Semantic resource extraction from Wikipedia category lattice

Olivier Collin, Benoît Gaillard, Jean-Léon Bouraoui
Orange Labs
Avenue Pierre Marzin, 22300 Lannion

Thomas Girault
Université de Rennes I
2, rue du Thabor 35065 Rennes

olivier.collin@orange-ftgroup.com,
benoit.gaillard@orange-ftgroup.com,
jeanleon.bouraoui@orange-ftgroup.com
toma.girault@gmail.com

This work is closely related to the domain of automatic acquisition of semantic resources exploiting Wikipedia data. More precisely, we exploit the graph of parent categories linked to each Wikipedia page to perform a hierarchical parent categories extraction, semantically and thematically related. This extraction is the result of a shortest path length computation applied to the global lattice of Wikipedia categories. So, each page can be indexed by its first level categories, and in addition within their parent categories. This resource has been used for two kinds of applications. The first one concerns semantic query expansion for a multimedia search engine. The second one is a query translator for a multimedia search engine. This last work has been performed by using English lattice of categories and Wikipedia translation tables.

Introduction

This work is closely related to the very large area of lexical semantic resources constitution. The aim is to provide featuring labels for each lexical entry, these labels being hierarchically organised within a taxonomy or a lattice. This representation should lead to a generalisation of the input lexical space (hyperonyms), but also to homonym differentiation and synonym clustering. This point of view is usually shared by the linguistic community which targets a precise and exhaustive modelling of lexical entries. An alternative representation is a vector space approach. Lexical entries are modelled by a vector of neighbour words counts, these neighbours being extracted in the same document within a local window or not, and by using linguistic processing or not. This way, standard vectorial or statistical techniques can project each entry on a "semantic distributed subspace" (Latent Semantic Analysis for example) or within clusters of semantically related entries (K-means for example).

A similarity measure is usually associated to these representations so that one can express a kind of proximity between lexical entries. Then, this measure enables semantic expansion treatment (semantic proximity) or ambiguity resolution (semantic differentiation). This paper proposes an alternative representation space: a subset of the Wikipedia category lattice.

Since Wikipedia creation, many authors (Medelyan and al., 2008), (Suchanek and al., 2008), (Mihalcea, 2007), (Ponzetto and al., 2007), (Zesch and al., 2007), (Strube and al, 2006), have explored means of exploiting Wikipedia data to make a usable semantic resource. This paper is closely related to these preceding works: we automatically extract, without any human help, a sub-lattice from Wikipedia category lattice. The sub-lattice structure is linguistically imperfect but shows, in many cases, relevant hyperonymic relations and thematic categories. We have evaluated this resource relevancy on two use-cases: a semantic expansion task and a query translation task where obviously disambiguation is required.

1 Resource extraction

1.1 The Wikipedia categories lattice

Each Wikipedia page is indexed by a set of visible parent categories, which can be clicked in at the bottom of each page. So, parent french categories for "*Tom Cruise*" page are: *Acteur américain, Producteur américain, Naissance dans l'État de New York, Naissance en 1962, Personnalité américaine d'origine allemande, Personnalité américaine d'origine britannique, Personnalité américaine d'origine irlandaise, Scientologie.* These categories usually express one or more semantic roles. Each of these bottom page categories is also a Wikipedia page which has parent categories. This hierarchy of categories is not a strictly build taxonomy since categories and hierarchical links are freely added by various contributors. These contributions constitute a part of Wikipedia richness as a kind of *folksonomy*,, however semantic relations are difficult to extract from this constellation (Guégan 2006), (Strube 2006). This space of categories rather constitutes a graph oriented towards a set of a unique parent category "*Article*", so, it rather constitutes a lattice. This lattice is linguistically organised, usually each parent category is generalising each child category following a hyperonymic or thematic axis. For the moment, we can't separate these two axes. In addition, for each category, several generalisations act in parallel. For example, in the case of "*Tom Cruise*" parent category *Acteur_américain*, a generalisation direction is: *Artiste_américain>Art_aux_Etats-Unis>Art_par_pays> Art>Article.* (son > parent).

1.2 Lattice fabrication

Raw data, page to category links and category to category links, come from two SQL tables[1]downloaded from the Wikipedia french resource site[2]. Suitable joints on these tables allow a direct relation between each Wikipedia page or category and its parents. This is a flat representation of the overall lattice which virtually enables us to list all the paths between articles and the terminal category "*article*". The page/category links are straight-forward, the combinatorial part of the lattice is mainly related to the category/category links sub-lattice. For computational purpose, we have separated page/category links from category/category links. Finally, for French data, we get 873 468 pages linked to 3 770 343 parent categories (4.31 category per page in average), and 119 492 categories linked by 244 817 edges (2.04 parent

[1] *Frwiki-latest-page.sql and frwiki-categorylinks.sql*
[2] http://download.wikimedia.org/frwiki/latest/.

category per category on average). Our goal was to expand the first level of parent category pages like *Acteur américain* or *Producteur américain* for "*Tom Cruise*" example. So we only processed the category/category sub-lattice.

However, even if the connectivity of this graph is not so high and shows "small world" properties (Guegan, 2006), the quantity of such paths is too great (dozens or even hundreds of paths for each category) to be used without pre-processing. In addition, the flat representation of the lattice doesn't match with our navigational needs. So, we have used NetworkX package[3], which allows us to upload the flat representation tables up to memory. This package also provides many useful functions for a quick navigation in a graph. The overall lattice of categories (119 492 nodes and 244 817 edges) has been uploaded in memory and we have been able to test several quick search algorithms

So, the main challenge of our work has consisted in making a relevant selection among all paths for providing useful semantic data, especially for disambiguation purpose.

1.3 Sub-lattice extraction

The sub-lattice extraction is based on a strong assumption: the relevant information is carried by the shortest paths that link each of the pages to the terminal category. In fact, after some testing we realized that paths linking to the "*Article*"category were less relevant than the paths reaching the set of categories pointed to by the top level category page *Wikipedia:Catégorie*[4]. This set contains 150 pseudo terminal categories such as (in French): *"Mouvement culturel", "Art contemporain", "Artisanat","Design", "Art par pays", "Rayonnement culturel", "Artiste"…* Given the overall lattice and a shortest path calculus provided by NetworkX package, the filtering process is :

```
For each page
    For each parent category page
        Select the shortest paths that reach
            all terminal categories
```

We kept all the shortest paths of the same length which can occur for one category. For a same page, we don't keep paths of length greater than 8 and we only keep the 15 shortest paths. We have filtered a few initial categories like date or place of birth brought noisy paths. Here are results obtained for the two different French pages related to *"Avocat"* (fruit versus occupation):

Avocat_(fruit) (only one path)
1-Fruit_alimentaire>Plante_alimentaire>Plante_utile>Agriculture

Avocat_(métier)(two paths)
1-Métier_du_droit>Droit
2-Personnalité_du_droit>Droit

We observe paths that reach a global thematic category through hyperonymic relations. In this example, filtered data provides quite a good result from a linguistic point of view and shows obvious disambiguation possibilities. In other cases we get more noisy paths, but they are still relevant. The following paths form a sub-lattice related to *"Tom Cruise"* page:

acteur_américain>acteur_par_nationalité>acteur>personnalité>médias
acteur_américain>artiste_américain>art_aux_tats-Unis>art_par_pays>art
acteur_américain>artiste_américain>artiste_par_pays>artiste>personnalité>art
producteur_américain>cinéma_américain>cinéma_aux_états-Unis>cinéma_par_pays>art_par_pays>art
producteur_américain>producteur_de_cinéma_par_nationalité>producteur_de_cinéma>producteur>artiste>personnalité>art
scientologie>groupement_spirituel>spiritualité_autres>spiritualité
scientologie>groupement_spirituel>petit_mouvement_religieux>religion>spiritualité
portail:cinéma>portail:art>portail:culture
portail:états-unis>portail:Amérique>portail:géographie>portail_du_domaine_géographique

The sub-lattice attached to each page is not always very large, especially for French pages and these data are not well formed on a strict linguistic point of view. However, they give a good trade-off between quantity and quality of features. For application purpose, the hierarchy can be broken and these data can be used as a useful "bag of categories". They can help us to treat a disambiguation task by using classical constraints based on hyperonyms and thematic classes: *fruit, agriculture / personnalité, droit*.

A similar job has been done for English Wikipedia pages and categories[5]. The lattice of English categories is bigger than the French one (524 313 nodes and 1 206 219 edges) but NetworkX allows us to get a memory mapping and navigational functions still work quite well. We have applied the same filtering process and created an English resource. However, we didn't use pseudo terminal categories which are not relevant for English data. In addition, many administrative categories are polluting our representation space. For now, we have not filtered all these noisy categories but this first level of English resource is already useful.

These semantic resources (French, English) have been used and evaluated on two different tasks. In each case, the goal was to increase the recall or the relevance of a proprietary multimedia search engine. The first task consists in a kind of query expansion. We have re-structured our French semantic space with a concept lattice technique. The result enables to generate new queries close to the initial user query. The second task is a query translation task for cross lingual information retrieval (French to English). Our English semantic resource enables us to perform a choice between the different translation hypotheses by using a cosine measure in an associated vector space.

[3] http://networkx.lanl.gov/
[4] http://fr.wikipedia.org/wiki/Wikipédia:Catégories
[5] http://download.wikimedia.org/enwiki/latest

2 Query expansion task

The issue of query expansion is to add to the initial query some words that are “similar” to it, or even to use them to replace it. The goal is to give the user more relevant documents in regard to his query, even if they don’t match with the initial terms. For example, for the query “car”, the search engine will also retrieve documents that contain “automobile”. A similar application is the content recommendation. It consists in proposing to the user some documents that do not match directly his query, but that should nonetheless interest him. In both cases, the aim is to model the similarity of the proposed terms. To overcome this problem, we use the resource described in section 1. A query corresponds to the name of a Wikipedia page. A concept lattice made from the resource allows us to extract relevant terms that are similar to the query.

2.1 Resources

Our resource enables us to index the Wikipedia pages, not only from their parent categories but also directly from the categories that take part in the associated sub-lattice. Thus, we can directly get all the pages that are indexed with *Acteur américain* but also all of the “acteurs” or the “personnalités”. This is a first application of our work, and for now Wikipedia does not propose this level of indexation on all the pages. In order to restrict the quantity of data to process for this study, we selected the subset of Wikipedia pages indexed with the *“informatique”* topic (25 140 pages). The paths of all the categories associated to a page have been changed into one single “bag of categories”. Finally, each page is represented by a vector of categories that contains all its parent categories that reach the terminal category (hyperonyms, topics). Though, this representation loses the initial hierarchy, it allows us to use some standard techniques of classification or “data-mining”, that rely on vectors of features. However, the combination of specific categories such as *“Matériel_informatique”* and generic categories such as *"Informatique"* is still a very structuring information. For example, a part of the vector attached to *“Disque dur multimédia”* contains, among others:
Matériel_audio-vidéo, Audiovisuel, Médias, électronique, Multimédia,Informatique,Stockage_informatique,Matériel_informatique,Techniques_et_sciences_appliquées,Stockage_informatique, Industrie, économie…
The similarity between the pages is then computed from this representation.

2.2 Method overview

We generated a concept lattice from the data previously described by using an implementation from Girault (Girault, 2008). The concept lattice is made of “formal concepts”. In this formalism, each formal concept is described by an extension and an intension. The extension is an enumeration of the set of the members of the same category. The intension is the set of the properties shared by the member of this category. In our work, a formal concept extension is a set of page names. The intension corresponds to the categories names, which are the items shared by the vector attached to the extension pages. That means that the names of the categories are used as features that will define the common points between the page names. We obtained a lattice made of 293 636 formal concepts that described the computer science domain. In our framework, an instance of a formal concept is:
Extension: *['Ethernet', 'Segment_de_réseau']*
Intension:*['électronique','Informatique','Télécommunications','Portail:Science','Protocole_réseau','Portail:Informatique','Normes_et_standards_informatiques','Portail:Technologie','Composant_électronique','Protocole_de_técommunication','Normalisation_des_técommunications','Techniques_et_sciences_appliquées','Matériel_informatique','Connectique']*
We obtained a pool of pages that share common categories. In this first study, the similarity of a page is defined by taking into account formal concepts which corresponds to the following criteria:

- The extension includes strictly 2 items; one of these items is the considered page;
- The intension includes at least 8 items.

This choice allows to link two pages that share more than eight categories. Our previous example respects this criterion; thus the page *"Segment_de_réseau"* is similar to the page *"Ethernet"* since they share 14 categories. The preliminary results that follow have been produced according to this computation.

2.3 Results

The first results concern query expansion. We chose as an example the query *"Ethernet"*. Within the relevant[6] formal concepts, this term is included in the extensions with: *Chiffreur IP, RS-232, IEEE 802.3, Protocole réseau, Informatique, Réseau informatique, Matériel informatique, IEEE 802, Segment de réseau, Architecture informatique, Carrier Sense Multiple Access with Collision Detection, Medium Attachment Unit.* All these terms have a neighbourhood link with the initial query. By the way, this link sometimes corresponds to some semantic link: notably synonymy (*IEEE 802.3*) and hyponymy (*Protocole réseau*). Most of the other terms are used in the context of "*Ethernet*", such as *Segment de réseau*, or *Carrier Sense Multiple Access with Collision …* Some of these expansions can be too specific in regard to the initial query, and add noise in the retrieved documents. A way to carry out this problem is to filter the proposed terms with the application index, as done in (Gaillard and al. 2010); thus, only the terms actually existing in the available documents are used. The following results use the same strategy but are obtained from queries about products. The applicative framework then becomes a recommendation system: the obtained data allows to the user, from an initial query, to be proposed other products likely to interest him. An emerging and promising feature of our results is their structuring: it allows to sort them according to several thematic dimensions. The figure 1 displays the thematic context of the query *"Super_Mario_Bros"*, as well as the associated directions of thematic associations.

[6] According to the criteria described in section 2.2.

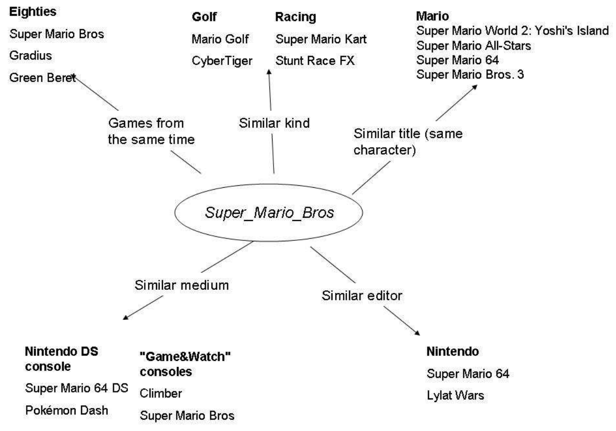

Figure 1: Game recommendations according to different thematics

Items of the extension *['Mario_Golf', 'CyberTiger']*, are linked to the intension *['Informatique', 'Projet_jeu_vidéo','Golf','Jeu_Nintendo_64','Jeu_vidéo'," Application_de_l'informatique",'Jeu_vidéo_sorti_en_1999','Jeu_vidéo_de_golf','Sport_individuel','Techniques_et_sciences_appliquées','Sport','Audiovisuel','Projet:Jeu_vidéo', 'Médias']*. We plan to automatically sort the extensions thanks to some items in their intension (for instance, “Golf” is explicitly referred to in the intension above). These first results, very promising, show that Wikipedia data constitutes a first resource for the expansion and recommendation techniques. This work is confirmed by other works that use this encyclopedia for this objective (Tien-Chen and al. 2007), (Peng and al. 2008).

3 Query translation task

3.1 Resources

In addition to our English semantic resources, we have constituted a bilingual (French/English) dictionary from the translation table for French pages[7]. This table lists multilingual links from French articles to their equivalents in various languages of Wikipedia, provided they exist. Only French to English links have been kept. A joint between the table of French articles[8] and the translation table enabled us to get a direct relation between French pages and English pages. We ended up with a table that directly associates titles with their various translations: *Avocat (fruit)/Avocado or Avocat (métier)/Lawyer*, for example. This translation table is comparable to a bilingual dictionary having 540 920 entries. Its specificity is that it contains many named entities and phrases, such as: *Avocat du diable/ Devil's advocate; L'Avocat du diable (film)/Guilty as Sin*. This bilingual dictionary can therefore be used directly but offers no solution to make a choice among the various translation alternatives.

3.2 Method overview

First of all, queries are segmented in lexical units which can be simple lexical entries or different kind of multi-words (terms, locutions, named entities). These selected lexical units are translated thanks to Wikipedia bilingual dictionary and we get one or several translated candidates for each lexical unit of the query. However, some lexical units don't get any translation at all. For a given query, we keep solutions of segmentation that give the maximum number of translated units and the longest units: we want to give priority to multi-words translation. The second phase is the disambiguation. Since there are often several alternatives for each lexical unit, many combinations can be candidates to the final translation. We choose the best combination according to a criterion of thematic homogeneity (Gledson and Keane, 2008). We use our English semantic resource (shortest category paths) to represent the semantic field of each selected lexical unit. Like for the query expansion task, we transform all the paths linked to a lexical unit to a "bag of categories". Finally, each lexical unit is represented by a flat vector of categories. The proximity between two category vectors is given by a cosine measure. This calculus is done between two adjacent units. In the

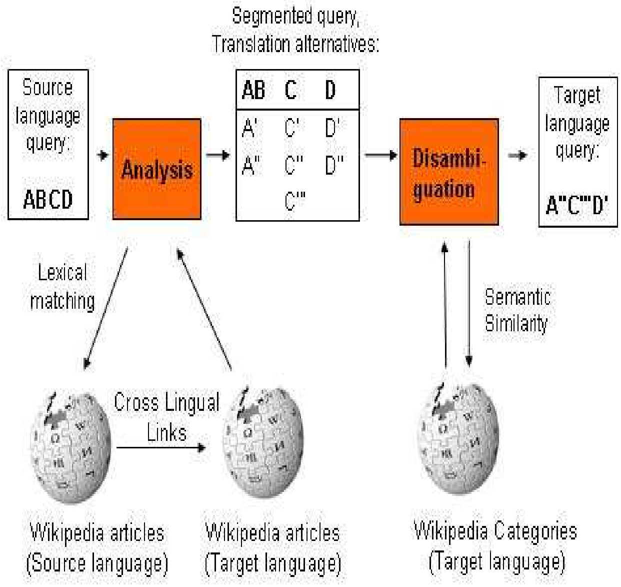

Figure 2: Wikipedia-based query translation. A query consisting of 4 words A, B, C and D is analyzed into 3 lexical units AB, C and D that have several candidate translations. After disambiguation, the A"C'"D' combination is deemed the most consistent.

Figure 3 shows the proximity between selected lexical units of the query "*avocat Tom Cruise*" and their associated shortest path reaching "*Article*" (only pages and terminal categories under "*Article*" are shown). We can see the overlap of "*Tom Cruise*" categories with "*Lawyer*" categories: "*People by occupation*" and "*People*". There is no category overlap with "*Avocado*" categories and "*Tom Cruise*" categories.

[7] frwiki-latest-langlinks.sql
[8] frwiki-latest-page.sql

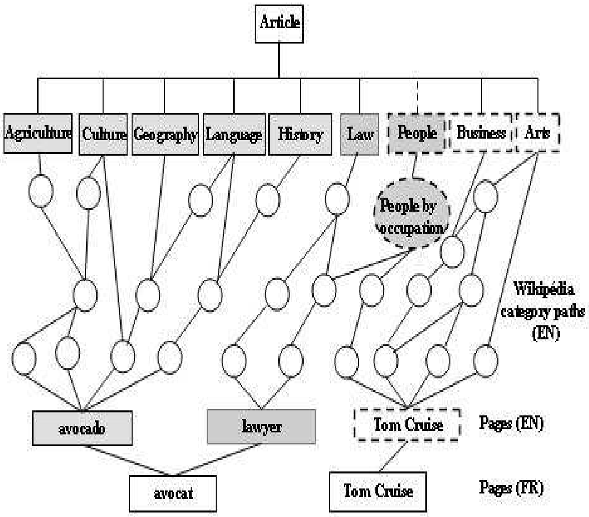

Figure 3: Disambiguation of the query "avocat Tom Cruise"

3.3 Results

We measured the accuracy of the translation of the prototype on a corpus of 750 queries issued from a monolingual multimedia search engine over three days during November 2009. Many of these 750 queries were typed in on several occasions. So the total number of queries in the corpus is about 7000.

Our goal was to get an idea of the performance of this system compared to standard solutions We compared the translations of these queries by our prototype with the translations of three well known MT services of the market, available online, freely: the online Systran solution[9], the ProMT online application[10], and the Google CLIR service[11]. We evaluate the Error Rate (ER) of each translator on the corpus. Our manual evaluation method was the following: each translation was given an accuracy score of success (0 for a wrongly translated query or not translated at, 0.5 for a partially correct translation and 1 for a good translation). The mean score M is computed over all these scores and weighted by the query frequency. The ER is defined by the formula: ER=1-M. Table 1 gives the following results:

M can be computed based on the 750 queries or based on each occurrence of each query (over the 7000 occurrences). We call it a weighted mean and the resulting ER is a called the weighted ER (ERw). Our prototype has no spelling mistake processing module and no grammatical processing at all either. Therefore, in order to compare its score with the three other state-of-the-art translators, we also measured the ER over the subset of queries that have no spelling mistake and no grammatical feature. For example the query "dog of Obama" would be grammatical because of the "of" genitive marker, as well as plurals. Each MT service or prototype was therefore given 6 different scores: ER over all the queries, ER over all the queries that have no spelling mistake or grammatical feature (ER-sg) and ER over the queries that do have spelling mistakes or grammatical features (ER|sg), these three rates weighted (ERw) or "flat". The results are presented in Table 1 (a lower ER means a more accurate translation):

	"Wiki "	**Systran**	**ProMT**	**Google**
ERw	**0.131**	**0.132**	**0.170**	**0.077**
ER	**0.331**	0.245	0.298	0.177
ERw- sg	**0.100**	**0.118**	**0.156**	**0.064**
ER- sg	**0.175**	0.155	0.225	0.111
ERw |sg	0.713	0.373	0.410	0.286
ER | sg	0.711	0.461	0.477	0.340

Table 1: ER Comparison of various MT solutions.

Several results can be highlighted. On the subset of queries that have no spelling mistake or grammatical feature, our ER is equal or slightly lower than the ER of other MT solutions, except Google. Since our system has no spelling or grammatical features, results on the spelling and grammatical queries (ER-sg) show that our prototype is very sensitive to spelling mistakes and grammatical features, its ER-sg query is higher than the others. This result shows that other system probably use spelling or grammatical features.

The fact that our accuracy is consistently much better with the weighted mean accuracy measure means that the most frequent queries are easier for our prototype to translate.

The weighted ER (ERw) keeps all the queries and measures the real performance of our system from the user point of view. Our result (13.1%) is comparable and even better than those of Systran or ProMT standard on line translation solutions, but worse than the specialised Google CLIR solution (nearly half errors). We have shown that our system is penalized in different ways (misspelling, grammatical parser and the lack of bilingual dictionary is far from being exhaustive, especially for standard lexical entries (verbs, common nouns). We think that the quite good performance of our system is partially du to the named entities frequency in our corpus. Nearly 60% of the queries contain a named entity and the Wikipedia bilingual dictionary contains many translated named entities.

We have designed a quite simple query translation system which only relies on Wikipedia data. These data are Wikipedia bilingual dictionary and a "bag of categories" for disambiguation purpose. This system is operational (not yet part of our search engine) and gives performance close to on line standard systems, the more adapted one being Google CLIR service. We have shown some of its weaknesses that we will soon improve.

Conclusion

We have filtered the Wikipedia categories lattice by the mean of a shortest path strategy. The result is a sub-lattice which extends each page with several parent category paths. These extensions generalize the semantic of category pages along hyperonymic and thematic axis. These linguistic relations are not explicitly labeled but the

[9] http://www.systran.fr/
[10] http://tr.voila.fr/
[11] http://www.google.fr/language_tools?hl=fr

generated representation space is useful enough to perform semantic query expansion and query translation disambiguation for queries related to a multimedia search engine. We will soon test its validity on a monolingual disambiguation task. Results seem to confirm our simple winning strategy which supposes that shortest categories paths are the most relevant.

In a future work, we will try to formalize this hypothesis within a theorical framework, the Minimum Description Length theory, which seems to be a logical way to follow. On a second hand we will further compare our resource to existing semantic resources, especially on a linguistic point of view. Automatic labeling of linguistic relations within the extracted sub-lattice is also in the scope of our next work.

References

Gaillard, B., Bouraoui, J.L., Guimier de Neef, E., Boualem, M. (2010). Expansion de requêtes pour la recherche d'information multilingue. *In Proceedings of the Conférence en Recherche d'Information et Applications (CORIA 2010).*

Girault, T., (2008). Exploitation de treillis de Galois en désambiguïsation non supervisée d'entités nommées. *In Proceedings of 15ème conférence sur le Traitement Automatique des Langues Naturelles, TALN'08*,260--269

Girault, T., (2008). Concept Lattice Mining for Unsupervised Named Entity Disambiguation. Concept *In Proceedings of Lattices and their Applications, CLA'08*, 32--43

Gledson, A., Keane, .J. (2008). Measuring Topic Homogeneity and its Application to Dictionary-Based Word-Sense Disambiguation. *In Proceedings of the 22nd International Conference on Computational Linguistics, Coling 2008, Manchester, UK,* pp 273--280.

Guegan, M. (2006). Catégorisation par les contributeurs des articles de l'encyclopédie Wikipedia.fr. *Mémoire de master de recherche informatique université paris XI, LIMSI CNRS*

Medelyan, O., Legg, C,Milne, D, Witten I.H.,(2008). Mining Meaning from Wikipedia. *Working Paper September 2008.*

Mihalcea, R. (2007). Using Wikipedia for Automatic word Sense Disambiguation. In *Proceedings of the NAACL 2007*, pp 196-203

Nastase, V. Strube, M. (2008). Decoding Wikipedia catégories for knowledge acquisition, *In Proceedings of the ,23rd national conference on Artificial intelligence (AAAI 2008)*, 1219-1224

Peng Y., Mao M. (2008), Blind Relevance Feedback with Wikipedia: Enterprise Track, *Proceedings of The Seventeenth Text REtrieval Conference (TREC 2008)*, 18-21

Ponzetto, S.P., Strube, M. (2007). Deriving a large scale taxonomy from Wikipedia. AAAI'07. *In Proceedings of 22nd national conference on Artificial intelligence*, 1440-1445.

Schönhofen, P., Benczur, A., Biro, I. and Csalogany, K. (2008). Cross-Language Retrieval with Wikipedia. *Advances in Multilingual and Multimodal Information Retrieval, Revised selected paper of CLEF 2007, Springer*, 72-79

Strube M., Ponzetto S. P. (2006). WikiRelate!: Computing Semantic Relatedness Using Wikipedia. *In proceedings of AAAI 2006*, 1419-1424

Suchanek, F.M., Kasneci, G., Weikum, G., (2008). Yago: A large Ontology from Wikipedia and WordNet. *Journal of Web semantics,Elsevier,*203-217

Tien-Chien L., Shih-Hung W.(2008), Query Expansion via Link Analysis of Wikipedia for CLIR, *Proceedings of NTCIR-7 Workshop Meeting*, 125-131

Zesch., T., Gurevych, I.,(2007). Analysis of the Wikipedia Category Graph for NLP Applications. *In proceedings of the Workshop TextGraphs-2:Graph-Based Algorithms for Natural Language Processing,* 1-8

Zesch., T., Gurevych, I. and Mühlhäuser, M. (2007). Analysing and Accessing Wikipedia as a Lexical Semantic Resource. *In Data Structures for Linguistic Resources and Applications,* 197-205