s'agit d'un domaine

complexe pour quelqu'un qui n'a pas de réelles connaissances techniques médicales. Consciente du temps supplémentaire nécessaires à l'acquisition des bases de ces connaissances, j'ai tout de même pris le risque de me lancer dans cette aventure car le projet me tient à cœur.

En effet, je trouve essentiel de pouvoir utiliser l'informatique afin de résoudre une réelle problématique ; c'est d'ailleurs pour cela que j'ai choisi d'en faire mon métier. En discutant avec mon professeur et avec les clients, je me suis rendu compte du réel besoin d'un outil d'analyse de qualité des rapports ; qu'ils soient radiologiques ou d'un autre domaine médical. Effectivement, étant dans le domaine de la santé plus particulièrement, la mauvaise qualité d'une conclusion ou, pire, une conclusion totalement erronée peut avoir de graves conséquences et représenter des pertes financières. C'est pourquoi, portée par l'enthousiasme et la passion émanant des personnes m'ayant entourées dans ce projet, j'ai fait de mon mieux pour apporter un début de solution future possible afin d'éviter que ce problème ne se produise. Il est très gratifiant de savoir que ce projet a de fortes chances d'être poursuivi et que son but sera peut-être un jour atteint.

En ce qui concerne les outils utilisés, HeTOP et ECMT, je trouve que ce sont de très bons outils et, suite aux échanges avec l'équipe de CISMeF, une réelle volonté d'amélioration de leurs outils s'en dégage. Cela me porte à croire que leurs outils n'en deviendront que meilleurs et plus complets avec le temps, et qu'ils pourront donc parfaitement s'adapter à l'amélioration de ce projet.

Finalement, ce travail de Bachelor m'aura également permis de mettre en pratique les connaissances Java basiques acquises durant ma formation et d'apprendre à travailler de manière plus autonome. Le contact avec de vrais clients permet également une transition vers le monde professionnel.

Malgré toutes les difficultés rencontrées durant ces trois mois, j'ai eu plaisir à pouvoir donner le meilleur de moi-même dans ce projet. J'ai espoir qu'il ait pu démontrer la réelle nécessité d'un tel programme et qu'une suite sera apportée.

RÉFÉRENCES

Roche, C. (2005, Janvier). Terminologie et ontologie. *Langages*, pp. 48-62.

Rodier, M., & Zuber, J. (2013). *Guide de présentation et de réalisation des travaux écrits*. HES-SO Valais-Wallis.

Savova, G., & Zheng, J. (2011, Mars 11). Natural Language Processing and Information Extraction for Biomedicine.

Declerck, T., & Lendvai, P. (consulté le 20.07.2018). *Extraction de concepts et relations sémantiques à partir de labels d'ontologies*. Récupéré sur <http://ceur-ws.org/Vol-580/paper3.pdf>

Annexe I : Formulaire de demande de licence UTS

*Username

[See username/password requirements](#)

*Password

*Password Confirmation

*Complete First Name (no initials)

Middle Name or Initial

*Complete Last Name (no initials)

*E-mail Address

*Confirm E-mail Address

[SPAM Filtering software notice](#)

Organization

Required if the address below is for organization, office or company.

☐ I am requesting this license on behalf of the above named organization.

Department

Department or office name. Required for university addresses.

*Postal Address (Line 1)

Street address with building, office/suite/apartment number, if applicable.

Postal Address (Line 2)

Province and locality can be included here.

*City

*Country

Select One ▼

State/Province

Select One ▼

Required for US and Canada.

*Postal/Zip Code

Enter "None" if your country does not use postal/zip codes.

*Phone Number

Non-US applicants please include country code.

Phone Number Extension

*1. What is your affiliation? (Select one)

Select One ▼

*2. What is the primary activity of your institution/organization? (Select one)

Select One ▼

*3. What is your role in your institution/organization? (Check all that apply)

☐ Administrator/Manager
☐ Educator
☐ Health care provider
☐ Researcher
☐ Software developer
☐ Other

4. ☐ Check this box if a third-party application or system requires you to have a UMLS license. Select from the list below.

Select vendor:

Select One ▼

5. ☐ Check this box if you plan on using the UMLS in an application. If so, what kind? (Check all that apply)

Select application types:

☐ Clinical - Clinical Decision Support
☐ Clinical - Clinical Research Applications
☐ Clinical - Electronic Health Records
☐ Clinical - Laboratory Information System
☐ Clinical - Personal Health Record
☐ Clinical - Pharmacy System
☐ Non-Clinical - Billing
☐ Non-Clinical - Bioinformatics
☐ Non-Clinical - Linking to Literature/Information Retrieval
☐ Non-Clinical - Research
☐ Other

* ☐ I confirm that I read and agreed to the terms of the UMLS License Agreement and its Appendices. I will provide an annual report, and my name, affiliation, address, and phone number are true and complete. I understand my application may be rejected without further notice if I provided insufficient or false information.

Sources : Screenshot_UTS_License_CompleteForm [523x885, PNG] Illustration réalisée par l'auteur sur <https://uts.nlm.nih.gov//license.html>

Annexe II : Cahier des charges

Etapes

1. Analyse de l'état de l'art
2. Prise de décision sur les outils et langage
3. Analyse de l'extraction de texte interne depuis des rapports et nettoyage nécessaire
4. Nettoyage du texte non-utile (stop word removable = mots plus fréquents, termes qui n'apportent pas d'informations)
5. Gestion de la négation du texte
6. Mapping entre du texte libre (.rtf) avec une certaine terminologie et évaluer les résultats
7. Gestion de traduction pour utilisation de RadLex
8. Gestion de la latéralité de la zone du corps depuis le titre
9. Création d'une interface simple afin d'accéder à la BDD du client
10. Implémentation d'un outil d'analyse de correspondance entre la question et la conclusion d'un rapport radiologique (prototype)
11. Inclusion de l'outil dans le workflow du client

Importance

- Obligatoire
- Obligatoire
- Obligatoire
- Obligatoire
- Obligatoire
- Obligatoire
- Facultatif
- Obligatoire
- Obligatoire
- Obligatoire
- Facultatif

14.05.18



Sources : CahierCharges1 [1056x727, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel
OrganisationTB_SandyMillius.xlsx imprimé et signé en date du 14.05.2018

Annexe III : Journal de bord (1)

Date	Nombre heures	Ce que j'ai fait	Ce que je vais faire	Problèmes rencontrés	Heures /s
07.05.2018	8h	Rdv client 07h45 Sion Rdv TB 11h00 Technopôle Mise en place du cahier des charges détaillé selon indications client, structure de l'écrit, début recherches diverses selon discussion avec le client	Recherches et rédaction	Aucun	
08.05.2018	4h	Début de rédaction et recherches sur les outils et langages Choix méthodologie à adopter et éventuelles adaptations nécessaires	Amélioration du cahier des charges en PB pour signature le 14.05.2018	Recherches bibliographiques trop larges	
11.05.2018	4h	Amélioration du cahier des charges complet en PB adapté	Rédaction de l'avant-propos et de l'introduction	Aucun	
12.05.2018	4h	Rédaction de l'avant-propos et de l'introduction	Suite des recherches afin de finaliser mon choix et rédaction de celui-ci	Aucun	
14.05.2018	3h	Rdv TB	Rdv TB	Aucun	20
17.05.2018	3h	Recherche Metamap, recherche solutions alternatives (Migrep, cTakes)	Lecture de l'exemple de TB donné par le professeur	Aucun	
18.05.2018	4h	Recherches diverses solutions, réception d'une alternative à tester avec l'API Metamap sans installation complète	Suite des recherches afin de trouver une solution et pouvoir proposer un début de solution au client pour mardi 24.05	Problèmes de compatibilité avec Windows 8.1	
20.05.2018	2h	Installation du dossier de l'API ok, tentatives de la faire fonctionner et de trouver les causes du problème	Tentative d'installation de API Metamap et tester la compatibilité, et si ok, s'habituer aux fonctionnalités	Problèmes de compatibilité avec Windows 8.1	
21.05.2018	7h	Tentative de faire fonctionner l'API mais il me faut un accès UTS (en attente car beaucoup de contraintes US et prix)	Tentative de faire fonctionner l'API en modifiant des classes exemples	Je n'arrive pas à faire fonctionner l'API	12
22.05.2018	3h	Début d'installation de cTakes, recherches	Trouver une alternative (installation de cTakes) afin de présenter quelque chose au client demain et si réponse du professeur avec une licence, continuer avec Metamap	Licence UTS nécessaire, en attente	
23.05.2018	3h	Rdv client, recherches Metamap Lite et licence	Rdv client, recherches sur retour email pour Metamap Lite et licence UTS	Licence UTS nécessaire, en attente	
24.05.2018	2h	Rédaction Metamap, souscription licence	Rédaction recherches, souscrire licence UTS	Aucun	
25.05.2018	4h	Rdv TB, réception de la réponse de UTS	Rdv TB, suite rédaction	Aucun	
28.05.2018	7h	Renvoi d'une demande de licence plus complète, rédaction de Licence UTS, Metamap Java API	Résoudre mon souci de licence refusée, analyser les différentes solutions reçues par email et voir quel choix est judicieux, rédaction des précédentes recherches (Licence UTS, Metamap Java API, Metamap Lite, cTakes)	Licence UTS refusée, donc pas de dev tant que pas ok	19
29.05.2018	2h	Analyse solutions proposées, recherches, rédaction Metamap Lite	Analyser solutions email, recherches HeTop, rédaction Metamap Lite	Pas de nouvelles pour la licence, donc pas de dev	
30.05.2018	3h	Rédaction cTakes	Rédaction cTakes, méthodologie, langage programmation	Pas de nouvelles pour la licence, donc pas de dev	
03.06.2018	3h	Rdv TB, recherches solutions alternatives	Rdv TB, rédaction méthodologie, langage programmation	Pas de nouvelles pour la licence, donc pas de dev	
04.06.2018	1h	Préparation pour rdv client avec critères, lecture des rapports donnés et analyse de ces derniers (négation, littéralité, etc), recherches	Préparation pour rdv client avec critères, lecture des rapports donnés et analyse de ces derniers (négation, littéralité, etc), tests	Pas de nouvelles pour la licence, donc pas de dev	15
			Rdv client	Pas de nouvelles pour la licence, donc pas de dev	

Sources : JournalBord1 [1349x815, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel
OrganisationTB_SandyMillius.xlsx

Annexe IV : Journal de bord (2)

06.06.2018	1h	Rdv client	Se renseigner pour la licence, contacter Ivan Eggel pour prendre rdv pour tester sa licence ou Metamap, retranscrire retours clients sur ordi	Pas de nouvelles pour la licence, donc pas de dev	
08.06.2018	4h	Contact de Ivan Eggel qui me prête sa licence en attendant, retranscription du retour des clients	Installer Maven et Metamap Lite	Documentation complexe et peu précise	6
11.06.2018	7h	Maven installé et Metamap Lite installé	Tester Metamap Lite avec les 10 rapports que j'ai déjà, installer cTAKES et le tester également pour comparer les deux outils		
12.06.2018	4h	Installation de cTAKES version développeur et utilisateur, recherches sur fonctionnement de cTAKES, recherche pour résolution de problèmes d'installation	Continuer recherches utilisation Metamap Lite, et utilisation cTAKES. Tester l'outil cTAKES avec des parties de rapports pour voir l'output et l'utilisation possible future.	Pas compréhension du fonctionnement de Metamap Lite, problème cTAKES dev	
13.06.2018	6h	Test de cTAKES utilisateurs, essai des possibilités	Recherches QuickUMLS, chercher comment faire fonctionner Metamap Lite, faire une liste des éventuels outils selon les divers besoins de fonctionnalités attendues dans cahier des charges		
17.06.2018	2h	Liste des éventuels outils, recherches QuickUMLS	Tests cTAKES avec nouveau rapports pour rdv 18.06.2018		19
18.06.2018	1h	Tests cTAKES avec nouveau rapports pour rdv client	Rdv client		1
25.06.2018	8h	Rdv client avec debriefing avancement	Envoi email pour prise de Rdv Müller, envoi d'email à Federica suite à la demande du client, prise en main/installation/test des outils envoyés par Rouen (HeTop, ECMT) pour voir s'ils satisfont aux besoin	Outils testés pas suffisamment complets, complexité du TB et du domaine médical	
26.06.2018	7h	Email Müller ok (rdv 27.07 à 9h), email Federica envoyé, prise en main du ECMT HeTop avec test sur phrases de rapports	Faire ma liste de points à aborder lors du rdv, recherches, rédaction HeTop + ECMT + chapitre des ontologies (Mesh, RadLex, SNOMED, UMLS)		
27.06.2018	8h	Liste points à aborder, recherches	Rdv avec Müller, fixer prochain rdv, mise en place d'un nouveau file rouge si besoin selon discussion, rédaction début partie tests et recherches déjà effectués	Fixer des objectifs à court terme afin d'être dans les temps et priorisation	
28.06.2018	7h	Rdv avec Müller, prochain rdv fixé, mise en place des objectifs pour mercredi prochain, maj du file rouge rédactionnel, rédaction des résultats de recherches (Intro, Licence et début Metamap)	Rédaction suite Metamap + cTAKES + HeTop		
29.06.2018	6h	Suite rédaction	Rédaction de la partie choix méthodologie		36
02.07.2018	8h	Rédaction des choix méthodologiques	Partie dev : Tester manuellement mes 10 rapports dans HeTop et voir résultats, Me renseigner sur API HeTop, Coder début du programme Java avec extraction du texte et appel à l'API		
03.07.2018	13h	Partie dev : Test manuel des 10 rapports (résultats dans doc word), début du code java pour extraction de texte	Suite dev extraction d'un morceau précis du doc + call à l'API HeTop	Format des documents étaient en .doc de Windows 97-2003	

Sources : JournalBord2 [1347x771, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel

OrganisationTB_SandyMillius.xlsx

Annexe V : Journal de bord (3)

04.07.2018	10h	Rdv Müller pour debriefing, suite dev : call API + récupération des données, téléchargement MeSH et analyser construction du dictionnaire, dev : début code pour comparaison des outputs et résultats	Rdv Müller pour debriefing, suite dev : call API + récupération des données, téléchargement MeSH et analyser construction du dictionnaire, dev : début code pour comparaison des outputs et résultats	
06.07.2018	10h	Rdv Müller (continuer tel quel), suite dev call API, Envoi email au chu-rouen pour demande infos car bloquée	Dev : fin call API + test, récupération données et structuration si besoin, Rédaction : fin de rédaction des choix	
08.07.2018	6h	Dev : fin call API Hetop + test	Dev : traitement de données "exemples" avec HeTOP (boudes) -> ajout à l'URL, enregistrement des outputs XML	ECMT qui est essentiel pour le début de mon traitement comporte un souci de dev (nouvelle version disponible la semaine prochaine (date non-définie))
09.07.2018	10h	Dev : traitement HeTOP avec boucle, ajout à l'URL	Dev: Gestion des outputs XML + Début méthodes de comparaison, Rdv client 14h00 pour debriefing, Téléchargement MeSH	47
10.07.2018	10h	Rdv client pour debriefing, dev : recherches gestion outputs XML	Ecriture retours client, Suite dev : Gestion outputs XML, tests ECMT	
11.07.2018	12h	Ecriture retour clients, Suite dev	Dev : résoudre problème affichage, implémentation ECMT, gestion des XML, début de la recherche dans l'arborescence pour comparaison	
12.07.2018	12h	Ecriture résultats dans un txt, connexion à ECMT réussie, implémentation et transformations pour l'indication et la conclusion dans ECMT	Gestion outputs XML (téléchargement de JDOM, tests), code pour enregistrement des bons outputs ECMT dans un tableau afin d'utiliser au mieux HeTOP, recherches sur écriture MeSH pour compréhension de l'output XML, gestion des outputs de HeTOP et écriture de ces derniers dans un csv comme demandé par les clients	
13.07.2018	12h	Compréhension des outputs ECMT et HeTOP (recherches pour check arborescence), début gestion outputs XML pour ECMT, téléchargement de JDOM	Rédaction : Planification des différentes parties manquantes selon avancement, Ajouter images (compléter écrits), check des citations de paraphrases corrects, rédaction de HeTOP + API, ECMT, ECMT + début ontologies (définition + MeSH)	
14.07.2018	12h	Rédaction : planification ok, images ajoutées, début rédaction	Suite rédaction	
15.07.2018	12h	Rédaction	Rédaction	80
16.07.2018	12h	Rédaction	Rédaction + Suite Dev : gestion outputs XML	
17.07.2018	12h	Rédaction, Envoi du doc au professeur, Dev : gestion des outputs XML (récupération des bonnes données)	Dev : Suite gestion des outputs afin de réussir à récupérer du XML et de pouvoir rechercher ce qui est nécessaire dans le document	Problèmes de conversion de données avec le XML, compréhension des différents éléments difficile avec le document ECMT
18.07.2018	12h	Dev: récupération du XML en output, enregistrement dans un fichier .xml afin de le stocker, recherche nœuds enfants afin de retourner les matchterms	Dev : Fin de récupération des outputs pour la partie ECMT avec stockage des termes (pas à double!) dans un tableau de string, recherche dans hetop avec ce tableau de string et gestion des outputs de sortie afin de récupérer le bon nœud pour la comparaison	Difficulté de compréhension de la construction XML des documents en outputs

Sources : JournalBord3 [1348x790, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel
OrganisationTB_SandyMillius.xlsx

Annexe VI : Journal de bord (4)

19.07.2018	10h	Dev : Récupération des outputs XML de ECMT terminée et fonctionnelle, gestion des termes à double, stockage dans un tableau, début recherche méthode de comparaison pour HeTOP avec étude du XML en output. Envoi d'un e-mail à l'équipe de dev de HeTOP afin de mieux comprendre le document XML en sortie et de pouvoir appliquer une méthode de comparaison	Rdv client, rédaction	Non-compréhension du XML en sortie de HeTOP, attente sur une réponse de l'équipe de Rouen pour savoir comment gérer la comparaison	
20.07.2018	12h	Rdv client, rédaction	Rédaction		
21.07.2018	12h	Rédaction	Rédaction		
22.07.2018	12h	Rédaction	Rédaction		92
23.07.2018	9h	Rédaction	Rédaction + Dev : trouver une solution pour comparaison	Problème de hiérarchie des termes pour comparaison	
24.07.2018	8h	Rédaction + Dev : début de solution pour hiérarchie trouvée, à voir comment faire la comparaison encore	Rédaction + chercher imprimerie pour TB		
25.07.2018	8h	Rédaction + imprimerie trouvée	Envoi email Henning + Dev (essayé d'aller au plus loin possible, 100% dev)	Complexité des hiérarchies et d'un moyen de comparaison	
26.07.2018	8h	Envoi email + Dev	Rdv client, rédaction	Dev non-fonctionnel, manque de temps	
27.07.2018	16h	Rdv client, rédaction pv, rédaction	Rédaction, rdv Henning pour derniers détails		
28.07.2018	10h	Rédaction, rdv Henning	Fin de rédaction, envoi du dossier terminé pour correction et review, application des retours de corrections, réalisation du poster		
29.07.2018	3h	Rédaction terminée, poster terminé	Graver le cd, dernier check que tout soit rempli, dépôt sur la zone O., envoi sur Urkund		
30.07.2018	2h	Cd gravé, dépôt sur la zone O., Urkund ok	Impression et remise officielle du TB au secrétariat		64

411

Heures totales

Sources : JournalBord4 [1352x586, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel
OrganisationTB_SandyMillius.xlsx

Annexe VII : Journal des rendez-vous

Rdvs Henning Müller	Rdvs clients	Points à aborder avec Henning Müller	Points à aborder avec les clients
30.04.2018 9h00		Informations TB générales	Rencontre client, souhaits et attentes, explication du contexte d'un TB et des limites que celui-ci peut avoir
07.05.2018 11h00	07.05.2018 07h45	Recherches, cahier des charges	Version Windows/Linux, Metamap ou cTAKES?, Licence UMLS utilisable?, VMs?
14.05.2018 11h00	22.05.2018 07h45	Peu de sources livre, MetaMap, Ordre de réalisation, priorités	
24.05.2018 11h00		MetaMap Lite, Licence UTS demandée, critères client, VMs?, Acronyme TB	Critères spécificité (-) / sensibilité (+), rapports, voir si la question doit matcher avec la réponse ou si la réponse doit matcher avec la question, voir si le degré de précision joue un rôle
30.05.2018 13h00	04.06.2018 07h45	Retours Dev, avancement	Tests effectués avec cTAKES sur un panel de rapports reçus
	18.06.2018 07h45		
27.06.2018 9h00		Organisation TB selon retours client et avancement outils, analyse de la faisabilité du projet, mise en place d'échéances strictes afin d'avancer au maximum	
04.07.2018 9h00	09.07.2018 14h00	Avancement rédaction, présentation de la réalisation partie dev	Problème d'anonymisation des rapports reçus, avancement du TB (démon dev)
12.07.2018 14h00	19.07.2018 8h00	Avancement général	Présentation de l'application (développement devrait théoriquement être terminé ou quasiment)
27.07.2018 13h00	26.07.2018 07h30	Derniers points pour TB, corrections et avis	Présentation finale de l'avancement

Sources : JournalRdvs [1750x505, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel

OrganisationTB_SandyMillius.xlsx

Annexe VIII : PV réunion client du 07.05.2018

Date	07.05.2018
Lieu	Institut de radiologie de Sion
Participants	M. Cyril Thouly, M. Dr Henning Müller, Mme Sandy Millius

Points abordés	
N°	Détails
1	Serveur interne -> je m'y connecterai via un lien sécurisé (VPN)
2	Windows ou Linux à choix
3	Pas de restriction de langage/outil
4	Uniquement pour une utilisation backend
5	Informations à chercher directement sur BDD (Oracle) -> documents en format rtf avec déjà les 4 zones définies (donc pas de séparation à effectuer mais voir nettoyage des balises) -> Voir pour automatiser depuis la BDD -> Ils effectuent une sorte de copie de leur BDD et c'est plus là-dessus que je vais travailler
6	SNOMED déjà dans leur système mais non-utilisé (gratuit en Suisse) -> dans leur BDD ils ont Cim10 (parti de UMLS donc devrait être ok avec Metamap) + ACR
7	Code ACR également déjà présent (par zone anatomique ou autre, à approfondir)
8	Possibilité de rechercher un outil de traduction pour utiliser RadLex directement plutôt que MeSh
9	Tout en français chez eux (sauf rares cas à ne pas prendre en compte, ou peut-être éventuelles futures améliorations)
10	Un des buts essentiels est de voir si on a assez d'infos dans le rapport et qu'on réponde bien à la question (=analyse de qualité des rapports en vue d'une éventuelle future formation interne pour leurs radiologues)
11	Gestion de la négation faisable? -> Outil NLPTK (python)
12	Pour question/conclusion, voir pour travailler sur la sémantique et syntaxe des phrases pour éliminer mots courants inutiles (/!\ ne pas enlever négation)
13	Pour le reste du texte, également supprimer les mots qui ne sont pas utiles dans la comparaison -> outils existants
14	Tests se feront directement dans leurs bureaux et sur leur BDD (tester sur des mois passés par exemple)
15	Amélioration future souhaitée -> Ils aimeraient voir s'il y a possibilité de rendre le système intelligent et qu'à chaque erreur, il s'améliore -> But de tendre vers l'IA
16	Remplir mon cahier des charges avec des tâches obligatoires et facultatives, mettre accent sur priorités
17	Pour info, ils ont environ 400 à 500 rapports entrés dans leur BDD par jour
18	On ne touche plus à la problématique de la facturation qui sera une amélioration future possible
19	Disponibilités -> le client est indisponible qu'une semaine en juillet alors voir si possible de commencer le dev plus tôt et se concentrer sur l'écrit plus tard mais avancer le plus rapidement possible

Sources : PVReunionClient090718 [1063x606, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel
OrganisationTB_SandyMillius.xlsx

Annexe IX : PV réunion client du 22.05.2018

Date	22.05.2018
Lieu	Institut de radiologie de Sion
Participants	M. Dr Hugues Brat, M. Cyril Thouly, M. Dr Henning Müller, Mme Sandy Millius

Points abordés	
N°	Détails
1	Clients nous fourniront environ 200 rapports pour vérifier la fiabilité des outils que je teste (100 ok, 50 avec négation, 50 mauvais)
2	Ce qui est important pour eux maintenant, c'est que je définisse mes critères (avec quoi on est satisfaits, un pourcentage attendu ou certains mots à rechercher ou autres)
3	Voir si les dictionnaires traduits sont assez complets sans RadLex.
4	Domaines principaux : Ostéo-articulaire et Cénologie (pathologies mammaires)
5	Voir comment gérer si indications sont un peu floues mais que la conclusion est très complète afin que le système comprenne que c'est bon et répondu. Il existe des cas particuliers (du type Bilan qui comprend pleins de choses) mais nous on se focuse uniquement sur les questions bien posées et complètes avec une réponse adéquate. Il faut que ça trouve qu'il y ait une bonne réponse
6	On doit leur donner un lien vers dico contextuel qu'on aurait "trié" pour qu'ils voient ce qui pourrait les intéresser
7	Henning envoie un email à une connaissance pour voir s'il existe d'autres outils sans licence, pour un éventuel plan B
8	Ils vont voir s'ils ont un dico à eux (Nuance peut-être) que j'importerais
9	Ils ont du Windows 7 et quelques Windows 10, mais possibilité de faire sur Linux mais en VMs du coup
10	Tous leurs serveurs sont virtualisés (Linux)
11	Négation pour eux doit être claire (positif ou négatif, pas de nuance)
12	Il faut que ça gère la littéralité (ou on le fait nous-mêmes pour matcher) -> outil NLP simple pour tester ça si besoin
13	Leurs rapports dans BDD ne sont pas anonymes (il faudra d'ailleurs que je signe une chart) voir pour peut-être faire l'extraction en 2 phases : 1) extraire les docs et les anonymiser 2) les traiter dans l'outil
14	On doit leur donner un acronyme pour mon TB (un nom) -> ACCORD (Automatic Classification of 3R RaDiology)

Sources : PVReunionClient220518 [725x762, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel
OrganisationTB_SandyMillius.xlsx

Annexe X : PV réunion client du 04.06.2018

Date	04.06.2018
Lieu	Institut de radiologie de Sion
Participants	M. Dr Hugues Brat, M. Cyril Thouly, M. Dr Henning Müller, Mme Sandy Millius

Points abordés	
N°	Détails
1	Ils vont nous envoyer les 200 rapports cette semaine
2	C'est la conclusion qui doit matcher sur l'indication
3	Ils veulent aussi qu'on puisse faire une recherche d'un terme précis dans les rapports (avec import de plusieurs fichiers à la fois et une certaine section) -> A voir si ça sera faisable et si j'aurais le temps car ce n'était pas dans le cahier des charges de base.
4	Ils voudraient bien qu'on ajoute une manière de valider les résultats avec possibilité d'ajout des termes justes aussi dans une bdd
5	En termes de stats, il faudra faire sur un panel de 1000
6	Dico NUANCE -> pas possible de l'avoir car payant
7	Pour les critères et % ok, ils disent qu'en principe 80 à 90 % devraient être ok mais il faudra voir en fonction des résultats qu'on aura obtenus d'une manière générale.

Sources : PVReunionClient040618 [1074x644, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel OrganisationTB_SandyMillius.xlsx

Annexe XI : PV réunion client du 18.06.2018

Date	18.06.2018
Lieu	Institut de radiologie de Sion
Participants	M. Dr Hugues Brat, M. Cyril Thouly, M. Dr Henning Müller, Mme Sandy Millius

Points abordés	
N°	Détails
1	Priorité : trouver un outil qui soit suffisamment complet car il y a effectivement des lacunes avec cTAKES car manque la partie matching
2	Contacter Federica Zanca -> fedezanca@gmail.com -> physicienne à qui je peux écrire de la part du client en expliquant le contexte, voir si elle connaît des outils comme nécessaire.
3	Voir si on peut trouver des outils de matching médicaux, ou moyen de financer des traductions.

Sources : PVReunionClient180618 [1304x537, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel
OrganisationTB_SandyMillius.xlsx

Annexe XII : PV réunion client du 09.07.2018

Date	09.07.2018
Lieu	Institut de radiologie de Sion
Participants	M. Dr Hugues Brat, M. Cyril Thouly, M. Dr Henning Müller, Mme Sandy Millius

Points abordés	
N°	Détails
1	Explication de l'avancement du TB -> Connexion à HeTOP qui fonctionne avec test, ECMT est cependant en révision et va avoir une mäj dans la semaine
2	Clients se rendent compte que le projet global de base est trop complexe pour aussi peu de temps, mais souhaiteraient néanmoins que la fonctionnalité de comparaison fonctionne et que si possible, ils puissent avoir accès à l'outil de manière graphique et simple (à voir selon le temps restant). Accent mis sur la fonctionnalité principale. Souhaiteraient que les résultats soient notés dans un csv ou autre fichier avec si possible le numéro de référence du dossier commençant par S et suivi de 11 chiffres.
3	Discussion pour une éventuelle participation à la traduction de RadLex (projet hors-TB) pour que ça puisse aider à l'avenir, notamment s'il y a une suite à mon développement

Sources : PVReunionClient090718 [1063x606, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel
OrganisationTB_SandyMillius.xlsx

Annexe XIII : PV réunion client du 19.07.2018

Date	19.07.2018
Lieu	Institut de radiologie de Sion
Participants	M. Dr Hugues Brat, M. Dr Henning Müller, Mme Sandy Millius

Points abordés	
N°	Détails
1	Présentation du prototype fonctionnel pour la première phase (ECMT)
2	Client satisfait du résultat -> avec les outputs trouvés, il peut directement s'imaginer quel type de rapport a été testé (expérience et connaissances métier que je n'ai pas), espère que la comparaison sera néanmoins faisable d'ici la fin du TB avec si possible un certain taux de concordance à calculer.
3	Client est d'accord sur le fait que toute la partie analyse qui se fera par la suite sur la comparaison, je ne peux pas l'implémenter car ça a attrait à l'humain et demanderait trop de temps supplémentaire
4	Dernière réunion fixée à jeudi 26.07.2018 afin de faire un point avant le rendu

Sources : PVReunionClient190718 [1306x619, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel
OrganisationTB_SandyMillius.xlsx

Annexe XIV : PV réunion client du 26.07.2018

Date	26.07.2018
Lieu	Institut de radiologie de Sion
Participants	M. Dr Hugues Brat, M. Dr Henning Müller, Mme Sandy Millius

Points abordés	
N°	Détails
1	Avancement du TB -> résultat auquel je suis arrivée -> client comprend que le temps pour la fin de la réalisation du projet était court et donc qu'il ne soit pas tout à fait terminé (l'idée étant d'avoir un simple prototype). Aurait bien sûr aimé qu'il puisse être terminé.
2	Apport de mon éventuelle solution de calcul de distance et idée de futur développement pour une éventuelle amélioration -> Réflexion semble correcte, à creuser
3	Le client voudra certainement amener ce projet plus loin -> A voir s'ils continueront le développement en interne ou s'il y a moyen de collaborer avec l'institut et des assistants de Henning, voir même pour une idée de futur TB
4	Client souhaiterait pouvoir être présent durant la défense orale
5	M. Cyril Thouly du service informatique étant absent, un rdv sera fixé avec les clients afin que je puisse également lui montrer le résultat du TB d'un point de vue plus technique

Sources : PVReunionClient260718 [880x550, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel
OrganisationTB_SandyMillius.xlsx

Annexe XVI : Résultat XML pour ECMT partiel et détaillé sur indication

```

1 <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2 <?xml-stylesheet type="text/xsl" href="/CISMeFecmtservice/xsl/ecmtservice.xsl"?>
3 <cis:sentences xmlns:cis="http://vs.ecmt.cisnef.org/" timemillis="8">
4 <cis:sentence idsentence="0" position="0" iddoc="0">
5 <cis:text>Douleurs basi-thoraciques gauches et au niveau de l'hypochondre gauche, avec irradiation vers le dos., , Technique</cis:text>
6 <cis:indexations>
7 <cis:indexation end="70" idterm="20392" idcisnef="FMA_CO_20392" offset="50, 64" start="50" ter="FMA" matchterms="l'hypochondre gauche" typeid="T_DESC_FMA_COMPOSANT"></cis:indexation>
8 <cis:indexation end="70" idterm="RID30002" idcisnef="RAD_CO_RID30002" offset="50, 64" start="50" ter="RAD" matchterms="l'hypochondre gauche" typeid="T_DESC_RADLEX_CONCEPT"></cis:indexation>
9 <cis:indexation end="93" idterm="264745005" idcisnef="SCT_CO_264745005" offset="64, 89" start="64" ter="SCT" matchterms="gauche vers" typeid="T_DESC_SNOCT_CONCEPT"></cis:indexation>
10 <cis:indexation end="100" idterm="14181" idcisnef="FMA_CO_14181" offset="97" start="97" ter="FMA" matchterms="dos" typeid="T_DESC_FMA_COMPOSANT"></cis:indexation>
11 <cis:indexation end="100" idterm="71938" idcisnef="FMA_CO_71938" offset="97" start="97" ter="FMA" matchterms="dos" typeid="T_DESC_FMA_COMPOSANT"></cis:indexation>
12 <cis:indexation end="100" idterm="70002114" idcisnef="ISH_L_70002114" offset="97" start="97" ter="ISH" matchterms="dos" typeid="T_DESC_MESH_CONCEPT"></cis:indexation>
13 <cis:indexation end="100" idterm="0001415" idcisnef="MESH_D_0001415" offset="97" start="97" ter="MESH" matchterms="dos" typeid="T_DESC_MESH_DESCRIPTOR"></cis:indexation>
14 </cis:term>
15 <cis:label lang="fr">dos</cis:label>
16 </cis:term>
17 <cis:altterms/>
18 <cis:category idcategory="T029" idcisnef="UML_ST_T029" origin="umis">
19 <cis:label lang="fr">localisation ou région du corps</cis:label>
20 </cis:category>
21 <cis:category idcategory="T029" idcisnef="UML_ST_T029" origin="umis">
22 <cis:label lang="fr">localisation ou région du corps</cis:label>
23 </cis:category>
24 <cis:related relationtypeid="T_REL_PTS_TO_PTS_VALID" relationlabel="Alignements automatiques CISNeF supervisés" idcisnef="FMA_CO_14181">
25 <cis:label lang="fr">dos</cis:label>
26 </cis:related>
27 <cis:related relationtypeid="T_REL_PTS_TO_PTS_VALID" relationlabel="Alignements automatiques CISNeF supervisés" idcisnef="NCI_CO_13062">
28 <cis:label lang="fr">dos</cis:label>
29 </cis:related>
30 <cis:related relationtypeid="T_REL_PTS_TO_PTS_VALID" relationlabel="Alignements automatiques CISNeF supervisés" idcisnef="SNO_NO_T_D2100">
31 <cis:label lang="fr">dos</cis:label>
32 </cis:related>
33 <cis:related relationtypeid="T_REL_PTS_TO_PTS_VALID" relationlabel="Alignements automatiques CISNeF supervisés" idcisnef="TSP_DE_003965">
34 <cis:label lang="fr">dos</cis:label>
35 </cis:related>
36 <cis:related relationtypeid="T_REL_PTS_TO_PTS_VALID" relationlabel="Alignements automatiques CISNeF supervisés" idcisnef="ICN_L_10003106">
37 <cis:label lang="fr">dos</cis:label>
38 </cis:related>
39 </cis:indexation>
40 </cis:sentence>
41 </cis:sentences>
42 </cis:sentences>
43 </cis:sentences>
44 </cis:sentences>
45 </cis:sentences>
46 </cis:sentences>
47 </cis:sentences>
48 </cis:sentences>
49 </cis:sentences>
50 </cis:sentences>
51 </cis:sentences>
52 </cis:sentences>
53 </cis:sentences>
54 </cis:sentences>
55 </cis:sentences>
56 </cis:sentences>
57 </cis:sentences>
58 </cis:sentences>
59 </cis:sentences>
60 </cis:sentences>
61 </cis:sentences>
62 </cis:sentences>
63 </cis:sentences>
64 </cis:sentences>
65 </cis:sentences>
66 </cis:sentences>
67 </cis:sentences>
68 </cis:sentences>
69 </cis:sentences>
70 </cis:sentences>
71 </cis:sentences>
72 </cis:sentences>
73 </cis:sentences>
74 </cis:sentences>
75 </cis:sentences>
76 </cis:sentences>
77 </cis:sentences>
78 </cis:sentences>
79 </cis:sentences>
80 </cis:sentences>
81 </cis:sentences>
82 </cis:sentences>
83 </cis:sentences>
84 </cis:sentences>
85 </cis:sentences>
86 </cis:sentences>
87 </cis:sentences>
88 </cis:sentences>
89 </cis:sentences>
90 </cis:sentences>
91 </cis:sentences>
92 </cis:sentences>
93 </cis:sentences>
94 </cis:sentences>
95 </cis:sentences>
96 </cis:sentences>
97 </cis:sentences>
98 </cis:sentences>
99 </cis:sentences>
100 </cis:sentences>
101 </cis:sentences>
102 </cis:sentences>
103 </cis:sentences>
104 </cis:sentences>
105 </cis:sentences>
106 </cis:sentences>
107 </cis:sentences>
108 </cis:sentences>
109 </cis:sentences>
110 </cis:sentences>
111 </cis:sentences>
112 </cis:sentences>
113 </cis:sentences>
114 </cis:sentences>
115 </cis:sentences>
116 </cis:sentences>
117 </cis:sentences>
118 </cis:sentences>
119 </cis:sentences>
120 </cis:sentences>
121 </cis:sentences>
122 </cis:sentences>
123 </cis:sentences>
124 </cis:sentences>
125 </cis:sentences>
126 </cis:sentences>
127 </cis:sentences>
128 </cis:sentences>
129 </cis:sentences>
130 </cis:sentences>
131 </cis:sentences>
132 </cis:sentences>
133 </cis:sentences>
134 </cis:sentences>
135 </cis:sentences>
136 </cis:sentences>
137 </cis:sentences>
138 </cis:sentences>
139 </cis:sentences>
140 </cis:sentences>
141 </cis:sentences>
142 </cis:sentences>
143 </cis:sentences>
144 </cis:sentences>
145 </cis:sentences>
146 </cis:sentences>
147 </cis:sentences>
148 </cis:sentences>
149 </cis:sentences>
150 </cis:sentences>
151 </cis:sentences>
152 </cis:sentences>
153 </cis:sentences>
154 </cis:sentences>
155 </cis:sentences>
156 </cis:sentences>
157 </cis:sentences>
158 </cis:sentences>
159 </cis:sentences>

```

Sources : ECMTXML2 [1517x605, PNG] Illustration réalisée par l'auteur depuis <https://codebeautify.org/xmlviewer#>

Annexe XVII : Résultat XML pour HeTOP pour le terme « dos » (1)

```

1 <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2 <cis:dboResp xmlns:cis="http://ws.cismef.org/">
3   <cis:dbo id="MSH_D_005490">
4     <cis:id>MSH_D_005490</cis:id>
5     <cis:ti id="T_DESC_MESH_DESCRIPTEUR" la="fr">Descripteur MeSH</cis:ti>
6     <cis:lb la="fr">infection focale</cis:lb>
7     <cis:an>&lt;sys&gt;last medline count=1810&lt;/sys&gt;&lt;rem&gt;MeSH2017&lt;/rem&gt;</cis:an>
8   <cis:hies>
9     <cis:hie>
10       <cis:id>H_4948101</cis:id>
11       <cis:ti id="BTNT_MSH" la="fr">Hiérarchie MeSH</cis:ti>
12       <cis:an>&lt;sys&gt;post-treatment&lt;/sys&gt;&lt;rem&gt;MeSH2017&lt;/rem&gt;</cis:an>
13     <cis:bt id="MSH_D_001423">
14       <cis:id>MSH_D_001423</cis:id>
15       <cis:ti id="T_DESC_MESH_DESCRIPTEUR" la="fr">Descripteur MeSH</cis:ti>
16       <cis:lb la="fr">infections bactériennes et mycoses</cis:lb>
17       <cis:an>&lt;sys&gt;last medline count=0&lt;/sys&gt;&lt;rem&gt;MeSH2017&lt;/rem&gt;</cis:an>
18     </cis:bt>
19   <cis:nt id="MSH_D_007239">
20     <cis:id>MSH_D_007239</cis:id>
21     <cis:ti id="T_DESC_MESH_DESCRIPTEUR" la="fr">Descripteur MeSH</cis:ti>
22     <cis:lb la="fr">infection</cis:lb>
23     <cis:an>&lt;sys&gt;last medline count=30304&lt;/sys&gt;&lt;rem&gt;MeSH2017&lt;/rem&gt;</cis:an>
24   </cis:nt>
25   <cis:path>MSH_D_ARBO/MSH_D_C/MSH_D_001423/MSH_D_007239</cis:path>
26   <cis:tree>C01.539</cis:tree>
27 </cis:hie>
28 <cis:hie>
29   <cis:id>H_4920491</cis:id>
30   <cis:ti id="BTNT_MSH" la="fr">Hiérarchie MeSH</cis:ti>
31   <cis:an>&lt;sys&gt;post-treatment&lt;/sys&gt;&lt;rem&gt;MeSH2017&lt;/rem&gt;</cis:an>
32 <cis:bt id="MSH_D_C">
33   <cis:id>MSH_D_C</cis:id>
34   <cis:ti id="T_DESC_MESH_DESCRIPTEUR" la="fr">Descripteur MeSH</cis:ti>
35   <cis:lb la="fr">Maladies</cis:lb>
36 </cis:bt>
37 <cis:nt id="MSH_D_001423">
38   <cis:id>MSH_D_001423</cis:id>
39   <cis:ti id="T_DESC_MESH_DESCRIPTEUR" la="fr">Descripteur MeSH</cis:ti>
40   <cis:lb la="fr">infections bactériennes et mycoses</cis:lb>
41   <cis:an>&lt;sys&gt;last medline count=0&lt;/sys&gt;&lt;rem&gt;MeSH2017&lt;/rem&gt;</cis:an>
42 </cis:nt>
43 <cis:path>MSH_D_ARBO/MSH_D_C/MSH_D_001423</cis:path>
44 <cis:tree>C01</cis:tree>
45 </cis:hie>

```

Sources : HeTOPXML1 [871x676, PNG] Illustration réalisée par l'auteur depuis <https://codebeautify.org/xmlviewer#>

Annexe XVIII : Résultat XML pour HeTOP pour le terme « dos »(2)

```

46 <cis:hie>
47   <cis:id>H_4908914</cis:id>
48   <cis:ti id="BTNT_MSH" la="fr">Hiérarchie MeSH</cis:ti>
49   <cis:an>&lt;rem&gt;MeSH2016&lt;/rem&gt;&lt;sys&gt;post-treatment&lt;/sys&gt;&lt;rem&gt;MeSH2016&lt;/rem&gt;</cis:an>
50   <cis:bt id="MSH_D_ARBO">
51     <cis:id>MSH_D_ARBO</cis:id>
52     <cis:ti id="T_DESC_MESH_DESCRIPTEUR" la="fr">Descripteur MeSH</cis:ti>
53     <cis:lb la="fr">arborescence MeSH</cis:lb>
54   </cis:bt>
55   <cis:nt id="MSH_D_C">
56     <cis:id>MSH_D_C</cis:id>
57     <cis:ti id="T_DESC_MESH_DESCRIPTEUR" la="fr">Descripteur MeSH</cis:ti>
58     <cis:lb la="fr">Maladies</cis:lb>
59   </cis:nt>
60   <cis:path>MSH_D_ARBO/MSH_D_C</cis:path>
61 </cis:hie>
62 <cis:hie>
63   <cis:id>H_4940075</cis:id>
64   <cis:ti id="BTNT_MSH" la="fr">Hiérarchie MeSH</cis:ti>
65   <cis:an>&lt;sys&gt;post-treatment&lt;/sys&gt;&lt;rem&gt;MeSH2017&lt;/rem&gt;</cis:an>
66   <cis:bt id="MSH_D_007239">
67     <cis:id>MSH_D_007239</cis:id>
68     <cis:ti id="T_DESC_MESH_DESCRIPTEUR" la="fr">Descripteur MeSH</cis:ti>
69     <cis:lb la="fr">infection</cis:lb>
70     <cis:an>&lt;sys&gt;last medline count=30304&lt;/sys&gt;&lt;rem&gt;MeSH2017&lt;/rem&gt;</cis:an>
71   </cis:bt>
72   <cis:nt id="MSH_D_005490">
73     <cis:id>MSH_D_005490</cis:id>
74     <cis:ti id="T_DESC_MESH_DESCRIPTEUR" la="fr">Descripteur MeSH</cis:ti>
75     <cis:lb la="fr">infection focale</cis:lb>
76     <cis:an>&lt;sys&gt;last medline count=1810&lt;/sys&gt;&lt;rem&gt;MeSH2017&lt;/rem&gt;</cis:an>
77   </cis:nt>
78   <cis:path>MSH_D_ARBO/MSH_D_C/MSH_D_001423/MSH_D_007239/MSH_D_005490</cis:path>
79   <cis:tree>C01.539.392</cis:tree>
80 </cis:hie>
81 </cis:hies>
82 </cis:dbo>
83 </cis:dboResp>

```

Sources : HeTOPXML2 [986x572, PNG] Illustration réalisée par l'auteur depuis <https://codebeautify.org/xmlviewer#>

Annexe XIX : Exemple de rapport anonymisé (1)



IRS
INSTITUT DE RADIOLOGIE DE SION

Radiologues
FMH

IRS Sion
CIMA Martigny
IRMC Monthey

G3R103043073/RIB/MUD/18.04.2018

Accès examen en ligne :
<http://radiologie.groupe3r.ch/user/login.php>
N°examen : S110966830

[CT-scan thoraco-abdominal] du 18.04.2018

Indications

Douleurs basi-thoraciques gauches et au niveau de l'hypochondre gauche, avec irradiation vers le dos.

Technique: acquisition volumique. Contraste intraveineux (Iopamiro 370) bien toléré.

Description

Thorax

Médiastin : vaisseaux de calibre normal. Pas d'adénopathie. Quelques ganglions banals.
Hiles pulmonaires : pas d'adénopathie ni de masse.
Poumons : fine condensation sous-pleurale à la jonction du Fowler et du segment postéro-basal lobaire inférieur droit, aspécifique. Discret épaississement des parois bronchiques périhiliaires bilatérales, sans comblement endoluminal. Verre dépoli d'allure **gravito**-dépendante postéro-basal lobaire inférieur des deux côtés.
Plèvre : sans épaississement focal ni épanchement. Pachypleurite biapicale.
Régions axillaires : pas d'adénopathie.
Paroi thoracique : sans particularité.

Abdomen

Foie : taille et structure normales, pas de lésion focale suspecte. Microkyste biliaire centro-hépatique droit.
Voies biliaires : de calibre normal. Vésicule de parois fines, sans calcul radio-opaque.
Pancréas : sans anomalie.
Rate : de taille normale, homogène.
Reins-surrénales : de morphologie normale.
Rétropéritoine : pas d'adénopathie ni de masse. Aorte de calibre normal.
Tube digestif : stase stercorale du cadre colique.
Péritoine : pas de liquide libre.
Vessie : parois régulières et non épaissies.
Organes génitaux : d'aspect normal.
Squelette : spondylose antérieure dorsale moyenne et inférieure, sans fracture-tassement. Pas de sténose foraminale de la charnière dorso-lombaire, notamment à gauche. Fusion manubrio-corporeale sternale. Pincement et protrusion discale focale postéro-médiane L5-S1.

Autres constatations : minime hernie ombilicale de contenu graisseux prépéritonéal avec collet de 4 mm.

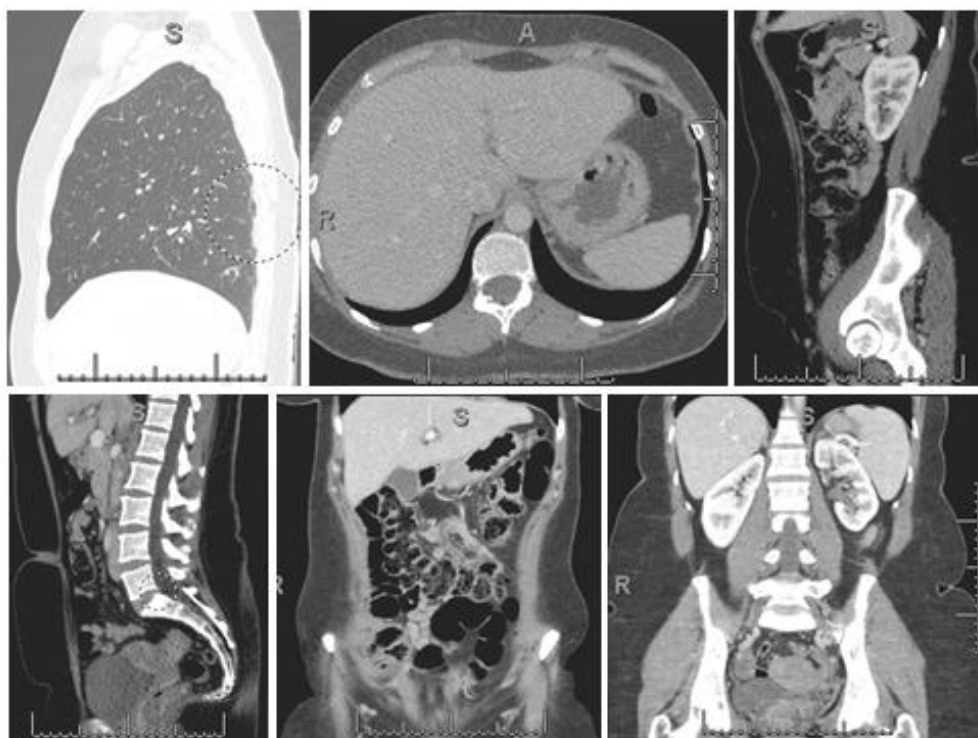
Sources : Rapport1 [580x749, PNG] Illustration réalisée par l'auteur depuis un rapport anonymisé fourni par les clients

Annexe XX : Exemple de rapport anonymisé (2)

Conclusion

Pas de foyer infectieux, inflammatoire ou tumoral thoraco-abdomino-pelvien, notamment à la base thoracique gauche et à l'hypochondre gauche.]

Avec nos remerciements et nos meilleures salutations.



Sources : Rapport2 [582x588, PNG] Illustration réalisée par l'auteur depuis un rapport anonymisé fourni par les clients

Annexe XXI : Détails tests cTAKES (1)

17.06.2018

Analyse de rapports dans cTAKES

Analyse très peu concluante au niveau du français, certainement l'obligation de les traduire systématiquement (à moins de pouvoir importer le RadLex partiellement traduit).

Outils ne font pas de matching

Voir comment gérer les concordances en termes de %.

Rapport (10) Français

Indication :

Douleurs de l'épaule droite depuis 3 jours. Tendinopathie calcifiante ?

Conclusion :

Radiographies de l'épaule droite confirmant une tendinopathie calcifiante inflammatoire active, en voie de dissolution avec migration de lait calcique dans la bourse sous-acromiale.

Proposition : infiltration échoguidée de la bourse sous-acromiale à visée antalgique.

Rapport (10) Traduit en anglais

Indication :

Right shoulder pain for 3 days. Calcifying tendinopathy?

- Right shoulder (AnatomicalSiteMention)
- Shoulder pain (SignSymptomMention)

Conclusion :

Radiographs of the **right shoulder** confirming an active calcifying inflammatory tendinopathy, in the process of dissolution with migration of **calcium** milk into the **subacromial bursa**.

Proposal: echoguided **infiltration** of the subacromial **analgesic bursa**.

- Radiographs (ProcedureMention)
- Right shoulder (AnatomicalSiteMention)
- Calcium (MedicationMention)
- Subacromial bursa (AnatomicalSiteMention)
- Infiltration (SignSymptomMention)
- Analgesic (MedicationMention)
- Bursa (AnatomicalSiteMention)

Rapport (120) Français

Indication :

Le 21 avril, chute à la maison.

Sources : TestcTAKES1 [582x739, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel
Tests_cTAKES.docx

Annexe XXII : Détails test cTAKES (2)

17.06.2018

Douleurs de l'aîne à gauche persistante.

Douleurs de la colonne dorsale en augmentation depuis hier.

Antécédents de fracture ischio et ilio-pubienne droite en 2014.

Fracture ilio – ischio-pubienne ? Tassement vertébral dorsal ?

Conclusion :

Bassin : fracture non déplacée de la branche ischio-pubienne gauche et probablement du fond du cotyle également.

Fracture de 50 % du corps de D9.

Pour une meilleure appréciation de la fracture du cotyle, un bilan CT complémentaire a été réalisé (cf. rapport séparé).

Rapport (120) Traduit en anglais

Indication :

April 21, falls at home.

Persistent groin pain on the left.

Back pain's been increasing since yesterday.

History of right ischio and ilio-pubic fracture in 2014.

Fracture ilio - ischio-pubic? Dorsal vertebral compression?

- Falls (SignSymptomMention)
- Persistent (DiseaseDisorderMention)
- Groin pain (SignSymptomMention)
- Groin (AnatomicalSiteMention)
- Pain (SignSymptomMention)
- Pubic fracture (DiseaseDisorderMention)
- Pubic (AnatomicalSiteMention)
- Fracture (DiseaseDisorderMention)
- Vertebral compression (SignSymptomMention)
- Vertebral (AnatomicalSiteMention)
- Compression (ProcedureMention)

Conclusion:

Pelvis: undisplaced fracture of the left ischio-pubic branch and probably of the bottom of the acetabulum as well.

50% fracture of D9's body.

For a better appreciation of the acetabulum fracture, an additional CT assessment was performed (see separate report).

- Pelvis (AnatomicalSiteMention)

Sources : TestcTAKES2 [583x770, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel
Tests_cTAKES.docx

Annexe XXIII : Détails tests cTAKES (3)

17.06.2018

- Undisplaced fracture (DiseaseDisorderMention)
- Fracture (DiseaseDisorderMention)
- Pubic (AnatomicalSiteMention)
- Probably (SignSymptomMention)
- Acetabulum (AnatomicalSiteMention)
- Acetabulum fracture (DiseaseDisorderMention)

Rapport (200) Français

Indication :

Chute sur poignet gauche. Exclure fracture.

Conclusion :

Fracture de l'extrémité distale du radius sans déplacement significatif.

Rapport (200) Traduit en anglais

Indication :

Fall on left wrist. Exclude fracture.

- Fall (SignSymptomMention)
- Left wrist (AnatomicalSiteMention)
- Wrist (AnatomicalSiteMention)
- Fracture (DiseaseDisorderMention)

Conclusion :

Fracture of the distal end of the radius without significant displacement.

- Fracture (DiseaseDisorderMention)
- End (MedicationMention) -> SNOMEDCT_US "Endoglin, human"
- Radius (AnatomicalSiteMention)
- Displacement (SignSymptomMention)

Sources : TestcTAKES3 [584x545, PNG] Illustration réalisée par l'auteur depuis son document Excel personnel
Tests_cTAKES.docx

Déclaration de l'auteur

« Je déclare, par ce document, que j'ai effectué le travail de Bachelor ci-annexé seul, sans autre aide que celles dûment signalées dans les références, et que je n'ai utilisé que les sources expressément mentionnées. Je ne donnerai aucune copie de ce rapport à un tiers sans l'autorisation conjointe du RF et du professeur chargé du suivi du travail de Bachelor, y compris au partenaire de recherche appliquée avec lequel j'ai collaboré, à l'exception des personnes qui m'ont fourni les principales informations nécessaires à la rédaction de ce travail et que je cite ci-après : Dr. Henning Müller, Dr. Hugues Brat, M. Cyril Thouly et Dr. Nathalie Casati. » (Rodier & Zuber, 2013)