

Auteurs : ZY	Documents de la recherche	Date de création : 31/05/2012
Référence : CDC_oper_Acronyme.docx	Fonctionnalité " Gestion de terminologie riche avec pivot et références multiples "	Version 1.0
Réviseurs :	Cahier des charges de l'opération " Gestion de terminologie riche avec pivot et références multiples "	Date de modification : 31/05/2012

## Cahier des charges de l'opération "Gestion de terminologie riche avec pivot et références multiples"

Version	Date	Personne	Description
1.0	31/05/2012	ZHANG Ying	Initialisation
1.1	06/06/2012	Mathieu Mangeot	Modification

### 1. Introduction

Cette opération concerne le travail de ZHANG Ying. Elle est motivée par le besoin initial de gérer les acronymes avec Libellex. Répondre à ce besoin constituera la partie "réalisation" de cette opération. Il existe déjà une fonction de gestion de terminologie dans Libellex mais elle n'est pas tout à fait satisfaisante. Il manque la possibilité de lier plusieurs termes différents à un seul référent via les notions de pivot et de prolexème. Ces notions permettent par exemple de faire le lien entre plusieurs termes qui désignent un même et unique référent : Jean-Paul II et Karol Jozef Wojtyla en français, ou en anglais John Paul II et Karol Jozef Wojtyla.

De même, certains liens évoluent avec le temps : le pape désignait Jean-Paul II en 2004 et Benoît XVI en 2012. Des pays parlant la même langue (ex : France et Suisse romande) peuvent également utiliser des mots différents pour le même concept. Ex : "chien renifleur" et "chien drogue". Inversement, le même terme peut désigner des concepts différents : Dans la province de langue allemande de Bolzano en Italie, le Landeshauptmann est le président du conseil provincial, avec des compétences beaucoup plus limitées que le Landeshauptmann autrichien, qui est à la tête de l'un des Etats (Land) de la fédération autrichienne.

Il existe également Jibiki, une plate-forme générique de gestion de bases lexicales permettant de gérer n'importe quel type de structure (macro et microstructure). On souhaite définir une structure de données permettant de gérer des terminologies riches reliées à des ontologies métier (onto-terminologies). Ce sera le thème de la partie "recherche" de l'opération.

Concernant la partie réalisation, un prototype fonctionnel de gestion de ces données sera implémenté. Ce document est une version préliminaire du cahier des charges (CDC) de cette opération. Ce CDC constituera le livrable de la tâche 1.

### 2. Situation antérieure

Le contexte de la tâche 2 ("réalisation") est constitué par l'état des lieux au moment de la demande de l'opération. Dans Libellex, il est possible de gérer des terminologies (importer, exporter, consulter, valider, afficher etc.) depuis la page « Gérer/Valider Terminologie ». Le projet Metricc a réalisé la fonction d'import/export de terminologie et la fonction d'extraction terminologique monolingue depuis un texte pour Libellex. Mais Libellex ne peut pas gérer des données plus riches. Par exemple, s'il existe un acronyme pour plusieurs termes ou différents acronymes pour le même terme (variant dans le temps), il n'est pas possible de créer les liens entre les acronymes et les termes. Il n'est pas non plus possible de distinguer des significations

Auteurs : ZY	Documents de la recherche	Date de création : 31/05/2012
Référence : CDC_oper_Acronyme.docx	Fonctionnalité " Gestion de terminologie riche avec pivot et références multiples "	Version 1.0
Réviseurs :	Cahier des charges de l'opération " Gestion de terminologie riche avec pivot et références multiples "	Date de modification : 31/05/2012

différentes pour un même terme dans une même langue (pour différents systèmes juridiques par exemple).

Le contexte de la tâche 3 ("recherche") est constitué de manière générale de l'état de lieux des thèses de Gilles Sérasset [1994], de Mathieu Mangeot [2001], et Hong-Thai Nguyen[2009]. Ces thèses ont déjà proposé des solutions pour gérer des données plus riches. De manière plus précise, il faudra étudier les notions de prolexème [Prolexbase, 2012], [Tran, 2006] et de gestion terminologique multilingue riche [LexAlp, 2006].

### 3. Buts

#### 3.1. Buts de la partie réalisation

Il s'agit de permettre à l'utilisateur de gérer des données plus riches, par exemple, consulter/chercher les acronymes de termes, et créer les liens entre les acronymes et leurs termes. Les structures (macro et micro) répondant aux seront définies dans la partie recherche. Ensuite, il sera nécessaire d'implémenter ces structures dans outil accessible depuis Libellex. Il faudra programmer des interfaces pour effectuer des recherches complexes à l'aide de filtres dans les structures précédemment définies.

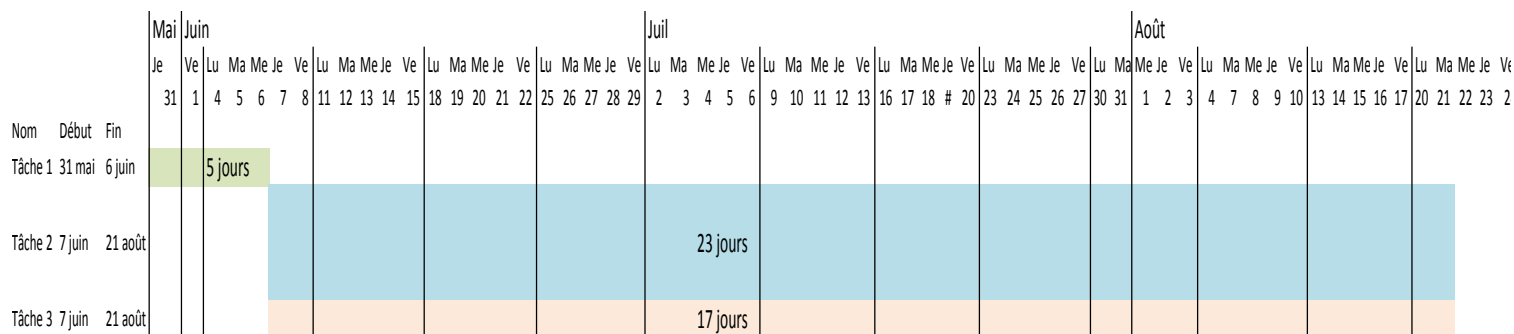
#### 3.2. Buts de la partie recherche

Il faudra tout d'abord définir une macrostructure (volumes et liens entre les volumes) permettant de gérer des données terminologiques riches. Au niveau de la microstructure, il faudra ensuite étudier le format TBX pour voir s'il peut être utilisé directement pour gérer les termes et leurs liens associés.

### 4. Organisation temporelle

Le temps total estimé pour la tâche est de 45 jours, répartis comme illustré ci-dessous.

- 5 jours de rédaction de l'opération
- 15 jours pleins pour la macrostructure (recherche)
- 2 jours pour la microstructure (recherche)
- 2 jours implémentation des structures dans l'outil (réalisation)
- 7 jours implémentation de la gestion des liens riches dans l'outil (réalisation)
- 7 jours communication entre l'outil et Libellex (réalisation)
- 7 jours pour les interfaces de recherche paramétrée (réalisation)



Auteurs : ZY	Documents de la recherche	Date de création : 31/05/2012
Référence : CDC_oper_Acronyme.docx	Fonctionnalité " Gestion de terminologie riche avec pivot et références multiples "	Version 1.0
Réviseurs :	Cahier des charges de l'opération " Gestion de terminologie riche avec pivot et références multiples "	Date de modification : 31/05/2012

## Bibliographie

- [Mangeot 2001] Mathieu MANGEOT, 2001, la thèse de docteur, *Environnements centralisés et distribués pour lexicographes et lexicologues en contexte multilingue*
- [Sérasset 1994] Gille SÉRASSET, 1994, la thèse de docteur, *Un système universel de bases lexicales multilingues et sa spécialisation aux bases lexicales interlingues par acceptions*
- [Nguyen 2010] Hong-Thai NGUYEN, 2009, la thèse de docteur, *Des systèmes de TA homogènes aux systèmes de TAO hétérogènes*
- [Prolexbase, 2012] visité en 2012 <http://www.cnrtl.fr/lexiques/prolex/doc/Prolexbase.pdf>
- [Tran, 2006] Mickaël Tran, 2006, la thèse de docteur, *Un dictionnaire relationnel multilingue de noms propres : conception, implémentation et gestion en ligne*
- [LexAlp, 2006] Gilles Sérasset, Francis Brunet-Manquat, 2006, *Multilingual Legal Terminology on the Jibiki Platform : The LexALP Project*