Qualitative and Quantitative Criteria for the Concept Evaluation Task

Lobna Karoui, Supélec France, Nabil El Kadhi, L.E.R.I.A. Lab France

I. *Abstract*—**Ontological concept evaluation is a difficult task. Till now, it is done either by domain expert or a knowledge base (thesaurus, ontology, etc.). In this research, we propose a new evaluation method based on a large web document collection, several context definitions deduced from it and three criteria. It provides a support for either a domain expert or a novice user. Moreover, it facilitates the semantic interpretation of the word clusters and consequently the ontological concept generation. Our contribution is to propose an evaluation framework that does not depend on a gold standard, could be applied to any domain even if expert intervention is not available and provides qualitative and quantitative criteria. Our experiments show how our method assists and facilitates the evaluation task for the domain expert.**

Index Terms—**Ontology, concept, evaluation, semantic web, context**

I. INTRODUCTION

Some current work in data annotation, information retrieval depends on ontologies. The development of these applications is related to the richness of the ontology's conceptualization. Ontology [14] is defined as "a specification of a conceptualization". The ontology learning researches [3, 7] are often more interested in building the ontology than in evaluating it. Moreover, there is not a standard evaluation method or some evaluation metrics by which we can confirm that the ontology is either good or bad. So, ontology evaluation is very difficult, since it depends more on some aspects than other and has a specific purpose. In order to deal with this issue, we are interested in the evaluation task. In this paper, we focus on the evaluation of the ontological concepts that are extracted from the web documents. We work on French documents related to the tourism domain. Our evaluation method is based on the concept of "Contextualization", a large collection of web documents and three revealing criteria which are the credibility degree, the cohesion degree and the eligibility degree. Each criterion provides complementary information for a better expert's interpretation of the word clusters. The credibility degree computation algorithm is based on two context types which are the linguistic context and the documentary one. It computes the credibility degree associated to each word cluster and to each context, extracts some useful information and tries to propose some suggestions to the user such as deleting one word, keeping it, etc. The cohesion degree algorithm provides quantitative information for each cluster in order to define the relation between the words inside a cluster. The eligibility degree algorithm orders the words and proposes one of them as a possible future concept or the semantically closest word to the future concept. Our ontological concept evaluation method helps either an ordinary user to evaluate the word clusters before the expert do it or the expert himself. It gives the possible word associations existing in these contexts, some semantic tag suggestions, some information about the relation between the words of each cluster, the order between them, deletes the noisy elements or moves them to their appropriate clusters, etc. Our evaluation method provides a qualitative evaluation task thanks to the word associations and a quantitative one thanks to the three degrees. These later are computed during the process and based on various contexts and the Google database. Our method does not depend on a gold standard and it could be applied in any domain even if expert intervention is not available.

II. RELATED WORK

Ontology evaluation remains a real problem in the area of the semantic web. There is no standard methodology or approach to evaluate ontology. This is due to the fact that the ontology learning depends on several aspects such as the purpose of the ontology building, the application (s) using the ontology, the entities constituting the ontology, the kind of ontology (domain ontology, task ontology, etc.), etc. Moreover, the ontology evaluation can take into account the evaluation of the extracted concepts (vocabulary and concept), the evaluation of the taxonomic and non taxonomic relations, the evaluation of the whole ontology based on its application, on existing data, on existing ontology or thesaurus or on the human intervention, etc. In this survey, we present research related to the evaluation of concepts which constitutes the ontology.
A concept is "A general idea derived or inferred from specific instances or occurrences" (http://www.thefreedictionary.com). Discovering concepts implies having a vocabulary for which the idea of the concept is deduced. Therefore, the ontology engineer can evaluate each or both vocabulary and extracted concepts. In order to evaluate the vocabulary, Meadche and Staab [8] proposed an approach that aims to evaluate the lexical and vocabulary level of an ontology. They have defined a similarity measure in order to compare two strings: one provided from the produced ontology and the other one from an existing ontology. In [1], the authors evaluate their lexical by using WordNet and precision and recall measures. Based on this vocabulary, ontology building approaches

applying clustering methods obtain word clusters as potential future concepts. So, the question is how to evaluate these clusters. In the clustering process, the quality of a cluster is generally based on homogeneity or compactness. In [12], some criteria for the statistical evaluation of unsupervised learners have been defined. However, the ontology learning applications cannot rely on these standards defined for other applications. Moreover, cluster homogeneity does not imply that the words in the cluster are semantically closer or that the associated label satisfies the domain expert. For the concepts extraction, the evaluation is more challenging. In [4], the authors proposed an evaluation method based on a collaborative manual ontology engineers in order to maintain the suggestions resulting from the maximum number of experts. Navigli and al [9] proposed a qualitative evaluation by multiple domain experts that answer to a questionnaire in which they evaluate the quality of the discovered concepts. In [5], the authors present the results (produced by the clustering algorithm) to two domain experts. As a first step, individually, each of them evaluates and labels manually the word clusters. Then, they work together in order to discuss about the results, and their label propositions. They produce only one evaluation and labeling result in which they agree. Based on this evaluation, the authors define four criteria to analyze the word distribution, the semantic interpretation, the extracted concepts and the generality degree of these later. In [4, 5, 9], the concept evaluation is based on a human intervention which is a painful task. In other research [8, 1, 12], the evaluation and the labeling process is based on statistic measures, a thesaurus, an existing ontology or an application. Generally, we remark that thesaurus doesn't cover all specific aspects of a domain. Also, the evaluation based on an existing ontology is not sufficient because, in some fields, ontology does not exist or does not contain all the concepts founded in the produced ontology. Moreover, it is not evident to find an application that uses the produced ontology. So, the evaluation is a difficult task. That's why; we deal with this issue.

III. THE FUNDAMENTAL DIRECTIVES OF OUR EVALUATION METHOD

Our evaluation method is based on the concept of contextualization. It exploits the richness of the web documents. In the following sections, we explain these ideas.

A. The Contextualization Concept

Our idea is as follows: "looking in the Web in order to understand the meaning of each word or two words together and so on" could be a solution but why?

Terms are selected from their contexts in order to group them (computing occurrences component section [5]) but they are presented to the knowledge engineer or to the expert domain without any context after a decontextualization process that's why the evaluation step is always difficult. In 1993, McCarthy says [6] that decontextualization is "to abstract a piece of knowledge from contexts into a more general context that cover the initial context". In our case, for each cluster the general context is the domain (tourism). But this information is not sufficient to evaluate a cluster and to give it a semantic tag that's why we decide to use the contextualization concept.

In our study, for the evaluation process, we use the contextualisation notion. So, our idea is to use information sources to extract contexts. But what type of information sources?

A solution for ensuring an easy analysis is to use a big collection of web documents related to the same studied domain. But, why do we use web documents?

B. The richness of the Web documents

The difference between thesaurus and web documents is that thesaurus is related to one domain and it is implemented by one scientific community who takes strategic decisions in order to focus on some domain aspects or uses generic expressions. However, domain web sites are written by persons having different opinions and purposes that's why there is not only one methodology or one goal (each web site is designed for one geographic place or government or association or private institution, etc.). In this case, domain vocabulary and situations deduced from it (constructed from the same words in multiple contexts) are so many. That's why; we could not limit them in comparison with a thesaurus or an existing ontology.

Another point is the domain limits versus the domain corpus coverage. In our case, choosing tourism as domain is not an easy task. In OMT thesaurus, scientific says that: "It is more difficult to discern the limits of tourism activities, as its facets or sectors extend beyond the field of tourism itself. Indeed, tourism overlaps with other social activities that are often wider in scope. Restricting the definition of tourism to leisure and holidays, would strip it of much of its sense". Moreover, concept extraction is based on a corpus which covers some domain aspects related to a generality degree (corpus speaking about tourism policies) or geographic elements (tourism in Sahara, tourism in Canada), etc. So, the manners and the interest to present or explain the tourism concepts from one site to another or from one country to another or even from one region to another are completely different. Consequently, the extracted concepts depend not only on the domain but also on the corpus. By proposing an evaluation based on web documents, we have the opportunity not only to understand the semantic connections between words but also to enhance the cluster by new words when they appear with the initial ones.

To obtain this domain web collection of French documents, we use a cleaner such as HTTrack Website Copier (Fig. 1: step2). Then we treat them thanks to the pre-processing step of our system [5]. This last cleans and structures the collected web pages. Moreover, some elements are coded using characters that are specific to browsers (é → é) or are without accents. We process these documents by deleting some elements (script), rectifying codes and correcting accents (Fig. 1: step 3). Then, by performing various corpus analyses, we determine corpus characteristics (Fig. 1: step4) in conjunction with the nature (term distribution), the structure (tags, phrases, etc.) and the linguistic elements (verb, noun, nominal group, etc.) and evaluate the corpus. Thanks to all these analyses, we can use the web collection into the following evaluation process.

IV. THE ONTOLOGICAL CONCEPT EVALUATION METHOD

In this section, we present our evaluation method. This new method is used for the evaluation of the ontological concepts related to one specific domain and produced by clustering techniques. It is based on three revealing criteria that help the domain expert during the evaluation task. These criteria are:
- The credibility degree: the character of what we can believe
- The cohesion degree: the character of a thing that all its parts are united with a logical relationship between its elements and without contradiction.
- The eligibility degree: the character of a word that combines the necessary conditions to be elected as a concept since it is the most representative word of the cluster or the word that can orient the reasoning, the interpretation or the labeling task.

In the following paragraphs, we will present separately the three revealing criteria with some examples from our results.

A. The credibility degree criterion

In the previous section, we have explained the reasons of chosen web documents as a base to the evaluation task. Now, our challenge is how we can evaluate word cluster s after a decontextualization task?

The context granularities degrees. Our hypothesis is "Having words in the same context imply that they share common information that gives indications about the appropriate concept for these words". In this case a context is an appropriate support for a semantic interpretation. It limits the associated knowledge of each word and gives a background for the evaluation and labeling task. In order to explain this idea, we take the sentence: "the possible accommodations in the east region of USA are hotels and residences. When we limit the context to the association of 'hotels' and 'residences' by the conjunction 'and', we deduce that 'hotels' and 'residences' belongs to the same concept. However, when we limit the context to the entire sentence, we find that the associated concept to these two words is 'accommodation'. So, thanks to the contextualization task, we can deduce either the meaning of each word, the semantic association between some words or the concept associated to some words. Taking into account a static context i.e only one such as a sentence for all the word clusters is not sufficient since in some cases the sentence does not contain all the words of a cluster. That's why; our evaluation is not restricted to a unique context. On the contrary, it depends on various granularity levels which are applied and considered consecutively.

By defining several contexts, we adopt the concept of "Progressive contextualization" defined by the Professor Andrew P. Vayda in 1979 in another completely different context (to understand cause of damage and destruction of forest). We integrate this interesting concept in the evaluation process since it focuses on diversity and it looks at how different words operate in their contexts through a variety of documents structures, word's organizations, designer's conceptions and intentions, etc. The progressive contextualization rejects the assumption of using a unique context to understand an object. On the contrary, it assumes that ordinary speech or writing involves many contexts for each studied word and their interactions permit the right semantic interpretation. In our research, the several contexts defined from the domain web documents are provided by two sources. The first one is a linguistic analysis that gives us the various nominal groups and verbal groups. Also, it procures the various word associations by a preposition (of, on, etc.) or a co-ordinating conjunction (and, or, etc.). The second source is a documentary analysis that gives us the various sections of phrases (part of a phrase finished by a punctuation like ';' or ','), the sentences, the paragraphs and the documents. So, we have two types of contexts which are a linguistic context and a documentary context (Fig. 1: step 5 and 6). By using the first one, we obtain the close words of our target terms. The second context type is more generalized than the linguistic one. Consequently, the information deduced will be either complementary information or completely new information for the words of a cluster.

Our context definition is dynamic since it depends on the presence of the target words in each context. For example, for a cluster with four words and using two nominal groups, we find that these words are associated. So, we can give them a concept without looking for their documentary contexts. But, when the 8 words of a cluster do not belong to one of the four results of the linguistic contexts, we are obliged to look deeply into the documentary context. By the progressive contextualization, the expert evaluation is done by respecting this order for each word clusters: (1) Linguistic context: Nominal groups based context, Verbal groups based context Prepositional groups based context, Conjunctional groups based context; (2) Documentary context: Sections of phrase based context, sentences based context, paragraphs based context and documents based context.

For example, in the following sentence "the possible accommodations are hotels and residences", we find: "are hotels and residences" is the verbal group and "the possible accommodations" is the nominal group. The entire sentence which is a documentary context contains two linguistic contexts.

The Credibility Degree. Now, the problem is that the expert is enable, even when he is given all the results of the analyses, to find the best associations of each target word cluster. In order to facilitate this process, we define a semantic index which represents the credibility of the target words' association in relation with the different contexts. This index is named "credibility degree". It is computed for each word cluster and for each context definition in an automated way. Let us take the examples from our experiments to explain our idea:

Table 1. Examples of term clusters

Examples	Word Clusters
Example 1	academy, golf, golfer, club
Example 2	Civilization, archeology, ethnology, people
Example 3	Park, national, cliff, rock
Example 4	Cult, church, evangelization, memory, religious, sanctuary
Example 5	excursion, foot, person
Example 6	Hiker, gorges

Our « Credibility Degree Computation » algorithm is executed on a set of word clusters in order to compute their credibility degree. For instance, with the example 1 (Table 1) and according to one context definition (sentence), the algorithm finds all the possible combination in the context i.e tries to find the four words (academy, golf, golfer, club), then the association of three words and so on. For each found association, it presents the associated words and gives a degree representing the number of times this type of association is found. For example, with the same example, it finds two possible associations with three words which are {academy, golf, golfer} and {golf, golfer, club} so the credibility degree is 3_2 i.e two associations of three words.

Our algorithm has several functionalities which are:

- Finding the associations between some words in order to facilitate the labelling step. With the Example 5, our algorithm finds only the association that permits the user to give a label by himself like ‘excursion on foot’.
- Finding in the same time the available associations in the context and the concept (Example 2, the concept is ‘civilization’).
- Detecting the noisy elements in a cluster to delete them or move them to another cluster. For instance, in the Example 5, the word ‘hiker’ is found inside the several associations corresponding to ‘excursion’ and ‘foot’ but the problem is that this word belongs to the Example 6. Our algorithm decides to remove it from the example 6 to the example 5 because no association is found with ‘gorges’. If the words exist in two associations related to two clusters, the algorithm writes the word using the red colour to announce a possible ambiguous situation.
- Enhancing a cluster by other words from the associations. For Example 3, we can enrich the group by the word “nature” and “space” and find the concept which is ‘natural space’.

Since our evaluation task is based on various context definitions, if the user does not find a connection between some words by using the linguistic contexts, he can look to the other associations provided by the documentary contexts. In these contexts, the probability to find more relation is bigger than with the first context type (Example 4).

Thanks to the credibility degrees computed for each cluster and for each context, the user obtains useful information and, in some cases, sufficient to manipulate (delete word, remove word, etc.), evaluate and label the cluster. For example, for a same cluster, if he finds the three credibility degrees (5_1, 4_3, 3_8, 2_{15}), he starts by analysing the association with 5 words. If it is not sufficient, he analyses the three associations of four words and so on. If the information returned by our algorithm to this cluster and for one context is not enough, he can look to the other credibility degrees provided by the other contexts by respecting the previous order (linguistic context type then documentary context and inside the same context analysis, from the highest degrees to the lowest ones).

Table 2. Examples of term clusters with some nformation

Cluster	Ref_Word	Doubt_Word	Partial N-Hits	Total N-Hits
Cluster 1 : {relaxation, remise, forme, détente, santé, oiseau}	relaxation	oiseau	222000	55600*6 = 333600
Cluster 2 : {concentration, équipement, province, vignoble, berceau}	vignoble	berceau	151000	46750*5 = 233750
Cluster 3 : {mollusque, ciel, mer, marin}	mollusque	ciel	192000	97100*4 = 388400

In this paragraph, we explain the sections 2 and 3 of our algorithm by giving some examples from our results. (Remark: we maintain the examples in French language since in some cases when we translate the word the meaning of the word or of the group changes)

In the section 2 of our algorithm, we explore the results of the section 1 in order to deduce some important knowledge which are: the most present associations, the most frequent words of the clusters that appear in the contexts, the Ref-Word and the Doubt-Words. The most present associations are the first fifty associations which are the most repeated ones. The most frequent words of the cluster that appear in the contexts permit to have an idea about the relevance of each word in the cluster. The expert can analyse these words because they are presented in an order from the most frequent one to the least frequent one. So it is another manner to present the word clusters but with a useful information. The Ref-Word is the word (inside a cluster) that could be either the appropriate concept or a word that gives more information about it. It is the most present word inside the different associations collected from the different contexts. For instance, in the cluster 2 (Table 2), the Ref-Word is ‘vignoble’. The Doubt-Words are those that the expert can delete them, replace them by other ones or move them from this cluster to another one. They are selected because they do not exist in the different associations resulting from the various analyses (linguistics and documentary). Generally, if the Doubt-Words are numerous, we select only the two first because we suppose that some of them can exist but in other contexts which are absent in our analyses. For instance, in the cluster 1, the Doubt-Word is ‘oiseau’.

In the section 3 of the CDC algorithm, we exploit this information in order to better help the domain experts. For this purpose, we use the Ref-Word and the Doubt-Words. We try to decide if we should delete the Doubt-Word, keep it or move it to another cluster. In order to explain how our algorithm proceeds, we comment the Table 2. The first step of our algorithm is to take into account the Ref-Word and one of the Doubt-Words. Then, it deletes the two last letters from the Doubt-Word and tries to search it with the Ref-Word in the Google database. If Google search engine corrects the word and finds the initial complete Doubt-Word, the CDC algorithm decides to keep the word. For example, if we consider the cluster containing “île” and “archipel” and we suppose that the word “archipel is a doubt-word. When we search the six first letters of “archipel” with “île, the search engine proposes the word “archipel” instead of “archip”. In this case, the word archipel is kept in the cluster.

Google search engine can detect the Doubt-Word even if it is incorrectly written and this is du to its association with the Ref-Word.

In some other cases, Google search engine does not correct the word. For this, we try to search the Ref-Word and the Doubt-Word (without any modification) and take the number of hits (named ‘Partial-N-Hits’). In the same time, we compute the number of hits related to all the words inside the cluster which

ISBN:978-988-98671-5-7

is named 'N-Hits'. To obtain the 'Total-N-Hits', we multiply the N-Hits by the number of words in the cluster. Then we compare the Partial-N-Hits with the Total-N-Hits. Three cases exist:

- If partial-N-Hits < Total-N-Hits, we have the choice between moving the Doubt-Word or delete it:

-if we find the Doubt-Word in other contexts related to a different cluster, we decide to move the Doubt-Word to this cluster. For example, we take the cluster 1 and we find that 'oiseau' is found in the contexts related to the cluster 8 :{zoo, animal, faune}. In this case, the CDC algorithm proposes to move the word 'oiseau' to this cluster.

- Else we delete the Doubt-Word. For example, within the cluster 3, the algorithm decides to delete the word 'ciel'.

- If partial-N-Hits > Total-N-Hits, then the CDC algorithm decides to keep the word because the relationship between the Ref-Word and the Doubt-Word is close. A good example is the cluster 2, in which we decide to maintain the word 'berceau' inside the cluster. Semantically, in French language, these two words are frequently used together.

B. The Cohesion degree criterion

Our idea is based on the fact that there is a big amount of web pages which are indexed by Google (8 058 044 651). For each domain, site designers explain differently the domain's knowledge according to their needs, objectives, geographic position, etc. Consequently, the web represents the knowledge provided by several persons or organisms that have different experiences in this field. In this section, we are interesting on computing the cohesion degree of each word cluster. That's why, we define a semantic distance between a set of words, based on the Google database. This latter is the documents indexed by Google. Our measure named Cohesion Degree Criterion and noted Coh-D, uses the number of documents in which the words occur together. Our criterion is defined after some experiences and modifications based on the normalized Google distance [2]. The Google distance is only used for two words. In our case, the cluster can contain more than two words. We present the defined formula:

Coh-D = (Min (log (NBH (Wi)))/Max log (NBH (Wj)) * (NBH ({W1, W2, W.., Wj}))

- NBH (Wi) is the number of hits returned by google when we search the word Wi, Wi is one word from the cluster {W1, W2, W.., Wj}.

- NBH ({W1, W2, W.., Wj} is the number of hits related to all the words of a cluster.

Table 3. Some results with their cohesion degrees

N°	Word cluster (French language)	label	Cohesion_ Degree
1	académie, club, golfeur, golf	Golf	37135
2	bonheur, reste, touriste camping, caravane,	Camping	15759,59 / 131266,62
3	allant, canotage, passant, population, spécimen, échelle, réserve, vue	Unknown	368,2
4	chapelle, église, évangélisation, génération, mémoire, religieux, culte, lieu, sanctuaire	religion	883,21 / 21413,76
5	avion, banquise, musée, recherche, semaine offerte,	Unknown	518,69
6	boîte, clientèle, francophone, métropole, ville	service	76842,01
7	administration, attention, éducation, particulier, préoccupation, cervidés, faune, habitat, utilisateur	administr ation	223,52 /81433,26
8	brochet, feuille, jaune, doré, saumon	Fish	1257,59 / 34997,49

When evaluating the word cluster, domain experts note that there are three types of clusters which are advisable clusters, improper clusters and unknown clusters [5]. Advisable clusters are those for which the expert could associate a label and in which words belonging to the same group are close semantically. Improper clusters are clusters where either there is an amount of words without any relation with the principal concept extracted from this cluster or that this cluster contains more than one concept clearly remarked. Unknown clusters are clusters where words do not have any semantic relation and the expert could not find any semantic interpretation. For instance, the cluster number 1 and 6 (Table 3) are two advisable clusters since all the words are associated to the same concept. Clusters 3 and 5 (Table 3) contain noisy words (words that must not belong to the cluster). That's why the experts either delete many words in order to find a concept or cannot find a concept. For the other cases, the cluster can contain some noisy words but that are known easily ('feuille' in the cluster 8). The question is how we detect the word that could be deleted. By using Google search engine, we compute the number of hits of each word belonging the cluster associated to the name of domain (in this case is "tourism"). The word that appears the least is deleted and we look this effect on the cohesions degree of the new cluster. Other examples are present in the Table 2. The bold words are those that when we delete them, the cohesion degrees are visibly improved. For instance, in the cluster 7 (Table 3), by deleting the word "cervidés" we obtain a better cohesion degree (a real impact since the cohesion degree move from 223.52 to 81433.26) and consequently an easy evaluation task.

The cohesion degree is a quantitative criterion that helps the domain expert or a novice user to see if the cluster is semantically coherent or not. But the evaluator cannot make a judgment only by using this criterion. This latter is helpful when it is considered with the other ones (credibility and eligibility criteria).

C. The Eligibility degree criterion

The eligible degree represents the voting weight that indicates to the user: what is the candidate that can probably represent the cluster or initiate the reasoning process.

It is calculated on the base of the following formula:

$$ED = \mathrm{Min}\left(\bigvee_{i=1}^{n} x_i \;\middle|\; \left(\mathrm{Somme}_{i=1}^{n}\, NT(x_i)/n\right) - NT(x_i)\right|\right)$$

NT(x) is the number of the occurrences of x

n is the number of word in the cluster

The word that obtains a value of ED that is the most near to the average is the eligible candidate and can represent the cluster. The eligibility degree is computed only for the improper and advisable clusters.

The Table 4 shows some results.

Table 4. Some results with the eligibility degrees

N°	Word Cluster	E-D-Word	Quant-E-D	Ref-Word
1	académie, club, golfeur, golf	golf	59144750	golf

2	bonheur, reste, touriste camping, caravane,	caravane	42828000	camping
3	falaise, mètre, national, rocher, parc	parc	760824000	parc

For the cluster 1, our criterion finds the right word "golf" that represents the cluster. When the word chosen by the eligibility criterion and the one defined by the credibility criterion (Ref-Word) are identical, the algorithm presents the E-D-Word in a bold style to explain a strong suggestion. The cluster 1 is an advisable one. The cluster number 3 (Table 4) contains a noisy word that is "mètre". After removing it, our algorithm detects the right word "parc". For the cluster 2, the algorithm proposes "caravane" as a possible concept for the cluster. The Ref-Word of the cluster 2 (Table 4) is "camping". Since the words are different, the algorithm shows them without any more information. In this case, the evaluator should make the choice between them or define another one. But in the two first cases (1 and 3 (Table 4)), the evaluator can easily decide.

D. Discussion

Our web driven concept evaluation method provides three revealing criteria that help either the domain expert or a novice. Our method presents some indications which are the qualitative ones based on the word associations deduced from the various contexts and the quantitative ones resulting from the computed credibility degree index, the cohesion degree and the eligibility degree. The qualitative evaluation provides a semantic support for an easy interpretation. Moreover, our proposition, based on a large collection of domain web documents and several contexts definitions with different granularity degrees, permits to an ordinary user to help the expert by manipulating the word clusters and giving him semantic tags as suggestions. Consequently, the expert should decide on the appropriateness of these labels as well as clusters homogeneities which are not labeled. Moreover, our proposed algorithm assures the ontology reuse and evolution since the elements on which the expert's interpretations are based (the provided word associations) depend on the web changes. For example, when the web documents change, the various extracted contexts change too and the results of the mapping operation between the words belonging to the clusters and the contexts are updated. This resulting information about the word clusters is presented to the experts in order to help them during the evaluation task. Then, they are stored with the experts' comments in order to be reused by another expert either during the same period or later (after some months or years depending on the frequency of updates). So, our ontological concept evaluation method helps the user to understand the sense of a set of words in order to evaluate and label it.

V. CONCLUSION

Ontology evaluation task is not evident. In this paper, we have argued for the need of new evaluation methods that do not depends only on the gold standard (thesaurus or ontology) or on the human intervention. We have explained why it is difficult to evaluate data knowledge. Then, we have proposed a new evaluation method that helps either an ordinary user (knowledge engineer) or the expert to take the write decision about the semantic homogeneity of a cluster. In order to achieve this purpose, we have defined a new method based on three revealing criteria which are the Credibility Degree, the cohesion degree and the eligibility degree. Each criterion is computed by a separate algorithm. The credibility degree algorithm tries to eliminate or remove the noisy elements, propose some semantic tags and give several word associations. The cohesion and eligibility degree informs the user about the relationship between the words belonging to one cluster and which one can represent the cluster as a concept. Our automatic method guides the expert to an easier interpretation of the word cluster and avoids the ambiguous cases. Future research in this area should seek to develop further techniques for evaluating the other elements of ontology such as the relations between the concepts.

REFERENCES

[1] Brewster, C., Alani, H., Dasmahapatra, S. and Wilks, Y., Data Driven Ontology Evaluation, Proceedings of Int. Conf. on language resources and evaluation, Lisbon, 2004.
[2] Cilibrasi, R. and Vitanyi, P. Automatic Extraction of Meaning from the Web. IEEE International Symposium on Information Theory, Seattle, Washington, 2006.
[3] Faure, D., Nedellec, C. and Rouveirol, C. (1998). Acquisition of semantic knowledge uing machine learning methods: the system ASIUM. Technical report number ICS-TR-88-16, inference and learning group, University of Paris-sud.
[4] Holsapple, C. and Joshi, K.D.: A collaborative approach to ontology design. Communications of ACM, 45(2): 42-47, 2005.
[5] Karoui, L., Aufaure, M-A., Bennacer, N.: "A New Extraction Concepts based on Contextual Clustering". To appear in the IEEE International Conference on Computational Intelligence for Modelling, Control and Automation – CIMCA06, 2006.
[6] McCarthy, J., Notes on formalization context. Proceedings of the 13th IJCAI, Vol. 1, 1993, pp. 555-560.
[7] Meadche, A. and Staab S. : "Ontology learning for the semantic Web, IEEE journal on Intelligent Systems, Vol. 16, No. 2, 72-79, 2001.
[8] Meadche, A and Staab, S., Measuring similarity between ontologies. Proc. CIKM 2002. LNAI vol.2473.
[9] Navigli, R., Velardi, P., Cucchiarelli, A. and Neri, F.: Quantitative and qualitative evaluation of the ontolearn ontology learning system. In Proc. Of ECAI-2004 Workshop on ontology learning and population, Valencia, Spain, Aug.2004.
[10] Spyns, P., et al., Evalexon: Assessing triples mined from texts. Technical report09, STAR Lab, Brussels, Belgium, 2005.
[11] Thomas R. Gruber (1993). Toward principles for the design of ontologies used for knowledge sharing. Originally in N. Guarino and R. Poli, (Eds.), International Workshop on Formal Ontology, Padova, Italy. Revised August 1993. Published in International Journal of Human-Computer Studies, Volume 43 , Issue 5-6 Nov./Dec. 1995, Pages: 907-928,
[12] Vazirgiannis, M., Halkidi, M. and Gunopoulos, D.: uncertaintly handling and quality assessmen in data mining. Springer, 2003.