

Conception et peuplement d'une ontologie modélisant la notion de contexte enrichie par les fonctions lexicales pour la détection du sens dans le texte

Parler du Maroc central

Hammou Fadili (1) & Malika Chakiri (2)

(1) Laboratoire CEDRIC du CNAM de Paris

(2) Paris-Descartes-Sorbonne

This work focuses on the design and the development of a context ontology model, based on a domain ontology enriched, by lexico-semantic relations defining the lexical functions introduced by Mel'cuk in the Meaning-Text Theory, and by the concept of the context, in order to improve the analysis and the detection of meaning in text. This is motivated, by the fact that, unstructured data constitute the majority of produced contents, requiring for their exploitation, the development of tools and technologies allowing their integration into knowledge-based and reasoning-based systems. Existing technologies in the analysis and the extraction of semantic information from text can have a lot of imperfections; indeed, important elements such as concept of the context and use of all possible relations between terms, are not fully and formally supported. It is why we propose an extended model of a domain ontology as a context (in our case an ontology of fauna and flora), enriched by aspects that can help to address the cited problems. Such ontologies constitute the basis of an advanced approach for the detection and the extraction of contextual and semantic information from text, published in a parallel article (H.Fadili ACS/IEEE, 2013).

Introduction

Ce travail porte sur l'élaboration d'un modèle d'ontologie de contexte à partir d'une ontologie de domaine enrichie par des relations lexico-sémantiques définissant les fonctions lexicales de la Théorie Sens-Texte (Mel'cuk, 1988) et par la notion du contexte, dans un but d'améliorer l'analyse et la détection sémantique des données dans les documents de type textes. Cette démarche, s'intéressant au traitement sémantique des données non structurées, est motivée par le fait que, ce type de contenus constitue la majorité des données produites aujourd'hui, nécessitant pour leur exploitation la mise en place d'outils et des technologies spécifiques permettant leur intégration dans les systèmes à base de connaissances

et à base de raisonnements. Dans ce domaine, les technologies existantes relatives à l'analyse et à l'extraction sémantiques d'information peuvent représenter beaucoup de lacunes du fait que des concepts importants, comme la notion de contexte et l'exploitation de toutes les relations possibles entre les termes, n'y sont pas totalement pris en charge.

C'est dans ce contexte que nous avons proposé une nouvelle approche permettant d'améliorer certains processus de gestion contextuelle de la sémantique (Fadili, 2013), basée sur les ontologies de domaine que nous améliorons suivant un « modèle » présenté dans cet article. Le but est de détecter et de relever tous les « traits sémantiques » relatifs à un contexte donné et d'augmenter ainsi l'efficacité de la formalisation du sens d'un texte afin qu'il soit compris et interprété par la « machine ». La modélisation de cette « relation » entre le texte et le contexte consiste dans la formalisation du sens en extrayant des mots et des relations permettant l'obtention d'un réseau sémantique « contextualisé » reflétant fidèlement le sens des contenus étudiés. Concrètement, elle permet de faire « la projection » du texte représenté par l'arbre conceptuel (AC) (F.AMARDEILH, 2009) issu des différents processus du TAL « sur » le contexte représenté par l'ontologie du contexte (OC) en utilisant les « relations lexico-sémantiques » pour faire le « mapping » entre les concepts, mots, relations, instances, attributs, etc. des deux graphes.

Nous avons tenu à mettre l'accent sur l'utilisation de ce type de relations parce que nous considérons que c'est le moyen le plus sûr permettant d'éviter la déperdition du sens et de relever toutes les nuances dans un contenu en rapport avec le contexte. En effet, nous considérons que si un mot est important et qu'il doit être retenu dans un contexte, il en sera de même pour les mots qui lui sont liés, tels que ses synonymes, ses antonymes, ses converses, ses génériques, ses spécifiques, etc. La spécificité de chaque type de relations lexico-sémantiques est gérée dans les différents algorithmes de la projection. Outre les mots simples, notre ontologie contient des mots composés. Ils sont codés ou indexés pour que la machine puisse les extraire facilement et leur réserver un traitement spécial. Cette démarche permet d'éviter toute ambiguïté au niveau de l'interprétation. De par sa nature, cette approche peut être utilisée dans beaucoup de domaines liés à la notion du contexte comme l'indexation « sémantique » des données, la recherche linguistique ou sémantique d'information, l'extraction contextuelle d'information, la gestion de corpus multilingues, la génération automatique de textes, etc.

Le nombre des relations lexico-sémantiques étant important, une soixantaine environ, nous essayons, à travers cette analyse, de ne prendre en compte (dans un premier temps) que certaines relations pour valider notre approche. Bien évidemment, l'extension à d'autres relations est possible suivant le même principe.

Dans cette recherche, nous expérimentons le cas de la langue amazighe. La raison pour laquelle nous avons développé et peuplé, un modèle de contexte basé sur une ontologie du domaine de la faune et de la flore, enrichie par des relations lexico-sémantiques. Dans ce qui suit, nous présentons une analyse des motivations des

choix de conception du modèle, ainsi qu'un aperçu de l'implémentation de l'ontologie¹.

Ontologies

Définition

D'une manière générale, les ontologies reposent sur des outils de modélisation et de représentation des connaissances permettant à des communautés d'experts humains et logiciels d'un domaine donné de partager, d'une manière consensuelle, un vocabulaire et de parler un même langage, c'est-à-dire communiquer et se comprendre. Définir une ontologie d'une manière précise est une tâche très difficile. Plusieurs définitions existent suivant le contexte d'utilisation. Celle qui fait autorité aujourd'hui, au sein de la communauté scientifique, est celle de Gruber qui l'a définie comme suit :

« Une ontologie est la spécification d'une conceptualisation d'un domaine de connaissance ». En d'autres termes, une ontologie est un modèle de représentation de la formalisation d'une conception d'un domaine. La modélisation d'un domaine repose essentiellement sur deux aspects importants, la représentation des connaissances et le raisonnement qui peut leur être associé. La partie description permet de décrire et de formaliser le domaine en utilisant les notions de : classes, instances, attributs, relations, fonctions (ensembles de relations : simplification), restrictions (conditions sur certains éléments), etc. La partie raisonnement, quant à elle, décrit les aspects dynamiques liés à une ontologie, à travers des règles (règles mettant en valeur les éléments de l'ontologie), et des événements (événements modifiant certains attributs ou relations), etc.

Quelques éléments constitutifs

Nous présentons dans ce qui suit des éléments constitutifs de l'ontologie étudiée.

1. Concepts et instances

Les concepts représentent des objets ou des idées. Ils sont organisés en taxonomie au sein d'un réseau de concepts et structurés hiérarchiquement. Chaque concept est caractérisé par un ensemble de propriétés :

¹ Concernant la notation des entrées lexicales, nous utiliserons le protocole suivant :

- **Voyelles** : *a, i, u* et *e* pour noter le schwa.
- **Semi-voyelles** : *w, y*.
- **Consonnes** : *p, b, t, d, k, g, l, m, n, s, z, š, ž, h/ç* notent les fricatives pharyngales sourde et sonore, *x/g* les fricatives vélares sourde et sonore, *h* la spirante, *q* l'occlusive dorso-uvulaire, *r* la vibrante apicale. Le point sous la lettre indique l'emphase, le *w* en exposant note la labiovélarisation, le trait sous la lettre note la spirantisation, le dédoublement de la consonne indique la gémination.

- un concept est *générique* s'il exclut toute extension. Dans nos corpus, aucun concept ne relève de cette catégorie.
- un concept portant une propriété d'*identité* permet de différencier deux instances de ce concept.
- un concept est *rigide* s'il ne peut pas être une instance d'autres concepts. Exemple : *amuder* « être vivant » :

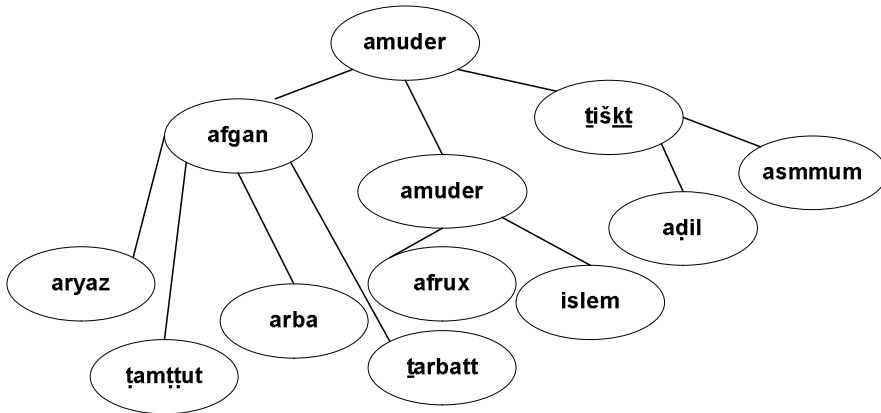


Figure 1 : Concept rigide

- un concept est *anti-rigide* lorsqu'il peut être une instance pour d'autres concepts. Dans l'exemple ci-dessus « afgan », « amuder » et « tışkt » sont des concepts anti-rigides.

Alors que le concept désigne l'intention du concept, l'instance renvoie à l'élément de l'ensemble constituant l'extension de concept.

2. Relations entre concepts

Au sein d'une ontologie, les concepts sont liés entre eux par des relations qui sont définies comme les liens entre des entités. Ces liens sont classés en deux catégories :

(1) Les *relations hiérarchiques* lient des éléments supérieurs dits « hyperonymes » et des éléments inférieurs dits « hyponymes ». De ce fait, l'hyperonyme englobe l'hyponyme. Exemple :

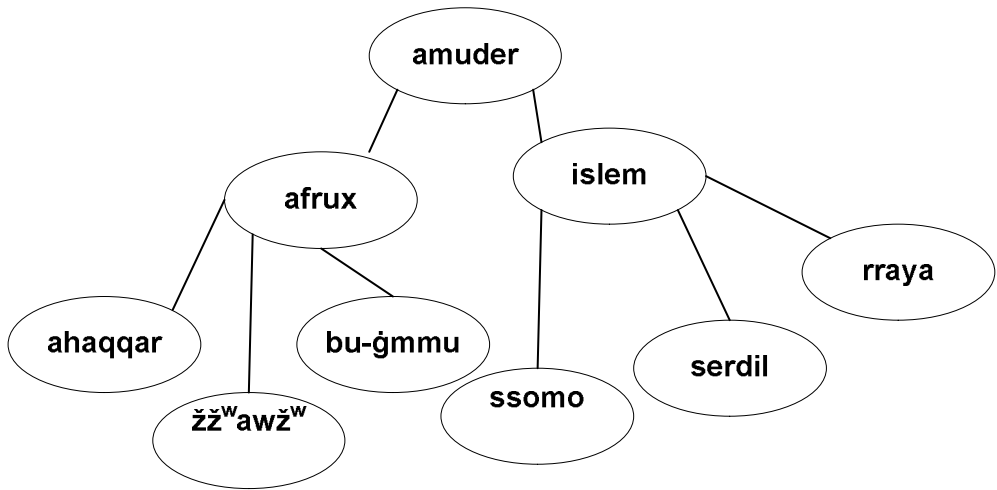


Figure 2 : Relations entre concepts

Ici, *amuder* est l'hyperonyme de *afrux* et *islem* ; *afrux* et *islem* sont les hyponymes de *amuder*.

afrux et *islem*, par rapport aux autres entités, peuvent, à leur tour, être considérés comme leurs hyperonymes.

A ne pas oublier de mentionner que *serdil* « sardine » dans certaines régions comme la région des Ait Wirra à Elksiba (Maroc), joue le rôle de l'hyperonyme (générique), dans ce sens qu'il renvoie à tout type de poisson.

(2) Les *relations sémantiques* correspondent à la structuration d'holonymie-méronymie. Elles peuvent être également considérées comme une relation hiérarchique liant un couple de concepts dont l'un dénote une partie de l'autre et l'autre comme un tout. Ce genre de relation est différent de la catégorie (1) dans la mesure où l'holonyme dispose des propriétés qui ne sont pas obligatoirement transmises à ses parties. Exemple : *aqbu* (tronc) est une partie de *tīškt* « arbre » mais n'est pas une sorte de *tīškt*. On peut représenter cette relation comme suit (1) faune, (2) flore :

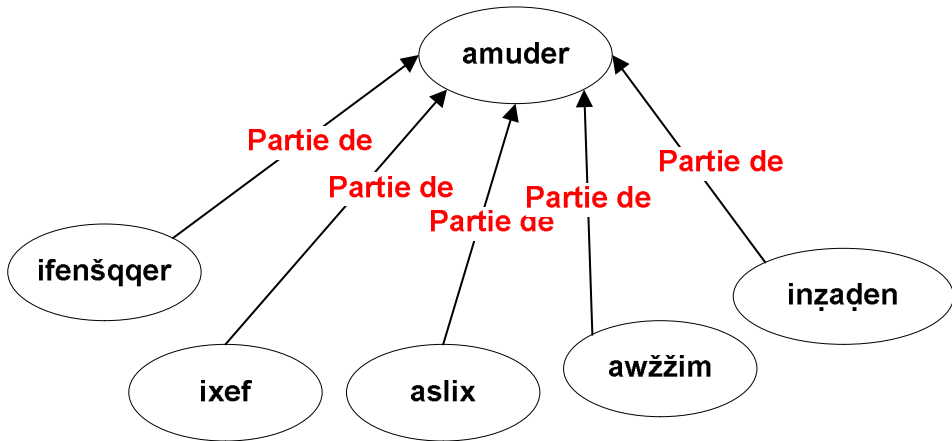


Figure 3 : Relations d'holonymie-méronymie

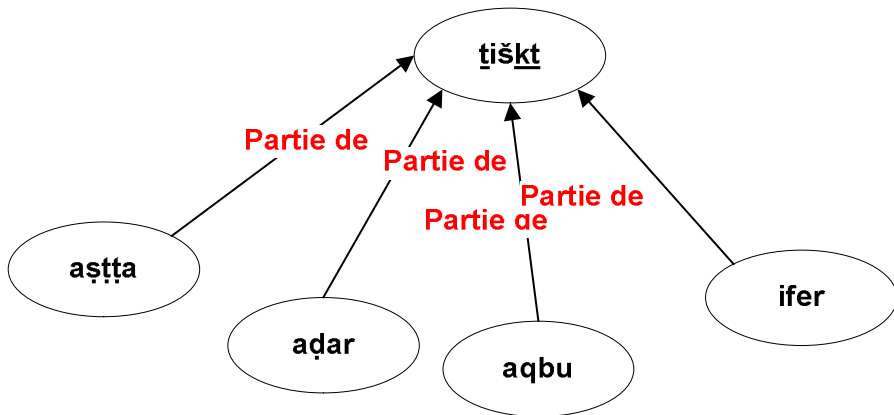


Figure 4 : Relations d'holonymie-méronymie

Outre ces relations sémantiques, les relations synonymiques et antonymiques, qui relèvent des liens horizontaux, seront également abordées dans l'élaboration de notre ontologie. Exemple : *iydi*, *buzaher*, *ammuter*, *amħdaw* « chien » ; *iddew*, *abağus* « singe ».

3. Propriétés

Ce sont des informations rattachées à chaque nœud du réseau sémantique. Exemple :

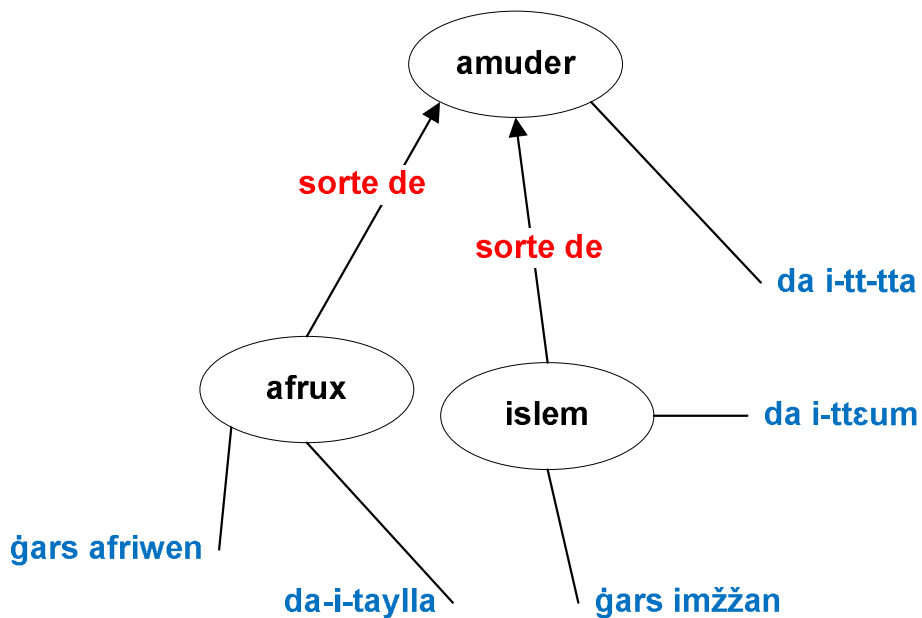


Figure 5 : propriétés

4. Attributs

Ce sont des relations qui relient un nœud concept à une valeur ou propriété. Ils définissent la structure de données. Exemple :

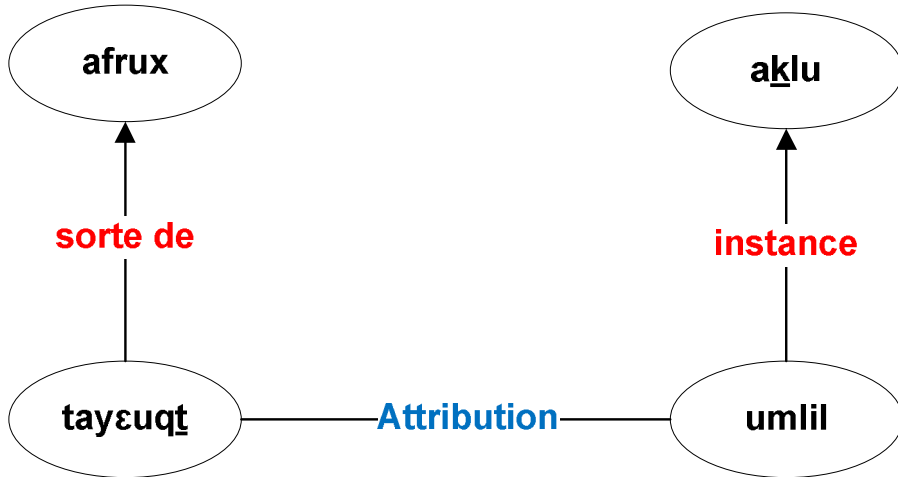


Figure 6 : Attributs

5. L'héritage

Il repose sur des liens de type « sorte de » ou « est un » qui relient un concept à un autre concept plus élevé. Ces liens sont étroitement liés à l'expression linguistique de la copule « est ». Ici, ils expriment soit l'appartenance d'un objet à une classe, soit l'inclusion. Exemple :

« *tayeuqt* » est une sorte de « *afrux* ».

« *tayeuqt* » est un « *afrux* ».

6. Types

En se basant sur certains travaux effectués dans le domaine de la classification d'ontologie (GómezPérez, 1999), (Guarino, 1997b), (Mizoguchi, 1998), (Mizoguchikeda, 1996), (VanHeijstAl, 1997), (VanwelkenhuysenAl, 1994), (VanwelkenhuysenAl, 1995), (WielingaSchreiber, 1993), etc., on peut en déduire la classification (dépendante des besoins d'utilisation) suivante :

- L'ontologie de haut niveau définit et représente les concepts de haut niveau qui sont des concepts de l'univers de modélisation commun aux ontologies des niveaux inférieurs, comme la notion du temps, de l'espace, etc. Son rôle est de réduire les ambiguïtés des termes entre les différentes ontologies et utilisations.
- L'ontologie générique se situe d'un point de vue abstraction entre l'ontologie de haut niveau et l'ontologie du domaine. Ses concepts sont

moins abstraits que l'ontologie de haut niveau et plus génériques par rapport à ceux de plusieurs ontologies de domaines, réutilisables par plusieurs domaines.

- L'ontologie de domaine représente et décrit les concepts se rapportant à un domaine donné.
- L'ontologie de tâches représente et décrit le vocabulaire propre à une activité décrivant une structure de résolution d'un problème.
- L'ontologie d'application représente et décrit les concepts propres à une application. Dans la plupart des cas, ces concepts proviennent de l'ontologie du domaine et de l'ontologie des tâches.
- L'ontologie de représentation représente et décrit les primitives des langages de représentation et de formalisation de connaissances : RDF, OWL, réseau sémantique, LD, LPO, etc.

7. Les langages

Les langages constituent une partie très importante des systèmes à base d'ontologies ; ils permettent la gestion de la base de connaissances et de raisonnements qui peuvent leur être associés. Les langages présentés dans ce paragraphe sont des recommandations du W3C, basés sur des normes, standards, etc. ayant pour but de permettre des traitements automatiques sur des contenus par des applications différentes. Nous verrons, un peu loin, que ceci fait de la notion d'ontologie un principe fondamental pour le partage des connaissances, la communication et la coopération au sein des systèmes distribués. Ci-après, nous reproduisons quelques éléments de description de ces langages, extraits de la pyramide du Web sémantique de Berners-Lee.

URI/IRI : les URIs (Uniform Resource Identifier) permettent d'identifier d'une manière unique les ressources sur le WEB et les IRIs (Internationalized Resource Identifiers) permettent aux personnes d'identifier des ressources Web dans leur propre langue.

XML : Extended Markup Language (XML) est un langage de description et d'échanges de documents et des données ne permettant pas leur présentation, appelé aussi format d'échanges standardisé. Il permet aux applications reconnaissant ce format d'échanger tous les types de données décrits dans ce langage, utilisé souvent pour assurer la compatibilité des données entre les applications hétérogènes. Exemple, on peut décrire la voiture 123 de la marque Renault et de couleur rouge.

RDF : Resource Description Framework (RDF) est un métalangage qui sert à décrire les ressources, leurs propriétés et les valeurs des propriétés sous forme d'un graphe (ressource, propriété, valeur). Il est considéré comme un modèle standardisé de description des métadonnées qui peut être défini et associé à des documents ou à des contenus. Ces annotations et métadonnées permettent d'associer du sens à des contenus qui peuvent être traités d'une manière automatique par les agents logiciels. Pour la gestion sémantique des données, les métadonnées RDF peuvent être des informations sémantiques associées à des mots du texte. Ces éléments peuvent être ensuite analysés et interprétés pour l'extraction du sens global. Il est à

noter que RDF peut être exprimé dans plusieurs langages, mais c'est XML qui est souvent utilisé, une version XML de RDF appelée RDF/XML a été même créée.

RDFS : RDF peut également être utilisé pour décrire des situations et des utilisations particulières avec un vocabulaire bien précis en utilisant la notion de RDF Schéma (RDFS). RDFS consiste à adapter RDF à des domaines modélisés particuliers décrivant des utilisations particulières au sein d'une communauté. La définition d'un schéma RDF consiste en une activité de typage et de classification des ressources, des propriétés et des relations sous forme de classes définies dans des « espaces de noms » servant principalement à désambiguïser les mêmes éléments d'un vocabulaire définis dans des utilisations ou espaces de noms différents.

OWL : c'est une extension de RDF enrichie avec des propriétés sémantiques, de contraintes, de comparaisons, de cardinalités, etc. pour décrire et manipuler les ontologies. C'est un langage recommandé par le W3C basé, comme RDF, sur XML qui permet à des moteurs d'inférences d'agents l'interprétation et le raisonnement automatiques sur les ontologies. En effet, OWL est basé sur les logiques de description utilisées dans les systèmes de représentation de connaissances et offrant de fortes possibilités de manipulation de prédicats de classes, de rôles et d'individus, donc d'ontologies, contrairement aux logiques de premier ordre classiques ne manipulant que des objets de même type.

Il existe trois versions d'OWL :

OWL Full : c'est la version la plus complète, elle combine toutes les primitives d'OWL avec RDF/S sans respecter aucune contrainte si ce n'est celle de RDF, cette version n'est pas décidable.

OWL DL : c'est un sous-ensemble d'OWL Full basé sur la logique de description n'autorisant que des constructions garantissant la décidabilité des inférences.

OWL Lite : c'est la version la plus simple, elle sert principalement à la création de hiérarchies de classes. Elle est dotée seulement de quelques propriétés de classes comme la comparaison, la restriction, la cardinalité (0 ou 1).

SPARQL (Protocol and RDF Query Language) : langage d'interrogation des ontologies représentées sous forme de graphes RDF/S. Il est pour les bases de connaissances RDF ce que SQL est pour les bases de données relationnelles. *Exemple : SELECT ?x, ?y, ?z FROM URI/IRI WHERE {Conditions1, condition2, etc.}*

RIF (Rule Interchange Format) : format d'échange de règles standardisé. Il permet de faciliter les échanges de règles utilisables par des systèmes distribués sur le WEB en assurant l'interopérabilité et la portabilité entre divers langages et moteurs de règles.

SWRL (Semantic Web Rule Language) : langage de raisonnement sémantique à base de règles sur les ontologies. C'est une combinaison d'OWL-DL et RuleML langage créé principalement pour le Web sémantique pour pouvoir développer des règles sémantiques au niveau des agents. Il permet, contrairement à OWL, de manipuler les instances par des variables, de définir des fonctions mathématiques, des types de données, etc. *Exemple : avec OWL, la relation oncle ne peut être*

définie que comme suivant : *intersectionOf(SubClassOf(Homme), estfrereDe(Pere))*, avec SWRL on peut la définir au niveau des instances représentées par des variables x, y, z comme suivant : $Personne(?x) \wedge Personne(?y) \wedge Personne(?z) \wedge pere(?x, ?y) \wedge frere(?x, ?z) \Rightarrow oncle(?z, ?y)$. Cela permet de définir qui est l'oncle de qui, impossible à définir avec OWL. SWRL se différencie, en plus, d'OWL par le fait qu'il ne peut créer ni de nouveaux concepts ni de nouvelles relations, excepté ceux créés par la manipulation des variables et par la satisfaction des règles d'inférences.

Le choix des fonctions lexicales et des relations lexico-sémantiques

Les relations sémantiques et conceptuelles

Ci-après, nous présentons un bref aperçu sur les relations sémantiques et conceptuelles qui nous ont permis d'introduire les relations lexico-sémantiques utilisées dans notre démarche.

Les relations sémantiques sont les liens entre les éléments du lexique dans le texte. Elles représentent les liens de sens que peuvent entretenir deux ou plusieurs mots par rapport à leurs significations, comme les relations de type synonymie, antonymie, hyperonymie, etc. Les relations conceptuelles sont des relations ou associations utilisées dans la modélisation informatique. Elles permettent de faire des liens entre les classes et les instances. Dans le cas de la modélisation ontologique, nous pouvons bien sûr utiliser ce genre de relation, mais ce sont les relations de types hiérarchie et classification : héritage, classification comme « is-a, sort-of, part-of, etc. » qui sont privilégiées et souvent utilisées.

Dans la littérature, une distinction est faite entre les relations sémantiques Ahmad et Fulford (1992) et les relations conceptuelles Sager (1990) et Condamines et Rebeyrolle (1998). Dans L'Homme (2004), on explique la différence entre les relations conceptuelles et les relations sémantiques. Dans d'autres travaux, on explique aussi que certaines relations sémantiques sont très proches des relations de représentation conceptuelle : c'est le cas, notamment, de celles qui interviennent dans les taxinomies et les méronymies. Ce type de relation concerne, le plus souvent, des termes qui renvoient à des entités.

On peut conclure qu'il est extrêmement difficile de faire une réelle différence entre ce qui relève du conceptuel et ce qui est lié au sémantique. Les relations lexico-sémantiques définies dans le cadre des fonctions lexicales de Mel'čuk est un modèle de relations complet intégrant au moins les deux types de relations définies précédemment. C'est une des raisons qui nous a permis de les adopter dans notre démarche.

Les relations lexico-sémantiques

C'est suite à ces différentes analyses et expériences menées dans ce domaine que le choix de l'utilisation des relations lexico-sémantiques définies par les fonctions lexicales (Mel'čuk 1988), (Mel'čuk et al., 1995) et (Polguère, 2003) s'est imposé. Car elles permettent de représenter tous les types de relations dont nous avons

besoin : sémantiques, ontologiques, lexicales, etc. Nous considérons que cela permet une modélisation fine de tous les traits sémantiques améliorant le processus d'analyse des textes et de détection de la sémantique à travers une ontologie de contexte. En effet, l'utilisation de ces relations est importante. Elle permet, dans le cas de la recherche ou de l'extraction d'information, par exemple, d'éviter d'ignorer des mots ou des phrases bien qu'ils soient pertinents par rapport à un contexte (silence). Elle permet également d'éviter d'extraire des mots ou des phrases non pertinents (bruit). On considère que si un mot est pertinent dans un contexte, alors tous les mots que l'on peut atteindre *via* les relations lexico-sémantiques sont aussi importants que le mot lui-même.

Suivant le nombre des relations utilisées, la probabilité qu'un mot et ceux qui lui sont sémantiquement liés soient considérés est supérieure à la probabilité ne tenant compte que du mot seul. L'utilisation de ces relations peut donc avoir un impact très positif sur l'amélioration des performances des applications où elles peuvent être utilisées.

Fonctions lexicales

Comme il a été mentionné ci-dessus, le présent travail porte sur l'élaboration d'un modèle d'ontologies de domaine enrichie par des liens lexico-sémantiques associés aux fonctions lexicales de la Théorie Sens-Texte (Mel'čuk, 1997) et la notion du contexte, dans un but d'améliorer l'analyse et la détection sémantique des données dans les documents de type textes. La Théorie Sens-Texte est un cadre théorique linguistique, développée à Moscou par Aleksandr Žolkovskij et Igor Mel'čuk, pour la construction de modèles du langage naturel Mel'čuk (1981) et (1988). C'est une des théories majeures pour le Traitement Automatique des Langues fournissant des bases solides pour beaucoup de domaines d'application liés au traitement automatique de la sémantique, comme la traduction automatique, la recherche sémantique ou linguistique, l'extraction d'information, etc. Elle est basée sur une formalisation globale de la langue utilisant des fonctions mathématiques définissant les fonctions lexicales, comme suivant :

Une fonction lexicale (FL) permet de définir la description et la modélisation des relations lexicales, de collocations et de dérivation sémantique entre les unités lexicales (UL) d'une langue. Elle se présente sous forme d'une fonction au sens mathématique : $f(L) = \{L1, L2...Ln\}$ où f = le nom de la FL, L = lexie qui est l'argument de la FL et $\{L1, L2...Ln\}$ = ensemble des lexies qui constituent la valeur de la FL auprès de L , cf. Žolkovskij & Mel'čuk 1967, Mel'čuk 1974, 1996, 1998, 2003, 2007, et Wanner (éd.) 1996.

Voici un extrait de la liste des fonctions lexicales :

Syn, Anti, Conv_{ij}, Contr, Epit, Gener, Figur, S₀, A₀, V₀, Adv₀, S_i, S_{instr}, S_{med}, S_{mod}, S_{loc}, S_{res}, Able_i, Qual_i, Sing, Mult, Cap, Equip, A_i, Adv_i, Imper, Result, Centr, Magn, Plus, Minus, Ver, Bon, Pos_i, Loc_{in}, Loc_{ab}, Loc_{ad}, Instr, Propt, Copul, Pred, Oper_i, Func_i, Labor_{ij}, Incep, Cont, Fin, Caus, Perm, Liqu, Real_i, Fact_i, Labreal_{ij}, Invol, Manif, Prox, Prepar_i, Degrad, Son, Obstr_i, Stop_i, Excess_i, Symp_{ijk}.

Parmi toutes les fonctions lexicales, deux groupes sont à distinguer : les FL PARADIGMATIQUES et les FL SYNTAGMATIQUES.

Dans ce qui suit, on trouvera un bref aperçu sur quelques fonctions, les principales sur lesquelles des tests ont été effectués dans le cadre de ce travail. Une étude détaillée de toutes les fonctions ainsi que leur intégration dans le processus d'analyse et de détection est prévue dans un travail ultérieur.

Les fonctions lexicales paradigmatiques permettent de définir les liens entre les lexies et la manière dont l'une peut être atteinte à partir de l'autre. Elles correspondent à des relations à la fois sémantiques et formelles entre deux mots. Exemples :

- **Syn**, qui met en relation une entité avec ses parasyonymes : *Syn(iydi)=bu-zaher, amħdar*.
- **Anti**, qui met en relation une entité avec ses antonymes : *Anti(amsksum)=amstuyya*,
- **Conv**, qui prend en considération les relations de conversions entre deux entités : *conv(isġa)=izzenza*.

La notion de contexte et l'ontologie de contexte

Le contexte

La notion de contexte constitue un élément fondamental dans plusieurs domaines, notamment dans le domaine de Traitement Automatique du Langage Naturel (TALN), l'analyse sémantique de données, la gestion de connaissances (Semantic Data Mining & Knowledge Management : SDMKM), etc. où l'on utilise plus particulièrement au niveau de la désambiguïsation et de l'interprétation automatiques de la sémantique des données. On peut le définir comme un ensemble de situations, où chaque situation est composée d'un domaine, d'un ensemble de contraintes d'utilisation et d'un profil utilisateur. Il peut être représenté par un ensemble de propriétés décrivant des situations d'utilisations particulières et un vocabulaire associé au domaine modélisé. Ces éléments peuvent être partagés au sein d'une même communauté et représentés sous forme d'une ontologie échangeable entre les différents agents humains et logiciels. Ce qui leur permet de parler un même langage et d'interpréter les éléments de la même manière.

Concrètement, le contexte est composé :

- D'un domaine défini par un schéma représentant une activité de typage et de classification des ressources, des propriétés et des relations sous forme de classes dans des « espaces de noms » servant principalement à désambiguïser les mêmes éléments d'un vocabulaire par rapport à des utilisations ou espaces de noms différents.
- Contraintes définies par des règles représentant les contraintes d'une utilisation particulière, comme par exemple inclure et exclure des éléments qui seraient respectivement exclus et y inclus du domaine, cela correspond à un réglage fin supplémentaire de l'ontologie du domaine. Des éléments

issus de la théorie du discours peuvent être intégrés dans cette partie. On peut citer dans ce cas-là et à titre d'exemple, l'interprétation de certains éléments suivant leur auteur, le message qu'on souhaite extraire du texte, etc.

- Profil utilisateur qui est un élément important de la notion de contexte. Cela peut correspondre dans le cas de l'analyse du texte à ce qui est perçu par le lecteur suivant ses compétences, expériences, etc.

Une vue synthétique du méta-modèle de la notion de contexte est représentée dans le graphe suivant.

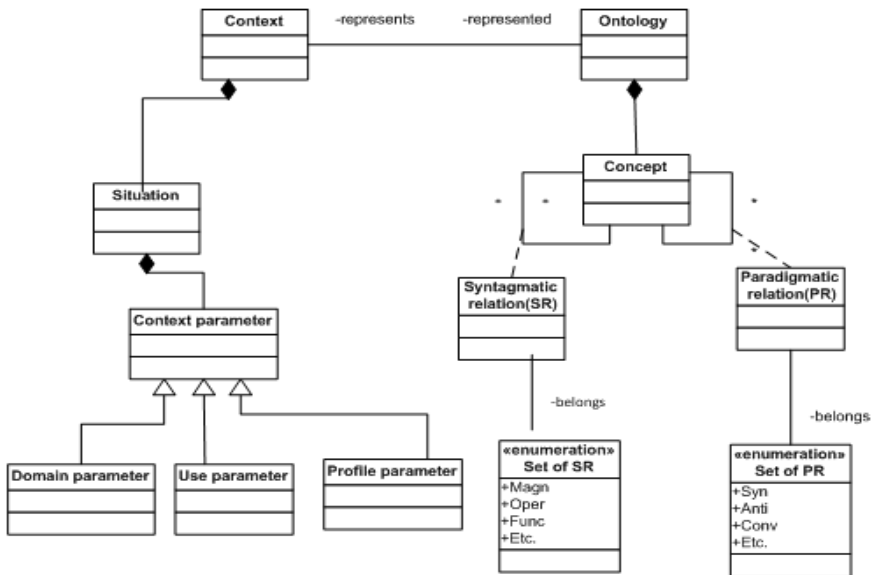


Figure 7 : Méta-modèle du contexte

L'ontologie de contexte

Dans cette étude, nous avons conçu et développé une ontologie du contexte intégrable dans le processus général de l'analyse sémantique et contextuelle du texte, proposé dans un travail parallèle (H.Fadili ACS/IEEE, 2013). C'est une ontologie du domaine que nous avons enrichie par la notion de contexte et les relations lexico-sémantiques décrites précédemment. En effet, étant donné que l'ontologie du domaine est seulement une ontologie conceptuelle du domaine étudié, nous avons estimé qu'il est indispensable, pour l'analyse sémantique à proprement dit, d'enrichir l'ontologie du domaine classique par des relations lexico-sémantiques permettant de prendre en compte, outre les relations conceptuelles entre les concepts, les relations lexico-sémantiques. Cette démarche rend possible une analyse linguistique de bas niveau entre les mots dans le contexte. Elle peut être considérée comme une ontologie du domaine fusionnée avec une ontologie linguistique, définie par deux types de liens : liens verticaux (contextualisés) qui sont des liens de catégorisation, et les liens horizontaux (décontextualisés) sont des liens linguistiques (relations lexico-sémantiques, actions, verbes, etc.). Tous ces éléments pris en compte, sont enrichis par des contraintes d'utilisation.

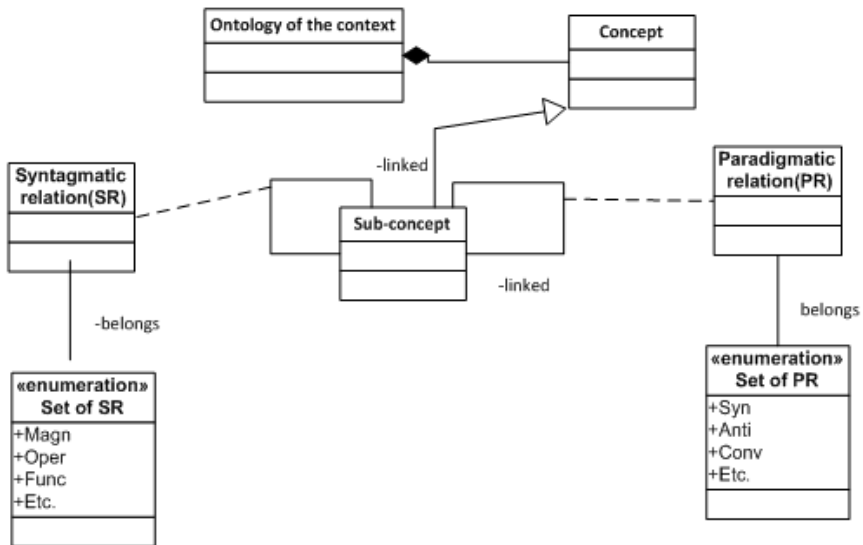


Figure 8 : Méta-modèle de l'ontologie du contexte

Constitution du corpus

D'une manière générale, la constitution d'une ontologie est une tâche très difficile ; cette tâche est encore plus difficile dans le cas de l'amazighe. Dans cette partie, l'objectif est de relever les exigences et les problèmes à résoudre pour la création

d'une telle ontologie en amazighe. Parmi lesquelles, on cite, les problèmes classiques liés à la création et à la gestion du corpus, d'une manière générale, puis ceux liés à la création et à la gestion de corpus amazighs, en particulier. A mentionner que la définition des termes scientifiques dans l'ontologie qui consiste dans la recherche, la création ou la réutilisation de termes scientifiques ou savants par rapport à un domaine ou à une discipline donnée n'est pas sans poser problème.

Dans ce qui suit, nous présentons les difficultés rencontrées, lors de la mise en place de l'ontologie scientifique faune-flore, en amazighe :

1. Nous nous sommes rendus compte que des données sources, que ce soit des données brutes y compris papier, ou des données électroniques, sont très rares :

Absence de sources de données, seulement quelques publications.

2. D'autres problèmes sont liés à la création de nouveaux contenus :

- Mots, concepts, etc.,
- Relations,
- Lexique scientifique,
- Problèmes de traduction,
- Traduction des mots scientifiques,
- Création de liens et des correspondances avec d'autres langues.

3. Applications avancées supportant le Tifinagh

- Absence d'applications totalement multilingues, problèmes rencontrés avec les éditeurs d'ontologies dans le Cloud2 par exemple.

4. Copyright

- Manque de données ouvertes et libres d'accès,
- Non institutionnalisation de la langue, ne permet pas la création de nouvelles données publiques.

Quelques solutions de contournements

Au niveau du contenu

Concernant la définition des concepts, nous nous sommes basés sur notre connaissance de la langue amazighe, sur des dictionnaires spécialisés et sur d'autres sources de données existantes.

Pour ce qui est de la définition des termes, il a été procédé ainsi :

1. Etant donné que les termes traités sont des entités du monde réel, nous avons défini chaque mot connu par ses relations et ses propriétés avant de l'insérer dans l'ontologie.

² Le cloud ou le cloud computing (le nuage en Français) est une technologie permettant la mutualisation des services et ressources informatiques ouverts en libre service.

Conception et peuplement d'une ontologie modélisant la notion de contexte enrichie par les fonctions lexicales pour la détection du sens dans le texte. Parler du Maroc central

2. Un mot non connu est traduit à partir d'autres langues. Dans certains cas, on a procédé par l'emprunt des termes les plus utilisés dans le langage courant : darija, arabe standard ou français, etc.

Au niveau technologique

Afin de rester conforme aux normes internationales, nous avons utilisé trois catégories de relations :

- les versions anglaises des relations de hiérarchisation OWL,
- les relations lexico-sémantiques dans leur version normalisée,
- puis les relations en amazighe pour les autres.

Ces règles ont été utilisées pour les propriétés (les propriétés hors la norme OWL).

Pour la graphie, faute de compatibilité avec le Tifinagh, nous avons opté, quant à cette première version de l'ontologie, pour la transcription phonétique. Une version en Tifinagh est prévue dans les versions futures.

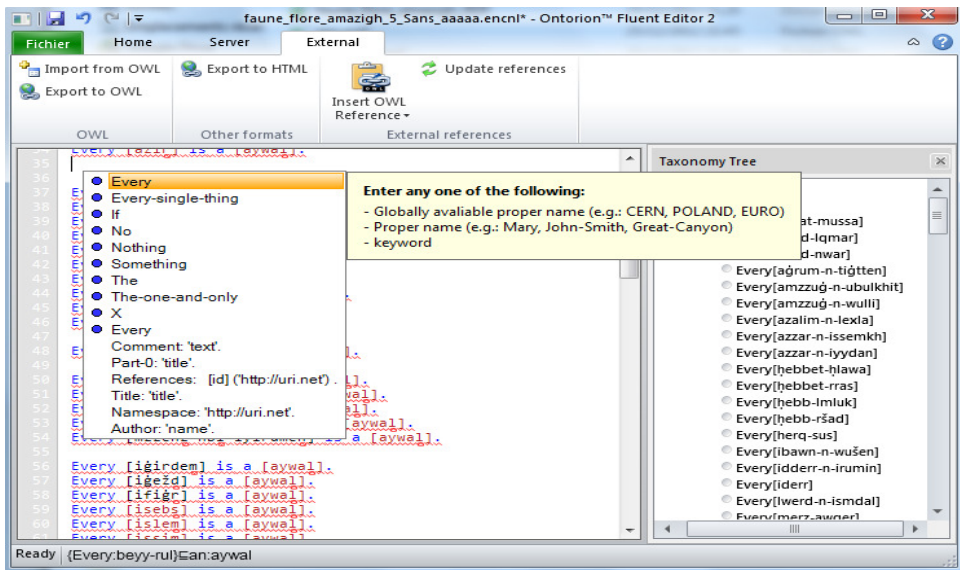


Figure 9 : Editeur facile

Afin d'accélérer la saisie, nous avons contourné, dans un premier temps, l'utilisation directe de l'interface Knoodl qui peut paraître difficile, nous avons utilisé l'application et la notation « *Fluent éditeur* » basé sur l'anglais contrôlé pour saisir facilement les données de l'ontologie, avant de les convertir en OWL et les importer dans la plateforme finale de travail (knoodl).

La saisie de fait en respectant la grammaire du CNL pour l'anglais comme suivant :

- (1) Every *abağus* is an *amuder*.
 Every *abrriđ* is an *amuder*.
 Every *abrriđ-n-taydwin* is an *amuder*.
 Every *abulxir* is an *amuder*.
 Every *ađil-n-wušen* is an *amuder*.
 Every *afrux* is an *amuder*.
 Every *afullus* is an *amuder*.
 Every *ağyul* is an *amuder*.
 Every *ağ^wilas* is an *amuder*.
 Every *ađerđay-n-lxla* is an *amuder*.
 Every *aħuliy* is an *amuder*.
 Every *ašitar* is an *amuder*.
 Every *ayis-n-lbħer* is an *amuder*.
- (2) Every *azmmur* is an *tiškt*.
 Every *tasafıt* is an *tiškt*.
 Every *tasmmumt* is an *tiškt*.
 Every *tašta* is an *tiškt*.
 Every *taylilut* is an *tiškt*.
 Every *tiqqi* is an *tiškt*.

Après cela, on procède à un export en RDF, facilement intégrable de dans la plateforme finale.

```
- <Class rdf:about="http://ontorion.com/unknown.owl/a3sat-mussa">
  <rdfs:subClassOf rdf:resource="http://ontorion.com/unknown.owl/ašklu"/>
</Class>
<!-- http://ontorion.com/unknown.owl/a3ud-lqmar -->
- <Class rdf:about="http://ontorion.com/unknown.owl/a3ud-lqmar">
  <rdfs:subClassOf rdf:resource="http://ontorion.com/unknown.owl/ašklu"/>
</Class>
<!-- http://ontorion.com/unknown.owl/a3ud-nwar -->
- <Class rdf:about="http://ontorion.com/unknown.owl/a3ud-nwar">
  <rdfs:subClassOf rdf:resource="http://ontorion.com/unknown.owl/ašklu"/>
</Class>
<!-- http://ontorion.com/unknown.owl/abağus -->
- <Class rdf:about="http://ontorion.com/unknown.owl/abağus">
  <rdfs:subClassOf rdf:resource="http://ontorion.com/unknown.owl/aywal"/>
</Class>
<!-- http://ontorion.com/unknown.owl/abrriđ -->
- <Class rdf:about="http://ontorion.com/unknown.owl/abrriđ">
  <rdfs:subClassOf rdf:resource="http://ontorion.com/unknown.owl/aywal"/>
</Class>
<!-- http://ontorion.com/unknown.owl/abrriđ-n-taydwin -->
- <Class rdf:about="http://ontorion.com/unknown.owl/abrriđ-n-taydwin">
  <rdfs:subClassOf rdf:resource="http://ontorion.com/unknown.owl/aywal"/>
</Class>
<!-- http://ontorion.com/unknown.owl/abulxir -->
- <Class rdf:about="http://ontorion.com/unknown.owl/abulxir">
  <rdfs:subClassOf rdf:resource="http://ontorion.com/unknown.owl/aywal"/>
</Class>
<!-- http://ontorion.com/unknown.owl/adar-n-ufullu -->
- <Class rdf:about="http://ontorion.com/unknown.owl/adar-n-ufullu">
  <rdfs:subClassOf rdf:resource="http://ontorion.com/unknown.owl/ašklu"/>
</Class>
```

Figure 10 : Version RDF

Eléments sur l'implémentation

Eléments de l'ontologie

Ci-dessous une présentation illustrative de cette méthode. Dans un premier temps, nous présentons des relations simples des êtres et dans un deuxième temps des relations complexes :

1. La hiérarchie simple des êtres. On parle ici des relations qui « spécifient les relations verticales de l'ontologie, et se résument à des formes de subsomption. de l'élément par la classe (relation is a) ou de la partie par le tout (relation has a). Ces relations sont déterminantes, car toute ontologie a une fonction unificatrice, l'Être étant défini par son unité à soi » (Rastier 2004).

Exemple :

1. Relations de base

Every *izem* is an *amuder*.

Pour aller plus loin.

Every *aserdun* is an *amuder*.

Every *amuder amsksun* is an *amuder*.

Every *amuder amstuyya* is an *amuder*.

Every *tafunast* is an *amuder amstuyya*

On traite également la flore.

Every *azmmur* is a *tiškt*

Every *tasmumt* is a *tuyya*

Every *tazdayt* is a *tiškt*

Every *bu-hmmu* is a *tuyya*

Every *alili* is a *tiškt*

Every *tamemt n waman* is a *tiškt*

Every *tiši t* is a *tiškt*

2. attributs et specifications

Every *tiškt ittuyatšan* is a *tiškt*

3. Disjointes

Every *amuder* is not a *tiškt*.

Every *tixsi* is not a *aserdun*.

Every *tiškt* is not a *amuder*.

4. Les relations simples

Every *izem da-i-tt-ta* a *tamlalt*.

Every *imiššəw da-i-tt-ta* an *ağərday*.

Every *tafunast* *da-tt-tta* a *tuyya*.

Every *iydi* *da-i-tt-tta* an *aksum*.

5. Les relations complexes

Every *lfssa* is something that *i-tt-tta* an *amuder* and *da-i-tt-tta* *timzin* and *da-tt-tta* a *ag^wlas*.

If X *i-yya* zi something that *i-ttu-yyan* zi Y then X is *-i-yya* zi Y.

6. Equivalence

Something is a *amuder* *amsksun* if-and-only-if-it *da-i-tt-tta* nothing-but *amuder* and-or *da-i-tt-tta* nothing-but thing that *i-yyan* zi an *amuder*.

Something is an *anarmu* if-and-only-if-it *da-i-tt-tta* an *amuder* and *i-tt-tta* an *tiškt* and *i-tt-tta* a thing that *i-yyan* zi an *amuder* and-or *i-yyan-zi* a *tiškt*.

Something is a *amstuyya* if-and-only-if-it *i-tt-tta* nothing-but *tiškt* and-or *i-tt-ttan* nothing-but thing that *i-yyan* zi *tiškt*.

7. Disjointnes

Anything either is an *anarmu*, is an *amstuyya* or is an *amsksun* or-something-else.

8. Relations entre les entités

X *i-yyanzi* Y if-and-only-if Y *i-ttuyya* zi X.

X *da-i-tt-tta* Y if-and-only-if Y *da-i-ttuyatša* zi X.

Spécifications techniques

Pour la mise en place de l'ontologie, nous avons utilisé la plateforme knoodl. C'est une solution de gestion d'ontologies dans le Cloud. Elle est dotée de plusieurs services permettant une gestion collaborative d'ontologies, de gestion de communautés, des wiki, des liens avec d'autres ontologies et données, etc. Ce choix a été motivé, entre autres, par le fait que cette technologie permet de profiter d'une part de la technologie du Cloud ne nécessitant aucune installation ni maintenance de solutions logicielles et matérielles, dans un esprit communautaire pour la création et la maintenance de ce type de contenus. Concrètement, knoodl permet notamment la création, la gestion et l'analyse des données de type RDF/OWL. Cette solution intègre en plus des fonctionnalités avancées comme : l'import/export de contenus de type RDF/OWL, l'interrogation SPARQL, visualisation des sources de données, des vues d'analyse, des visualisations graphiques, etc.

C'est une solution gratuite hébergée dans le cloud d'Amazon EC2.

Conception et peuplement d'une ontologie modélisant la notion de contexte enrichie par les fonctions lexicales pour la détection du sens dans le texte. Parler du Maroc central

Aperçu sur l'ontologie (la faune et la flore en amazighe)

L'ontologie a été conçue en se basant sur les relations conceptuelles et lexico-sémantiques, comme représentée dans le méta-modèle précédent.



Figure 11: Extrait de l'ontologie à partir du Cloud (Knoodl)

C'est une ontologie de domaine sur la faune et la flore en amazighe enrichie par des éléments linguistiques, intégrable dans notre démarche d'analyse et d'extraction sémantiques des données à partir du texte.

Quelques éléments statistiques sur le contenu de l'ontologie

A l'état actuel du développement de l'ontologie, nous avons intégré au format RDF :

- Plus que 1500 concepts,
- Plus que 3000 relations (conceptuelles, ontologiques, sémantiques, etc.),

- Quelques contraintes types.

Cette ontologie est prête à être utilisée dans notre processus d'analyse sémantique des textes. A la fin de son développement, nous envisageons de l'intégrer dans le Cloud du Linked Open Data (LOD).

Conclusion et perspectives

Bien qu'il existe des systèmes capables d'analyser et de traiter des contenus d'un point de vue sémantique, la relation qui lie le contenu à son utilisation est généralement peu ou pas du tout prise en compte. En effet, cette relation peut être d'une extrême complexité qui nécessite des approches et systèmes intelligents difficiles à mettre en œuvre et capable de s'adapter en fonction du contexte d'utilisation. Dans cet article, nous avons présenté la conception d'une ontologie de contexte à partir d'une ontologie de domaine enrichie par la notion de contexte et les relations lexico-sémantiques, intégrable dans une approche de gestion de la sémantique dans son contexte. Ceci, afin d'améliorer les performances de certains domaines d'application comme l'indexation, la recherche ou encore l'extraction d'information. On a pu, à travers quelques exemples de relations et de définition particulière du contexte, répondre aux soucis relatifs à des énoncés ou à des mots auxquels peuvent correspondre plusieurs et différentes structures sémantiques en analysant fidèlement les relations sémantiques que peuvent entretenir les mots simples, voire les expressions figées au sein d'un même texte et qui relève d'un même domaine, ainsi que les règles de raisonnement qui leur sont applicables. Une extension de cette étude à toutes les relations, une étude approfondie des spécificités de chaque relation ainsi que son utilisation constituent les perspectives du présent travail. On appliquera les résultats de ces améliorations pour affiner « la pertinence » et « l'optimisation » dans des domaines et des utilisations particuliers.

Références

- Ahmad K., Fulford H. (1992), "Knowledge Processing: Semantic Relations And Their Use In Elaborating Terminology", *Computing Science Report*, University of Surrey.
- Amardeilh, F et *al.*, (2005), « Annotation documentaire et peuplement d'ontologie à partir d'extractions linguistiques », *Actes de la Conférence Ingénierie des Connaissances (IC'05)*, Nice.
- Bachimont, B., (2001), « Modélisation linguistique et modélisation logique des ontologies : l'apport de l'ontologie formelle », *Actes des journées francophones d'Ingénierie des Connaissances (IC'2001)*, Grenoble.
- Chakiri, M. (2011), « La locution nominale entre opacité et transparence (parler des Aït Wirra, Moyen Atlas, Maroc) », *Etudes et Documents Berbères*, p. 97-108.
- Chakiri, M. (2010), « Analyse stylistique des locutions nominales en amazighe », *Revue Asinag*, IRCAM, Maroc, p. 201-210.

- Condamines A., Rebeyrolle J. (1998), "Ctkb : A Corpus-Based Approach For Terminological Knowledge Base ». in *Proceedings Of The First Workshop On Computational Terminology (COMPUTERM'98), Workshop of Coling'98*, Montréal.
- Cunningham, H. et al., (2002), « Framework and Graphical Development Environment for Robust NLP Tools and Applications». *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'2002)*, Philadelphia.
- Fadili, H., (2013), « Towards a new approach of an automatic and contextual detection of meaning in text », AICCSA'2013, Fès/Ifrane.
- Gómez-Pérez A. (1999), *Ontological Engineering: A State Of The Art*, Expert Update.
- Guarino N. (1997), "Understanding, Building And Using Ontologies". *International j. Human-computer studies*.
- Guerin, É., (2003), « Un exemple d'article dans les actes d'une conférence », *Actes de la conférence CORESA'03*, Lyon.
- Hernandez N., (2005), *Ontologies de domaine pour la modélisation du contexte en Recherche d'information*, Thèse de doctorat, Université Paul Sabatier de Toulouse.
- L'homme Marie-Claude (2004), *La terminologie : principes et techniques*, Montréal, les presses de l'Université de Montréal.
- Mel'čuk, I. (1981), "Meaning-Text Models: A Recent Trend In Soviet Linguistics", *Annual Review Of Anthropology*.
- Mel'čuk, I. (1995), *The Russian Language In The Meaning-Text Perspective*, Wiener Slawistischer Almanach/Škola "Jazyki Russkoj Kul'tury": Vienna/Moscow.
- Mel'čuk, I. (1997), *Vers une linguistique sens-texte. leçon inaugurale*, Paris: Collège de France.
- Mel'čuk, I. (1998), "Dependency Syntax: Theory And Practice", Albany, N.Y.: The SUNY Press.
- Michael, B., (1998). *Fractals everywhere*. Academic Press.
- Mizoguchi R. And Ikeda M. (1996), *Towards Ontological Engineering (Ai-Tr-96-1)*,. Osaka: Isir, Osaka University.
- Mizoguchi R (1998), "A Step Towards Ontological Engineering", paper presented at the 12th National Conference On Ai Of JSAI.
- Polguere, A., (2003), *Lexicologie et sémantique lexicale. notions fondamentales (paramètres)*, Montréal, PUM.
- Rastier, F. (2003), *De la signification au sens. Pour une sémiotique sans ontologie*. [En ligne]. Disponible sur : http://www.revue-texto.net/Inedits/Rastier/Rastier_Semiotique-ontologie.html.

Rastier, F. (2004), « Ontologie(s). Revue des sciences et technologies de l'information », Série : *revue d'intelligence artificielle*, 4.

Sager, J. (1990), *A Practical Course In Terminology Processing*, Philadelphia : John Benjamins publishing company, Amsterdam.

Taifi, M. (1991), *Dictionnaire tamazight-français* (parler du Maroc central), Paris, L'Harmattan-Awal.

Van Heijst G., Schreiber A. And Wielinga B. J. (1997), "Using Explicit Ontologies In Kbs Development". *International Journal Of Human And Computer Studies /Knowledge Acquisition*.

Vanwelkenhuysen J. And Mizoguchi R. (1995), "Workplace-Adapted Behaviors: Lessons Learned For Knowledge Reuse", *Paper presented at the KB&KS '95*.

Vanwelkenhuysen J. And Mizoguchi R.(1994), "Maintaining The Workplace Context In A Knowledge Level Analysis", *Paper presented at the Proc. of JKAW'94*, Hatoyama, Japan.

Wielinga B. And Schreiber A. (1993), "Reusable and Sharable Knowledge Bases: A European Perspective", *Paper presented at the KB & KS'93*, Tokyo.

The GATE platform: <http://gate.ac.uk/>, mars 2013.