

DIRO
Université de Montréal

Encadrant : Philippe LANGLAIS

**Extraction de relations d'hypéronymies et de
synonymies sur Wikipédia Français**
Rapport Final

Adeline GRANET

le 14 septembre 2013

Table des matières

1	Introduction	3
1.1	Etat de l'art dans d'autres langues	3
1.2	Pré-étude	4
2	Extraction par hiérarchie	6
2.1	Méthode d'extraction	6
2.2	Format des résultats	7
2.3	Analyse des premiers résultats	7
3	Extraction des relations par patrons	9
3.1	Présentation des patrons	9
3.2	Question de choix	9
3.2.1	Phase de rejet de certains patrons	10
3.2.2	Phase de recherche d'une syntaxe	11
3.2.3	Phase de produit cartésien	11
4	Résultats	12

1 Introduction

Le but de cette étude est de chercher automatiquement des relations d'hypéronymies et de synonymies sur un corpus aussi volumineux que celui de Wikipédia français.

Définition du concept d'hypéronymie L'hypéronymie désigne une relation d'inclusion entre deux lexèmes. Lorsque l'on dit d'un terme qu'il est dans une relation d'hypéronymie avec un autre, cela signifie que son sens est plus spécifique que l'autre, par exemple : "*lapin*" est un *hyponyme* d'"*animal*". (source : Wikipédia)

1.1 Etat de l'art dans d'autres langues

L'extraction automatique de relation d'hypéronymie a déjà été réalisée dans deux langues : japonais et anglais. Dans les différentes publications qui ont été réalisées [1] [2] [3], il est possible de se faire une bonne idée sur la méthode qui a été mise en place pour cela. Un résumé d'une des méthodes les plus intéressantes est présenté ci-dessous.

Méthode de Sumida et Torizawa Pour réaliser l'extraction des relations d'hypéronymies, ils utilisent la hiérarchie des titres et sous-titres de sections qui structurent les articles de Wikipédia. Leur méthode permet de collecter 6.3×10^5 relations en 3 étapes.

La structure hiérarchique sur laquelle ils se sont basée repose sur les différentes sections qui composent un article, qui elles-mêmes comportent des sous-sections. Dans le code source des pages, cette hiérarchie se lit facilement grâce à des balises de type `=*` où le nombre de `=` indique la profondeur dans le plan. Ils se sont proposés d'utiliser ces dépendances de section et sous-section pour automatiser l'extraction des relations d'hypéronymies. Le même système est utilisé avec des `*` pour créer des listes où le nombre d'astérisques permet de se placer dans la hiérarchie des sections. La figure 1 est un exemple d'article de wikipédia où l'on peut voir comment se présente la hiérarchie des sections dans un article. Le chiffre en début de ligne est un indicateur qui retourne le niveau où l'on se situe dans l'arborescence.

```

title Amigny

CATE : Commune de la Manche
2 == Géographie ==
2 == Toponymie et microtoponymie ==
2 == Toponymie ==
2 == Histoire ==
2 == Administration ==
2 == Démographie ==
2 == Lieux et monuments ==
3 * Église Notre-Dame, romane, remaniée au {{XVIIIe siècle}}.
3 * Château du {{XVIIe s}}.
2 == Activité et manifestations ==
2 == Personnalités liées à la commune ==
2 == Activité et manifestations ==
* [[Liste des communes de la Manche]]

```

FIGURE 1 – Exemple d’une structure hiérarchique d’un article de Wikipedia

Une méthode en trois étapes La méthode proposée pour extraire les relations hypéronymiques selon la hiérarchie des sections, se compose de trois étapes.

- La première consiste à extraire, les relations en utilisant les titres des différentes sections. Si l’on reprend l’exemple précédent, le titre va être couplé à tous ses subordonnés directs (niveau 2 des sections) pour créer des relations candidates comme "*Amigny/Géographie*", "*Amigny/Toponymie et microtoponymie*". Evidemment, des relations erronées sont aussi extraites. Les étapes 2 et 3 sélectionnent les meilleurs relations.
- La seconde étape sélectionne les relations d’hypéronymies par application de simples patrons. En combinant les résultats de la première étape avec celle-ci, il est possible de supprimer les mots en trop dans les relations obtenues à cette étape par comparaison avec les relations obtenues à la première étape.
- Pour finir, ils appliquent un algorithme d’apprentissage machine sur les relations qui ont été recalées à l’étape précédente. Pour chaque relation candidate, ils appliquent une analyse morphologique et ainsi obtiennent les caractéristiques de chaque hypéronyme et hyponyme candidat ; pour ensuite les placer dans un vecteur qui sera donné au classifieur.

L’article de Sumida et Torizawa décrit relativement bien la méthode, et au vu des résultats obtenus, il semble intéressant d’appliquer celle-ci sur la version française de Wikipédia.

1.2 Pré-étude

Une ouverture en français A notre connaissance, cette étude n’a encore jamais été réalisée en français. Avec l’article qui a été résumé précédemment, une ligne de conduite a pu être établie quant à la démarche à adopter. Malheureusement, le temps imparti à l’implémentation de cette méthode, ne laisse la possibilité que de réaliser seulement les deux premières étapes et de faire leur intersection. En effet, la seconde étape demandent beaucoup de temps.

Les outils à disposition Il existe une API WikipediaMiner (« wikipedia-miner.cms.waikata.ac.nz ») qui permet d'extraire un certain nombre de renseignements sur les différents articles formulés dans Wikipedia, par exemple : le titre, les liens entrants et sortants, le premier paragraphe ou bien le texte en entier... Dans notre étude, cela va être très pratique car il va être possible de récupérer facilement la page au complet avec le titre, la hiérarchie ainsi que le texte facilement. D'un point de vue performance, il a été possible de réaliser un test avec le dump en anglais : pour 5000 articles traités, il a fallu moins de 30 secondes pour créer 65 fichiers avec l'ensemble des articles à raison de 100 articles par fichier texte. Un premier traitement est déjà réalisé au cours de cette extraction afin d'enlever les articles vides comme les pages "Redirects".

2 Extraction par hiérarchie

Il est important de rappeler que cette étape consiste à extraire une base solide de relations en se basant sur la hiérarchie qui constitue les articles Wikipédia.

2.1 Méthode d'extraction

Le fichier obtenu grâce à l'API WikipédiaMinner, contient le texte source précédé du titre de l'article ainsi que les catégories (Wikipédia) auquel il était directement relié. Le fichier texte étant brut, c'est-à-dire contenant tout le texte, la hiérarchie ainsi que de nombreuses balises, celui-ci a nécessité un premier traitement qui a consisté à isoler le titre de l'article avec les sections, et sous-sections qui lui était associées (tout en gardant la partie sur les catégories associées intacte).

La méthode pour cette partie est restée relativement simple : pour chaque article, une relation a été créée entre une section et sa(ses) sous-sections directe(s). Par exemple, le titre de l'article a une relation d'hypéronymie directement avec les titres de section de niveau 1 (c'est-à-dire =section=). Des titres de sections qui eux-mêmes entretiennent une relation avec leurs sous-sections directes de niveau 2 (donc commençant par ==).

La même méthode a été appliquée dans le cas d'une liste, où chaque puce de la liste est associée directement au titre de la section dans laquelle elle se trouve.

Dans le cas des catégories, une relation correspond au titre de l'article avec le titre de la catégorie.

La figure 2 montre la hiérarchie d'un articles avec un ensemble de section, sous-sections et de puces. La figure 3 montre les relations candidates qui ont pu être réalisées à partir de la hiérarchie.

```

title Tokyo Skytree

CATE : Bâtiment de Tokyo
CATE : Bâtiment de Tadao Andō
CATE : Tour d'observation
CATE : Fondation en 2012
CATE : Wikipédia:Outil de retour des lecteurs

2 == Structure ==
2 == Financement et fonctionnement ==
2 == Dénomination ==
3 * {{japonais|Tokyo Edo Tower|東京EDOタワー|Tokyo Edo Tawā}}
3 * {{japonais|Tokyo Skytree|東京スカイツリー|Tokyo Sukaitsurī}}
3 * {{japonais|Mirai Tower|みらいタワー|Mirai Tawā}}
3 * {{japonais|Yumemi Yagura|ゆめみやぐら}}
3 * {{japonais|Rising East Tower|ライジングイーストタワー|Raijingu Īsuto Tawā}}
3 * {{japonais|Rising Tower|ライジングタワー|Raijingu Tawā}}.
2 == Évolution du chantier ==
2 == Notes et références ==
2 == Liens externes ==

```

FIGURE 2 – Extraction de la hiérarchie des sections

```

CATE | Bâtiment de Tokyo | Tokyo Skytree
CATE | Bâtiment de Tadao Ando | Tokyo Skytree
CATE | Tour d'observation | Tokyo Skytree
CATE | Fondation en 2012 | Tokyo Skytree
CATE | Wikipédia:Outil de retour des lecteurs | Tokyo Skytree
HISTO | Tokyo Skytree | Structure
HISTO | Tokyo Skytree | Financement et fonctionnement
HISTO | Tokyo Skytree | Dénomination
HISTO | Tokyo Skytree | Évolution du chantier
HISTO | Tokyo Skytree | Notes et références
HISTO | Tokyo Skytree | Liens externes
PUCE | Dénomination | Tokyo Edo Tower
PUCE | Dénomination | Tokyo Skytree
PUCE | Dénomination | Mirai Tower
PUCE | Dénomination | Yumemi Yagura
PUCE | Dénomination | Rising East Tower
PUCE | Dénomination | Rising Tower

```

FIGURE 3 – Relations candidates hyperonymiques

2.2 Format des résultats

Pour pouvoir savoir comment ont été extraits les relations, il y a un mot-clé en début de chaque ligne (où une ligne correspond à une relation candidate) :

- CATE : pour les relations basées sur les catégories associées à chaque article, un exemple : CATE | Fondation en 2012 | Tokyo Skytree ;
- HISTO : pour les relations liées à la hiérarchie des sections de l'article, un exemple : HISTO | Tokyo Skytree | Dénomination ;
- PUCE : pour les relations utilisant les listes contenues dans les articles, un exemple : PUCE | Dénomination | Tokyo Skytree.

2.3 Analyse des premiers résultats

L'observation des relations extraites nous a permis de tirer des premières conclusions sur la forme de Wikipédia Français.

Pour chaque type de relation, les hypéronymes les plus fréquents ont été cherchés ; puis pour quatre mots (choisis aléatoirement), leurs hyponymes respectifs ont été cherchés. La première recherche nous a permis de mettre en évidence que les thèmes les plus récurrents dans Wikipédia fr sont la géographie et les personnes (voir les résultats présentés dans le tableau 1).

HISTO	PUCE	CATE
Biographie (95208) Histoire (57847)	Distribution (103331) Biographie (114046)	Anglais langue officielle (64) Résolution 109 du conseil de sécurité des nations unies (62)
Géographie (51367) Palmarès (35673) Discographie (25130)	Articles connexes (121608) Listes des espèces (130867) Palmarès (152914)	Jacques Chaban-Delmas (52) Jacques Chirac (49) Michel Debré (45)

TABLE 1: Hypéronymes les plus vus pour chaque type

(Le chiffre entre parenthèse après chaque hypéronymes correspond au nombre d'occurrences trouvées dans le fichier résultat.)

Le tableau 2 présente les résultats obtenus pour les mots Vin, Calendrier, Saison et Population. Il apparaît que les relations données par le système de puce n'est rien d'autre que du bruit. De même pour les relations obtenues grâce à la hiérarchie, il est évident que les relations ne sont pas de l'ordre hypéronymique. Donc pour la dernière phase, nous avons choisi de ne garder que les relations liées aux catégories.

	HISTO	PUCE	CATE
Vin(s)	Vinification et élevage (7) Commercialisation(6) Gastronomie, durée de garde et température de service (6) Titre alcoométrique volumique (6) Liste des producteurs (5)	1er service (1) 2e service (1)	Vin Boisson alcoolisée Wikipédia : outil de retour des lecteurs
Saison	Avant saison (4) Résumé de la saison (4) Autres (1) Coefficient FIFA (1) Euro 2012 (1)	/	Calendrier Mécanique céleste Unité de temps Saison Climatologie
Calendrier	Articles connexes (1) Bibliographie (1) Combinaison des unités naturelles (1) Date des dernières et prochaines éditions (1) Février (1)	Nouveau menu "calendrier" dans les préférences (1) Possibilité d'avoir des calendriers multiples en couleurs (1) Push calendrier (1)	Calendrier Instrument de mesure du temps Wikipédia : outils de retour des lecteurs
Population	Ethnies et minorités étrangères (95) Evolution de la population en janvier de chaque année (94) Démographie (3) Articles connexes (2)	Femme : 69.171 (1) Homme : 68.115 (1) Précis : 137.286 (1) Rur : 6 950 (1) urbain : 130.366 (1)	

TABLE 2: Exemple de relations obtenus d'hypéronymie

WikipédiaMiner a permis de traiter 1 278 381 articles. A la fin de cette phase, c'est à peu près 4 203 146 relations qui ont été collectées.

3 Extraction des relations par patrons

Maintenant qu'une première base de relations candidates d'hypéronymies a été extraite, des patrons vont être appliqués sur les premiers paragraphes de chaque article. Pour cette phase, nous avons décidé de ne pas nous contenter que d'un seul type de relation mais de l'ouvrir à l'extraction de synonymie et d'antonymie.

3.1 Présentation des patrons

Après une recherche rapide, il est apparu qu'il n'existait pas de base libre contenant des patrons qui puissent permettre l'extraction d'hypéronymes. L'idée a été d'utiliser les différents articles déjà écrits sur le sujet et d'exploiter les patrons cités (après avoir été traduits selon les cas). Certains patrons comme *"est un" / "sont des"* restent très ambigus. En effet dans le domaine de l'acquisition automatique d'hypéronymie, ils sont reconnus comme étant un des patrons fondamentaux et restent les plus cités dans les articles (Marti A. Hearst [4]; Patrick Pantel, Deepak Ravichandran et Eduard Hovy [3], entre autres); pourtant certains linguistes estiment que ce patron peut également être reconnu pour la recherche de synonymie. Les autres patrons sélectionnés proviennent d'articles anglais qui traitent, pour la majeure partie, de l'extraction automatique de relations taxinomiques et sémantiques. Après lecture d'une dizaine d'articles traitant de ce sujet, il apparaît que les patrons utilisés sont toujours les mêmes (cela est vrai pour les articles qui travaillent sur des corpus anglais comme japonais ou encore français).

Une première liste a pu être établie grâce à cette étude :

- "par exemple";
- "une sorte de";
- "de la famille des";
- "un principe de";
- "tel que";
- "par N comme L" (où le N représente un nom, et L une liste de terme qui lui sont associés);
- "L ou autre N";
- "L et autre N";
- "N y compris L";
- "N spécialement L";
- "un genre de";
- "est un(e) / sont des".

Pour la relation de synonymie, voici les patrons retenus à travers différents articles français-anglais confondus :

- soit;
- il s'agit de;
- ou;
- comme;
- autrement dit;
- au sens de;
- à savoir;
- qui signifie.

Pour finir, il n'y a qu'un seul et unique patron pour l'antonymie qui est "contrairement à / aux".

3.2 Question de choix

Pour se faire une idée de la capacité des patrons sélectionnés, ils ont été appliqués sur le corpus à différentes reprises (sur le texte brut, sur le texte nettoyé) et ainsi permettre de faire un choix sur ceux qui semblent être les plus efficaces.

Les tableaux 3 et 4 présentent le nombre d'occurrence pour chaque patron à quatre moments différents. Les deux premières colonnes correspondent aux tests avec le texte brut et celui nettoyé, les deux suivants ont été réalisés après avoir rejeté une partie des patrons.

Voici un exemple lors des quatre phases :

- Observation sur le texte brut : "La japonais"Tokyo Skytree"Tokyo Sukaitur est une tour de [[radiodiffusion]] dans le quartier [[Sumida-ku]] à [[Tokyo]]. Mesurant unité634|mètres, elle est depuis son inauguration la deuxième plus haute structure du monde encore debout, derrière le [[gratte-ciel]] [[Burj Khalifa]]."
- Observation sur le texte nettoyé : "La "Tokyo Skytree" est une tour de radiodiffusion dans le quartier Sumida-ku à Tokyo. Mesurant 634 mètres, elle est depuis son inauguration la deuxième plus haute structure du monde encore debout, derrière le gratte-ciel Burj Khalifa."
- Observation avec une syntaxe recherchée : "La "Tokyo Skytree" est une tour de radiodiffusion dans le quartier Sumida-ku à Tokyo. Mesurant 634 mètres, elle est depuis son inauguration la deuxième plus haute structure du monde encore debout, derrière le gratte-ciel Burj Khalifa. "

Patron	Fichier brut	Fichier nettoyé	Recherche syntaxe	Produit cartésien
par exemple	4 449	4 338	/	/
une sorte de	899	890	/	/
de la famille	16 879	38 841	37 791	81 956
un principe de	125	136	103	87
tel que	953	927	/	/
par N comme L	/	/	/	/
ou autre	481	479	/	/
et autre	1 639	1 599	/	/
y compris	1 279	1 248	183	280
spécialement	967	950	/	/
un genre de	1 579	1 587	1 438	2 883
est un(e) / sont des	900 611	875 599	171 488	530 405

TABLE 3: Extraction possible d'hypéronymie par patron

Patron	Fichier brut	Fichier nettoyé	Recherche syntaxe	Produit cartésien
Soit	11 628	11 421	/	/
Il s'agit	806	795	/	/
Ou	822 881	787 866	82 120	147 220
Comme	87 061	83 832	/	/
Autrement dit	254	251	/	/
Au sens de	348	344	57	38
A savoir	718	697	/	/
Qui signifie	4 046	4 009	/	/

TABLE 4: Extraction possible de synonymie par patron

3.2.1 Phase de rejet de certains patrons

Le problème qui a été rencontré avec un certain nombre de patrons, est l'incapacité à détecter où se situe l'hypéronyme et l'hyponyme, les deux synonymes ou encore antonymes. En effet, il arrive souvent que l'un des deux se situe dans une autre phrase connexe mais sans étiquetage au préalable, il est impossible de sélectionner le(s) bon(s) terme(s).

En voici quelques exemples :

- "Les doigts des Tétrapodes sont initialement palmés, la région interdigitale est détruite par apoptose (sauf dans les pattes postérieures des canards ou des grenouilles **par exemple**."
- "Le mot néerlandais "hofje", diminutif de "hof", 'cour', désigne un ensemble Architecture, architectural composé de maisonnettes disposées en carré autour d'une cour intérieure ou d'un patio, formant ainsi **une sorte de monde clos**."
- "Le "Sikorsky Aircraft Corporation, Sikorsky S-61R" est un hélicoptère conçu et réalisé aux USA, États-Unis à partir du célèbre Sikorsky S-61, Sea King. Il est optimisé **spécialement** pour les missions de transport aérien, de sauvetage en mer, et de soutien aux forces spéciales."

Dans le cas où nous sommes en présence d'une liste, il est difficile de savoir où commence la liste réellement. Voici un exemple :

"Cette présence dans la conscience populaire est attestée par les références à ce personnage dans les musique folklorique, chansons folkloriques, les Folklore, contes **et autres** formes d'expression Folklore, folkloriques ou populaires, comme la littérature et le cinéma." Les exemples choisis sont uniquement dans les cas où il y a la présence d'une relation mais il existe beaucoup de cas qui ne sont que du bruit. Il est important de prendre en compte le fait que tous les patrons listés précédemment, sont appliqués sur des corpus annotés ce qui facilite la location des termes contrairement à notre étude. C'est pourquoi la majorité des patrons ont du être recalés.

Acceptation de certains patrons Après avoir analysé rapidement le fichier contenant le premier paragraphe, il est apparu évident que les patrons "est un(e)" et "sont des" étaient des patrons intéressants puisque chaque paragraphe (à quelques exceptions près) commencent par une définition qui contient un de ces deux patrons. Après applications du patron, on constate qu'environ 890 000 relations peuvent être extraites, ce qui est non négligeable.

3.2.2 Phase de recherche d'une syntaxe

Voici un exemple de paragraphe pour un article

"La "Tokyo Skytree" est une tour de radiodiffusion dans le quartier Sumida-ku à Tokyo. Mesurant 634 mètres, elle est depuis son inauguration la deuxième plus haute structure du monde encore debout, derrière le gratte-ciel Burj Khalifa. "

Volontairement, les balises qui indiquent le nom/sujet de l'article ont été conservées et vont servir pour délimiter la sélection autour du patron. Pour appliquer chaque patron, il a été recherché < "titre de l'article"> suivit du patron jusqu'à la rencontre d'un signe de ponctuation (à l'aide de commande grep, awk et sed) ; puis sauvegardé dans un fichier sous la forme

" type de la relation | patron utilisé | hypéronyme | hyponyme".

3.2.3 Phase de produit cartésien

Après observation des premiers résultats, il est apparu que souvent l'hypéronyme, l'hyponyme ou bien même les deux étaient constitués de plusieurs termes séparés par une virgule ou un "ou". Donc pour ne pas perdre d'information et dans le but d'améliorer la précision des relations obtenues, le choix de réaliser le produit cartésien entre les différentes composantes s'est avéré le plus prudent. Par exemple, la ligne

"HYP |est un| mathématicien, ingénieur et architecte Empire byzantin, byzantin ((- † avant 558).|Anthémius de Aydn, Tralles",

devient

HYP |est un| mathématicien|Anthémius de Aydn

HYP |est un| ingénieur|Anthémius de Aydn

HYP |est un|architecte Empire byzantin|Anthémius de Aydin

HYP |est un|byzantin ((- † avant 558).|Anthémius de Aydin

Ainsi ce n'est plus une mais quatre nouvelles relations que l'on aura pour le même hyponyme. Avec cette méthode, le nombre de relations est multiplié, et cela permet également de dissocier le bruit du reste de la relation.

Au final, pour cette seconde étape d'extraction par patron c'est 645 063 relations qui ont pu être extraites (32 535 synonymies et 612 093 hypéronymies).

4 Résultats

L'évaluation a du se faire manuellement par manque de base de données comparatives libres. Les données ont été modélisées dans un format qui facilite la validation manuelle.

Pour évaluer la qualité des relations obtenues, un fichier a été créé en associant un hyponyme à tous ses hyponymes dans une liste. Les listes de relations qui ont été créées sont du type : *nom de la catégorie = [nom des articles associés]*.

```
Médecin du XIXe siècle = [Abraham Jacobi| Adamántios Korais| Adam Kuhn|
Adolf Kussmaul| Adolf Strümpell| Adolf Wallenberg|
Adolphe de Barrau| Adolphe Leray| Adolpho Lutz|
Adolph von zur Lippe| Adrien Pozzi| Adrien Proust|
Aimé-François Legendre| Albert Adamkiewicz|...]
```

FIGURE 4 – Liste hypéronymique pour "Médecin du XIXe siècle"

Sur la figure 4, un exemple de liste que l'on peut trouver. L'hypéronyme dans ce cas est "Médecin du XIXe siècle" et les noms qui suivent sont les hyponymes trouvés. Une sélection aléatoire de 200 et 500 relations hypéronymiques a été réalisée pour pouvoir calculer le rappel et la précision sur les relations de la phase d'extraction via la hiérarchie des sections des articles et celles extraites par les patrons.

	% de relation hypéronymique	Précision (vrai relation)	Rappel
Phase hiérarchie	55,94%	91,87%	/
Phase patron	21,8%	96,23%	80,61%

TABLE 5: Résultat sur les relations extraites sur les trois phases

Pour ces relations extraites grâce à la hiérarchie des articles, le rappel n'a pas été calculé car seules les catégories associées à chaque article ont été conservées, or les articles sont automatiquement référencés sur les pages des catégories. Donc, le rappel aurait été de 100% puisque le lien entre articles et catégories est symétrique.

Un constat qui a pu être fait c'est qu'une majeure partie des hypéronymes trouvés, lors de la phase par extraction par patrons, sont des titres de catégorie. Et que les hyponymes se trouvant dans le vecteur associé ne sont autre que les titres des articles qui sont associés à ses catégories. C'est cela qui a permis de calculer le rappel dans certain cas. En effet, dès que l'hypéronyme était une catégorie, il suffisait de comparer le nombre d'hyponymes associés et le nombre de page associées dans Wikipedia. Le taux de rappel et de précision étant relativement haut, cela laisse penser que les patrons utilisés étaient corrects et judicieux

même si le nombre de relations finales est relativement faible par rapport à nos espérances sachant que l'on trouve moitié moins de relations que sur la base japonaise (voir [1]).

Liste des tableaux

1	Hypéronymes les plus vus pour chaque type	7
2	Exemple de relations obtenus d'hypéronymie	8
3	Extraction possible d'hypéronymie par patron	10
4	Extraction possible de synonymie par patron	10
5	Résultat sur les relations extraites sur les trois phases	12

Références

- [1] Asuka Sumida and Kentaro Torisawa. Hacking wikipedia for hyponymy relation acquisition. In *In Proceedings of IJCNLP 2008*, 2008.
- [2] Véronique Malaisé, Pierre Zweigenbaum, and Bruno Bachimont. Detecting semantic relations between terms in definitions. In *COLING 2004 CompuTerm 2004 : 3rd International Workshop on Computational Terminology*, pages 55–62. COLING, August 29 2004.
- [3] P. Pantel, D. Ravichandran, and E. H. Hovy. Towards terascale knowledge acquisition. In *Proceedings of COLING-04*, pages 771–777, 2004.
- [4] Marti A. Hearst. Automatic acquisition of hyponyms from large text corpora. In *In Proceedings of the 14th International Conference on Computational Linguistics*, pages 539–545, 1992.
- [5] LE Thi Hoang Diem and Jean Pierre Chevallet. Extraction et structuration des relations multi-types à partir de texte. 2006.
- [6] Erik Tjong Kim Sang and Katja Hofmann. Lexical patterns or dependency patterns : which is better for hypernym extraction? In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, pages 174–182, 2009.
- [7] Jean Fargues. Synthesizing a large concept hierarchy from french hyperonyms. In *COLING*, pages 112–117, 1990.