A semi automatic annotation approach
for ontological and terminological knowledge acquisition

Driss Sadoun

Laboratory MoDyCo / University of Paris-Ouest Nanterre.
driss.sadoun@u-paris10.fr

Abstract

We propose a semi-automatic method for the acquisition of specialised ontological and terminological knowledge. An ontology and a terminology are automatically built from domain experts' annotations. The ontology formalizes the common and shared conceptual vocabulary of those experts. Its associated terminology defines a glossary linking annotated terms to their semantic categories. These two resources evolve incrementally and are used for an automatic annotation of a new corpus at each iteration. The annotated corpus concerns the evaluation of French higher education and science institutions.

Key words : annotation, ontology, terminology, machine learning.

1 Introduction

For several years, French higher education and science institutions have been evaluated by an external institution. Most often this evaluation is conducted by the *High Council for the Evaluation of Research and Higher Education* (HCERES). Each year the HCERES recruits and trains academic experts participating in the evaluation of one or more institutions. These evaluations lead to the production of publicly accessible reports. These reports are rather standardized documents as their writing follows an established *evaluation template*. This *evaluation template* can be divided into ten *fields* : *Training*, *Governance*, *International Relations*, *Management*, *Piloting*, *Research*, *Student achievement*, *Scientific Culture* and *Valorization*. Each *field* may then be divided into several sub-fields (only twenty are explicitly named). Each report summarizes in its conclusion the strengths and weaknesses of the evaluated institution according to the *evaluation template*'s fields. Each year positive or negative assessments are manually classified by the HCERES experts according to the fields they refer to, in order to synthesise strengths and weaknesses of evaluated institutions over the same year.

In the reports, classifying an assessment means simultaneously identifying a term denoting a field and a term denoting an opinion. Sentences (1) and (2) below respectively contain a positive assessment on the field *training* and a negative assessment on the field *Valorization*. Sentences (3) and (4) contain more than one assessment, which is representative of the sentences of the conclusions. Thus, although sentences are generally well written, the aggregation of assessments can make them quite long and complex. Throughout this article, terms denoting a field appear in **bold** and terms denoting an opinion appear in *italic*.

1. une **formation doctorale** *très attractive*. / (A *very attractive* **doctoral training**.)

2. une **politique de valorisation de la recherche** *peu lisible*. / (A **most unclear policy of research valorization**.)

3. Une **présidence** *forte* mais une **gouvernance** *à revoir*. / (A *strong* **presidency** but **governance** *must be overhauled*.)

4. Une *difficulté* de **prévision des recettes** et un *manque* d'**approche politique dans la construction du budget**. / (Some *difficulty* in **forecasting revenues** and a *lack* of a **political approach in budget drafting**.)

This work of classification is a long, complex and subjective task for a human being. Given the amount of work, experts have to restrict their annotation to the ten major fields. Moreover, the work has to be shared out among several academic experts, hence no expert can have a global view of all reports. Since the number of reports keeps increasing, it has becomes necessary to automate this classification task by training an opinion mining system. The main issue of the described work is the classification of opinions into fields and sub-fields. Indeed, it appeared that for the *HCERES* experts the ambiguity of term denoting a polarity as *fort* (*strong*) is almost nil.

Identifying fields and their associated terms is a prerequisite to training an opinion mining system. Due to the number and the diversity of the evaluated institutions, a comprehensive and consensual listing of all possible sub-fields appeared to be hardly feasible for the *HCERES* experts. Hence, we propose to identify and structure the different fields empirically by performing an annotation task, during which each expert is allowed to suggest new sub-fields when he feels the need. Suggested fields are then consensually validated or rejected. The resulting consensual annotations are used to automatically build an ontology conceptualising the fields validated during the annotation task as well as a terminology linking the annotated terms to the fields they refer to. These resources serve to train an automatic annotation system. Afterwards, a new corpus is automatically annotated before being submitted to the experts who may validate, correct or add missing annotations. This whole process represents one iteration. The resulting ontology, terminology and annotated corpus are available on request.

2   Related work

Structuring extracted terms from corpora has been a topic of interest for many years (L'Homme, 2004; Claveau and L'Homme, 2005; Toledo et al., 2012; Szulman, 2011; Marciniak et al., 2016). One issue is to choose between either a *onomasiological* based approach (relating the term to its concept) or a *semasiological* based one (relate the term to its meaning) (L'Homme, 2004) which may consist in choosing an ontology or a terminology as a means of representation.

Ontology and terminology building is often based on textual corpora because texts carry shared and stable knowledge (Mondary et al., 2008). Building an ontology or a terminology from text in a completely unsupervised manner is hardly feasible. This is due to the nature of natural languages, whose meaning depends as much on the formulated sentence as its context. Therefore, proposed methods and tools offer an assistance to reduce human effort (Cimiano et al., 2009). They require human intervention for validating or rejecting the automatically extracted terminological, onto-logical or termino-ontological resources from texts as *Text2Onto* (Cimiano and Völker, 2005), *OntoLT* (Buitelaar et al., 2004), *OntoGen* (Fortuna et al., 2007), *Terminae* (Szulman, 2011) or TermoPL (Marciniak et al., 2016). The validation may then depend only on a single person, which makes it subjective. Our aim is to minimise as much as possible the inherent human subjectivity.

To do so, we believe that the validation of identified resources has to be done consensually. Moreover, to reduce confusion and distinguish between the ontological and the terminological level to which a term can be related, we propose to use texts to build both a consensual representation of knowledge -formalized by an ontology- and a framework for interpreting terms in the context in which they appear -formalized by a terminology.

Identifying terms often depends on corpus annotation. When compared to the need and generally speaking, very few annotated corpora for opinion mining have been proposed and this regardless of language (Wiebe et al., 2005; Steinberger et al., 2014; Hammer et al., 2014; Wachsmuth et al., 2014; Croce et al., 2013; Mele et al., 2014; Daille et al., 2011; Lark et al., 2015). The lack of annotated corpus is due to the complexity of a human annotation which is notoriously difficult even for domain experts (Bernier-Colborne and Drouin, 2014). This observation is not recent and many works have proposed (semi-)automatic annotation approaches (Erdmann et al., 2000; Swift et al., 2004; Dufour-Lussier et al., 2012; Christen et al., 2015) most often based on the use of an ontology. Indeed, an ontology may be particularly helpful to help define and use complex annotation schemas (Ogren, 2006). In our approach, we propose an ontology and terminology evolution approach resembling those proposed in (Taleb et al., 2009; Toledo et al., 2012), which allows both resources to evolve while the corpus is iteratively annotated.

Empirical results have shown that using the syntactic structure of sentences to capture contexts of formulation in text is relevant for the task of opinion mining (Wu et al., 2009; Jiang et al., 2011; Lapponi et al., 2012). In addition, experiments have shown that methods based on dependency graphs may perform significantly better than the word-based methods (Hammer et al., 2014; Vilares et al., 2015). So, we chose to add to our semantically annotated corpus some annotations related to the syntactic features of the words involved in the annotated terms. In the remainder of this article, we discuss the practical value of such a choice for term identification.

3   Description of the corpus

Our corpus is made up of sentences extracted from the conclusions of the 34 evaluation reports published in 2013. More precisely sentences belong to the subsections detailing the *strong points*, the *weak points* and the *recommendations* addressed to evaluated institutions. The corpus contains 692 sentences, which represents around 20 sentences per report. The writing style is not standardized and depends on the writer. Sentences can be quite long and complex. Indeed, the number of words in the corpus is 12171, which means an average of 17 words per sentence. This length is due to the use of complex terms and to the conjunction of several nominal and verbal groups. Moreover, the majority of terms ($\simeq$ 73%) referring to a field or an assessment are complex terms, *i.e.* formed by multiple words such as **gestion des ressources humaines** (human resources management), **formation continue** (lifelong training) or *très bas* (very low). These terms may contain contiguous words, such as the term **sentiment d'appartenance** (sense of belonging) in sentence (1) or non contiguous ones like the same term in sentence (2). Each sentence may contain more than one opinion, as is the case in sentences (3) and (4). In the following sentences arrows in bold indicate the link between *head* words of terms denoting an assessment with *head* words of terms denoting a field. The light arrow indicates the link between non contiguous words belonging to the same complex term. In both cases, these arrows link words that are *heads of terms*. Among the words that form a term, the *head word* is the one that determines the syntactic features of the term. The other words that belong to a term can be designated as the *dependent words* in a syntactic sense. In a semantic sense they may be considered as *modifiers*. The addition or the deletion of

dependent words does not change the syntactic distribution between *head words*. For example, sentence (5) highlights that deleting the words **d'appartenance** and *très* respectively from the terms **sentiment d'appartenance** and *très fort* does not modify the syntactic dependency between the terms *fort* and **sentiment**.

(1) Un *sentiment d'appartenance* **très fort** / A **very strong** *sense of belonging*

(2) Un *sentiment* **très fort** *d'appartenance* / A **very strong** *sense of belonging*

(3) Une *situation financière* **non maitrisée** et une **absence** de *sincérité budgétaire.* / A **non controled** *financial situation* and a **lack** of *budget honesty.*

(4) **Renforcer** le *pilotage* et le *contrôle de gestion* / **Reinforce** the *piloting* and the management control.

(5) Un *sentiment* **fort** / A **strong** *sense*

4   Semi-automatic term annotation

The aim of the semi-automatic term annotation is to bring out the shared vocabulary of the *HCERES* experts. First, a manual annotation involving the experts is performed and leads to a *consensual annotation*. Annotated terms and their associated fields are then used as a set for training an automatic annotation system. This system is based on the following automatically built resources : 1) an ontology structuring the evaluation fields ; 2) a terminology linking each annotated term to the field it refers to within the ontology and 3) syntactico-semantic patterns characterizing the features of the annotated terms. The trained system is used for the automatic annotation of a new corpus. The automatically annotated corpus is then submitted to a new manual annotation. This incremental process of the semi-automatic term annotation is illustrated in Figure 1.
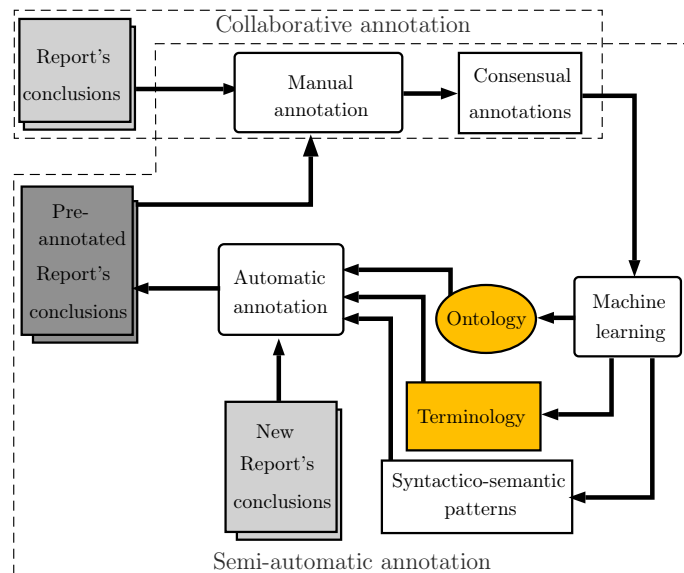

FIGURE 1 – The semi-automatic term annotation steps.

4.1 Experts' annotation

A total of 22 experts from the *HCERES* have participated in three successive annotations. They have been divided into six groups of 3 to 4 experts. Each group had a sixth of the corpus to annotate. The annotation has been done under the platform *Webanno* (Yimam et al., 2014) which enables online annotations. Many annotators can annotate the same corpus. Each annotator annotates their version of the corpus without viewing others' annotations. Then, elements of agreement and disagreement can be compared. Within *Webanno*, each annotation category may have an unlimited number of labels. In our framework, we defined the ten *main fields* as categories and their sub-fields as labels. Hence, the starting annotation tag set contained ten categories and twenty labels. To extend this tag set, annotators were allowed to offer new sub-fields for each major field when they felt it was necessary, *i.e.*, when the existing fields were not sufficient to characterize the terms to annotate. Indeed, *Webanno* authorizes the creation of new labels on the fly during annotation. When a new label is chosen, it is visible and usable by all other annotators. Moreover, if an expert identifies a new label which is better to annotate a term it can always change its previous annotations. At the end, each newly suggested sub-field can be validated or rejected, as it is based on the subjectivity of the annotator who proposed it.

Experts were all volunteers and motivated. However, as they had no previous annotation experience and due to their tight schedule, the first manual annotation took more than five weeks. Then, the use of a semi-automatic annotation approach quickly appeared as a need.

4.2 Inter-annotator agreement

Our annotation protocol is quite unusual since the tag set is not finite and evolves at each iteration as each annotator is able to propose a new label on the fly. As far as we know, there is no metric for calculating inter-annotator agreement (IAA) fitting that case. For the calculation of IAA, we then chose to calculate the F-measure by pairs of annotators of the same group. As expected, IAA for the first manual annotation was low ($\leq 40\%$). Hence, to reconcile divergent annotations, discussion sessions were organized. They took place between at least two annotators along with the authors of this article with the aim of finding common ground for each divergent annotation. In order to fit experts' schedules, each sessions duration was half an hour. Figure 2 is an example of a consensus reached on divergent annotations within the sentence 16. In this figure, the first line represents the consensual annotation, *i.e.* the one accepted by all. The next four lines correspond to the annotations of four different experts. A total agreement can be noted for the annotation of the term *développer* (develop) which refers to a *Recommendation*. However, concerning the term **sentiment d'appartenance** (sense of belonging) the associated sub-field differs for each expert. This highlights the subjectivity of the annotation task and by extension of the classification task. The annotation of *user-2* : **Identité** (Identity) is the one chosen, which is a sub-field **Gouvernance** (Governance). Once the consensual annotations are built they are used to build and change the ontology, the terminology as well as to train the automatic annotation system.

5 Automatic creation of the ontology

We chose to use an ontology to formalize the conceptual vocabulary of the *HCERES* experts as it is easily understandable by humans and readable by machines. It also allows experts to have a concise representation of their conceptual vocabulary. In addition, it offers the means to annotate its own elements in order to specify their meaning and give extra information and definitions about the set of semantic labels that will be used for the opinion classification.
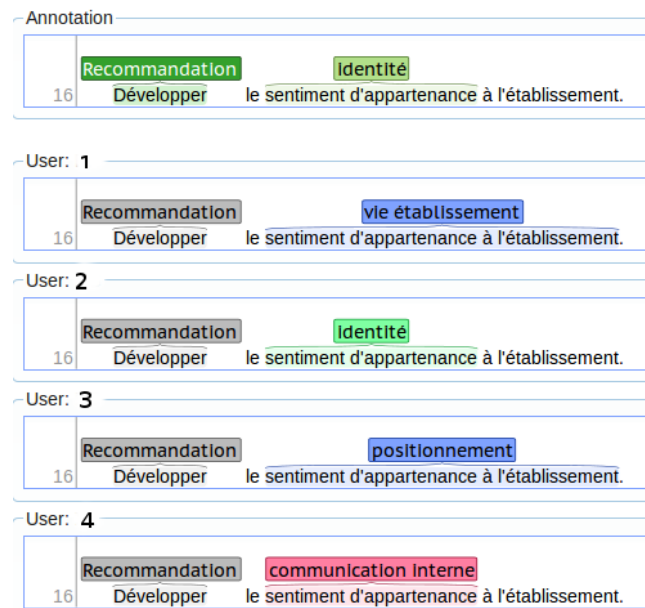
FIGURE 2 – A consensus on a divergent annotation.

After each annotation new sub-fields may be proposed by the annotators. Only fields that are agreed upon by the experts are kept i.e. added to the conceptual vocabulary. Each annotated corpus produced at an iteration is added to the collection of previously annotated ones. The whole annotated corpus is then analysed in order to automatically extract fields that have been used as annotation categories and labels. Each sub-field belongs to a field that is more general. Thus, a two-level hierarchy structure is extracted from the annotated corpus wherein a sub-field is a sub-concept of a field. This structure is represented in the form of an OWL ontology. This ontology is intended to serve the identification of opinions and their classification based on the several fields and types of assessments. Hence, the ontology also contains the concept *Assessment* and its three sub-concepts : *Positive*, *Negative* and *Recommendation*. Figure 3 is a fragment of the automatically built ontology. After a manual and two semi-automatic annotations, the ontology numbers 117 concepts including 113 fields. Among these fields, 83 are new sub-fields proposed and accepted by consensus by the experts during their collaborative annotation. These 83 new sub-fields are selected from over 100 proposed sub-fields over the three annotations.

The number of kept new sub-fields may appear high. It can be explained by the variety of evaluated institutions covered by the *HCERES* reports. This number tends to stabilize after the processing of the 34 evaluation reports published in 2013. However, it should increase (less significantly) in the future due to the particularity of certain *higher educational and scientific institutions* that reports were not yet annotated and the evolution of French institutions.

6 Automatic creation of the terminology

Opinion classification is largely based on the recognition of relevant terms in text and the conceptual categories they refer to. In order to link the linguistic knowledge to the conceptual vocabulary, we automatically create a terminology from the semi-automatic annotations. For higher recall, terms are stored in their lemmatized form. In addition, terms are linked to the sentences of the corpus they belong to. Thus, the terminology defines a glossary of the domain, with accompa-
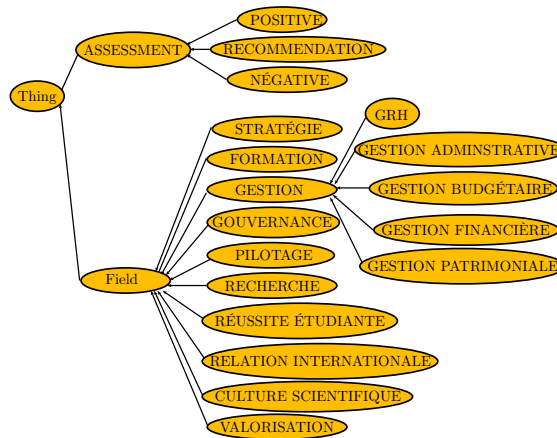
FIGURE 3 – A fragment of the ontology detailing the sub-fields of **GESTION**.

nying sentence examples for each term. For academic experts, this represents a precious resource for understanding the meaning behind the existing fields.

Term organization within the terminology is intended to allow the classification of opinions based on the organization of sub-fields within the ontology. For example, if the term *doctorate* is associated to the concept **Doctoral training** within the terminology while the concept **Doctoral training** is a sub-concept of the concept **Education**, then the identified opinion belongs to the field **Education**. If in the evolution of the ontology, the concept **Doctoral training** becomes a sub-concept of the concept **Research**, then the identified opinion will no longer belong to the field **Education** but to the field **Research**. Thus, the conceptual choices do not influence the way the opinion mining system is learned but only the opinions' classification. In practice, a sub-concept could be under two different Concepts. However, as the ontology hierarchy determines the distribution of opinions, at the moment, the choice has been made to avoid that kind of situation.

Throughout the three successive annotations, 1137 distinct terms were annotated then included in the terminology. Each was associated with the concept representing a sub-field of the ontology. When a term is not precise enough to be associated to a sub-field, it is associated directly to a field.

The ontology and the terminology do not prevent the portability of our approach. Indeed, in case they do not exist, they would be built from the first manual annotation and will evolve at each iteration. Otherwise, they will still evolve at each iteration.

7 Syntaxico-semantic pattern learning

The purpose of learning patterns is to capture the context of formulation of terms referring to a field or an assessment. These terms can be complex and contain non-contiguous words. Furthermore, words forming these terms may be subject to declension or conjugation. Hence, the characterization of these terms, must take into account both semantic and morpho-syntactic features. We propose to capture the formulation contexts of annotated terms using *syntactico-semantic patterns*. These patterns are used for automatic recognition of contiguous or non contiguous complex terms of the terminology. Constraining the identification of words based on their morpho-syntactic features, is meant to increase the accuracy of term recognition. To do so, at each iteration annotated sentences are analysed by the statistical dependency parser for French *Bonsai* (Candito et al., 2010). The syntactic features of sentences are then merged with their annotations. Figure 4

ID	FORM	LEMMA	CPOS	POS	FEATURES	HEAD	DEP	ANNOTATION	ANNOTATION LINK
1	Un	un	D	DET	g=m|n=s|s=ind	2	det	<>	<>
2	sentiment	sentiment	N	NC	g=m|n=s|s=c	0	root	<Identité[2, 5, 6] (Gouvernance)><>	<>
3	très	très	A	ADV	_	4	mod	< >	<>
4	fort	fort	A	ADJ	g=m|n=s|s=qual	2	mod	<Positive[3, 4] (Assessment)>	<>
5	d'	de	P	P	_	2	dep	< >	< 2]Suite_Gouvernance >
6	appartenance	appartenance	N	NC	g=f|n=s|s=c	5	obj	< >	< >

FIGURE 4 – A sentence annotated both syntactically and semantically.

illustrates the matching of syntactic and semantic features for the sentence : *Un sentiment très fort d'appartenance*. The first eight columns contain syntactic information while the last two columns come from the annotations. The 9th column contains the annotation features. For example, the feature *Identité[2, 5, 6] (Gouvernance)* means that the term (*sentiment d'appartenance*) composed of the words numbered 2, 5 and 6 has the annotation **Identité** which is a sub-field of **Gouvernance**. The 10th column contains the link with a previous part of a term containing non-contiguous words. For example, the feature *2]Suite_Gouvernance* of word number 5 (*d'*) means that this word is linked to word number 2 (*sentiment*) within an annotation of the field **Gouvernance**.

Annotations are represented on *head words* because syntactic dependencies linking terms referring to a field to those referring to an assessment are expressed on their *head nodes*. For example, the dependency *mod* between node 4 (*strong*) and its *head node* 2 (*sense*), which means that the term *very strong* modifies the term *sense of belonging*, is expressed over the head words *strong* and *sense*. The advantage of using syntactic dependency is that regardless of the order of words in the sentence, whether terms are simple or complex, or contiguous or not, the dependencies between head nodes remain unchanged. This property is very useful in our context in which most terms are complex and almost 15% are non contiguous. Semantic and syntactic levels are complementary without being interdependent. Thus, opinion mining system training, can be based on either the semantic level or the combination of both levels.

Syntactico-semantic patterns are acquired using an algorithm which calculates the shortest dependency path between two terms. They contain word lemmas, morpho-syntactic categories, morphological features and syntactic dependencies to check that identified words are syntactically linked to form a complex term of the terminology. For example, the pattern bellow represents the shortest dependency path between the terms *sentiment d'appartenance* (words of nodes 29,30 and 31) and *réel* (*real*). The involved dependency is *mod* (for modifier) that links the nodes of the terms *sentiment* (node number 29) and *réel* (node number 28).

mod({sentiment,sentiment,29,N,NC,g=m\|n=s\|s=c,26,Identité[29,30,31] (Gouvernance)},
{réel,réel,28,A,ADJ,g=m\|n=s\|s=qual,29,Positive[28] (Appreciation)})

Acquired syntactico-semantic patterns cover 116 semantic categories (3 types of assessment, 10 fields and 103 sub-fields). A total of 776 syntactico-semantic patterns are acquired. Among them, 728 are for complex term covering all the semantic categories and 48 are for simple terms covering only 37 semantic categories. These last numbers show how large the proportion of complex terms is in our corpus.

8   Semi-automatic term acquisition

The automatic term acquisition is guided by the ontology, the terminology and the syntactico-semantic patterns. Each combination of words of a sentence is checked looking for a match within the terminology. Terms that match a syntactico-semantic pattern are then annotated with the concept they are linked to within the terminology. Then, the automatically annotated corpus is submitted to the experts who may validate, correct or add missing annotations. Among the semi-automatic annotations, we distinguish three kinds of annotated terms :

— added ($\simeq$43%) : terms newly annotated by the experts.

— validated ($\simeq$32%) : terms automatically annotated that were kept by the experts.

— expanded ($\simeq$25%) : terms annotated by experts based on automatic annotations. For example, the annotated term **pilotage de la formation** (piloting of education) contains the two automatically annotated terms *pilotage* (piloting) and *formation* (education).

Almost a third ($\simeq$32%) of term annotations are automatic annotations and more than half ($\simeq$57%) of the annotations are based on the semi automatic annotations. Annotators confirmed that the automatic pre-annotation considerably eased and sped up their task. Indeed, at each new iteration annotators are less required to act as terms that have been consensually annotated once do not have to be manually annotated again. Moreover, automatic annotations serve as real examples for the annotators. These results show that automatic pre-annotation provides valuable assistance for expert annotators. In addition, automatic annotation reflects the consensual agreements between annotators, thus making it less subjective. After three iterations of the semi-automatic annotation, 1792 terms have been annotated : 932 terms referring to a field and 860 terms referring to an assessment.

9   Conclusion

We proposed a semi-automatic method for the acquisition of ontological and terminological knowledge. This method relies on incrementally building and tuning up these domain resources thanks to previous expert's annotations i.e. consensually approved knowledge. At each iteration these resources serve the automatic annotation of new corpora to ease and speed up experts' annotation work, decreasing the inherent subjectivity of such a task. The annotated corpus, the ontology and the terminology are built to train an opinion mining system for the evaluation of higher education and science institutions. In our method, domain dependent resources are built from scratch if they do not exist and evolve incrementally. So, we believe that it can be applied to other domains.

References

Gabriel Bernier-Colborne and Patrick Drouin. 2014. Creating a test corpus for term extractors through term annotation. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 20(1) :50–73.

Paul Buitelaar, Daniel Olejnik, and Michael Sintek. 2004. A protégé plug-in for ontology extraction from text based on linguistic analysis. In *The Semantic Web : Research and Applications*, volume 3053, pages 31–44.

Marie Candito, Benoît Crabbé, and Pascal Denis. 2010. Statistical French Dependency Parsing : Treebank Conversion and First Results. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, may.

Victor Christen, Anika Groß, Julian Varghese, Martin Dugas, and Erhard Rahm, 2015. *Annotating Medical Forms Using UMLS*, pages 55–69.

Philip Cimiano and Johanna Völker. 2005. Text2onto - a framework for ontology learning and data-driven change discovery. In *Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB)*, volume 3513, pages 227–238.

Philipp Cimiano, Alexander Mädche, Steffen Staab, and Johanna Völker. 2009. Ontology learning. In *Handbook on Ontologies*, pages 245–267.

Vincent Claveau and Marie-Claude L'Homme. 2005. Structuring terminology using analogy-based machine learning. In *Proceedings of the 7th International Conference on Terminology and Knowledge Engineering, TKE*.

Danilo Croce, Francesco Garzoli, Marco Montesi, Diego De Cao, and Roberto Basili. 2013. Enabling Advanced Business Intelligence in Divino. In *Proceedings of the 7th International Workshop on Information Filtering and Retrieval co-located with the 13th Conference of the Italian Association for Artificial Intelligence*, pages 61–72.

Béatrice Daille, Estelle Dubreil, Laura Monceaux, and Matthieu Vernier. 2011. Annotating opinion-evaluation of blogs : the Blogoscopy corpus. *Language Resources and Evaluation*, 45(4) :409–437.

Valmi Dufour-Lussier, Florence Le Ber, Jean Lieber, Thomas Meilender, and Emmanuel Nauer. 2012. Semi-automatic annotation process for procedural texts : An application on cooking recipes. *CoRR*.

Michael Erdmann, Alexander Maedche, Hans-Peter Schnurr, and Steffen Staab. 2000. From manual to semi-automatic semantic annotation : About ontology-based text annotation tools. In *Proceedings of the COLING - Workshop on Semantic Annotation and Intelligent Content*.

Blaz Fortuna, Marko Grobelnik, and Dunja Mladenic. 2007. Ontogen : semi-automatic ontology editor. In *Proceedings of the 2007 conference on Human interface : Part II*, pages 309–318.

Hugo Lewi Hammer, Per Erik Solberg, and Lilja Øvrelid. 2014. Sentiment classification of online political discussions : a comparison of a word-based and dependency-based method. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 90–96.

Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. 2011. Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies - Volume 1*, pages 151–160.

E. Lapponi, J. Read, and L. Ovrelid. 2012. Representing and Resolving Negation for Sentiment Analysis. In *Data Mining Workshops (ICDMW)*.

Joseph Lark, Emmanuel Morin, and Sebastián Peña Saldarriaga. 2015. CANÉPHORE : a French corpus for aspect-based sentiment analysis evaluation. TALN 2015.

Marie-Claude L'Homme. 2004. A lexico-semantic approach to the structuring of terminology. In *Proceedings of the 3rd Computerm*, pages 7–14.

Malgorzata Marciniak, Agnieszka Mykowiecka, and Piotr Rychlik. 2016. Termopl - a flexible tool for terminology extraction. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC)*.

Francesco Mele, Antonio Sorgente, and Giuseppe Vettigli. 2014. An Italian Corpus for Aspect Based Sentiment Analysis of Movie Reviews. In *First Italian Conference on Computational Linguistics CLiC-it*.

Thibault Mondary, Sylvie Després, Adeline Nazarenko, and Sylvie Szulman. 2008. Construction d'ontologies à partir de textes : la phase de conceptualisation. In *Actes des 19èmes Journées Francophones d'Ingénierie des Connaissances (IC'08)*, pages 87–98.

Philip V Ogren. 2006. Knowtator : a plug-in for creating training and evaluation data sets for biomedical natural language systems. In *Proceedings of the 9th International Protégé Conference*, pages 73–76.

Josef Steinberger, Tomáš Brychcín, and Michal Konkol. 2014. Aspect-Level Sentiment Analysis in Czech. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*.

Mary D Swift, Myroslava Dzikovska, Joel R Tetreault, and James F Allen. 2004. Semi-automatic syntactic and semantic corpus annotation with a deep parser. In *Proceedings LREC*.

Sylvie Szulman. 2011. Une nouvelle version de l'outil terminae de construction de ressources termino-ontologiques. In *22èmes journées francophones d'Ingénierie des Connaissances (poster)*, page 3 pages.

Nora Taleb, Sellami Mokhtar, and Michel Simonet. 2009. Knowledge acquisition for the construction of an evolving ontology : Application to augmented surgery. In *Proceedings of World Academy of Science, Engineering and Technology*.

C. M. Toledo, O. Chiotti, and M. R. Galli. 2012. An ontology evolution approach for information retrieval strategies with compound terms. In *Informatica (CLEI), 2012 XXXVIII Conferencia Latinoamericana En*, pages 1–10.

David Vilares, Miguel A. Alonso, and Carlos Gómez-Rodríguez. 2015. On the usefulness of lexical and syntactic processing in polarity classification of Twitter messages. *Journal of the Association for Information Science and Technology*.

Henning Wachsmuth, Martin Trenkmann, Benno Stein, Gregor Engels, and Tsvetomira Palakarska. 2014. A Review Corpus for Argumentation Analysis. In *Computational Linguistics and Intelligent Text Processing*, volume 8404, pages 115–127.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, 39(2-3) :165–210.

Yuanbin Wu, Qi Zhang, Xuanjing Huang, and Lide Wu. 2009. Phrase Dependency Parsing for Opinion Mining. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing : Volume 3*, pages 1533–1541.

Seid Muhie Yimam, Richard Eckart de Castilho, Iryna Gurevych, and Chris Biemann. 2014. Automatic Annotation Suggestions and Custom Annotation Layers in WebAnno. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. System Demonstrations*, pages 91–96.