



HAL
open science

Vers des pratiques collaboratives pour les systèmes d'organisation de connaissances

Amel Fraisse, Patrick Paroubek

► **To cite this version:**

Amel Fraisse, Patrick Paroubek. Vers des pratiques collaboratives pour les systèmes d'organisation de connaissances. 10ème colloque international du Chapitre français de l'ISKO – Systèmes d'organisation des connaissances et humanités numériques, International Society of Knowledge Organisation, Nov 2015, Strasbourg, France. hal-01617191

HAL Id: hal-01617191

<https://hal.science/hal-01617191>

Submitted on 21 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vers des pratiques collaboratives pour les systèmes d'organisation de connaissances

FRAISSE Amel

LIMSI-CNRS

Bât. 508 Université Paris-Sud, 91403 Orsay Cedex, France.

PAROUBEK Patrick

LIMSI-CNRS

Bât. 508 Université Paris-Sud, 91403 Orsay Cedex, France.

Résumé. Dans cet article nous proposons un nouveau paradigme de gestion collaborative et incrémentale pour la construction de connaissances basé sur la participation de plusieurs types d'utilisateurs (annotateurs non professionnels, des simples utilisateurs, etc.). Ainsi nous avons mené deux expériences: la première basée sur le *crowdsourcing* et la deuxième fait participer des simples utilisateurs à la construction de connaissances via la mise en place d'un jeu à but appelé *Climate Challenge*. Les résultats de nos deux expériences montrent que la qualité obtenue est équivalente à celle qui serait obtenue par des experts et que les coûts et les délais sont réduits.

Mots-clés: Système d'organisation de connaissances, crowdsourcing, jeu à but, approche collaborative.

1. Introduction

La problématique de l'usage et de la conception de systèmes d'organisation des connaissances (SOC) adaptés à la gestion de l'information documentaire numérisée qui ne concernait jusqu'alors qu'un petit nombre de spécialistes tend à devenir un enjeu majeur de la gestion de l'information numérique. En effet, les langages et les index qui organisent l'information, sont devenus des technologies intégrées aux dispositifs de stockage et aux usages de recherche de l'information conditionnant de manière déterminante les modalités d'accès à celle-ci dans les contextes professionnel et de loisir. Depuis les systèmes à base de simple liste de termes (glossaires, dictionnaires, etc.) jusqu'à ceux à base de thésaurus et d'ontologies, les SOC suscitent de plus en plus de travaux dans le domaine des humanités numériques.

Jusque là et quelque soit le type du système d'organisation de la connaissance, ce dernier est conçu et mis en place par des spécialistes de la connaissance et du domaine. Cependant avec l'avènement du Web 2.0, de nouvelles pratiques d'organisation de la connaissance sont apparues. En effet, les utilisateurs ne se contentent plus de la génération du contenu mais ils participent aussi à son indexation. C'est le cas par exemple des mot-dièses sur le réseau social Twitter qui sont utilisés comme des métadonnées afin d'indexer et faciliter l'accès aux messages.

Dans cet article, nous proposons un nouveau paradigme de gestion collaborative et incrémentale de l'information numérique qui est basé sur la participation de différents types d'utilisateurs. Ce dispositif consiste à faire participer des simples utilisateurs, via un jeu à but ou un site collaboratif, à la construction d'une base de connaissances factuelles et affectives.

2. Les systèmes d'organisation de connaissances

Les systèmes d'organisation de connaissances permettent de fournir une meilleure représentation de l'information qui est utile aux systèmes de recherche d'information.

Les méthodes de représentation de l'information peuvent être divisées en deux grandes catégories : celles qui sont basées sur l'abstraction de

sujets des documents et celles basées sur l'indexation du contenu des documents ; qui consistent à extraire depuis le document un ensemble de mots clés décrivant son contenu (Cleveland, 2001), (Lancaster, 2003). Différents types de SOC basés sur des méthodes d'indexation ont été développés comme par exemple les thésaurus, les classifications et les nomenclatures qui contiennent un vocabulaire contrôlé pour indexer les documents.

Plus récemment, les SOC incluent deux nouveaux modèles de représentation de la connaissance : les ontologies et les folksonomies (Weller, 2007). Les ontologies permettent d'étendre la structure formelle d'un vocabulaire en incluant plusieurs types de relations (synonymie, antonymie, etc.) entre les concepts avec un niveau de rigueur formelle variable.

Les folksonomies intègrent une nouvelle dimension sociale en donnant la possibilité à l'utilisateur d'indexer des documents afin qu'il puisse plus aisément les retrouver grâce à un système de mots-clés (appelés aussi Tag) (Marlow, Naaman, Boyd & Davis, 2006).

Ces différents systèmes d'organisation de connaissances peuvent être classés en fonction de deux aspects inversement proportionnels à savoir la complexité de la structure et la couverture du SOC. En effet, plus la structure est complexe plus la couverture de domaines est restreinte.

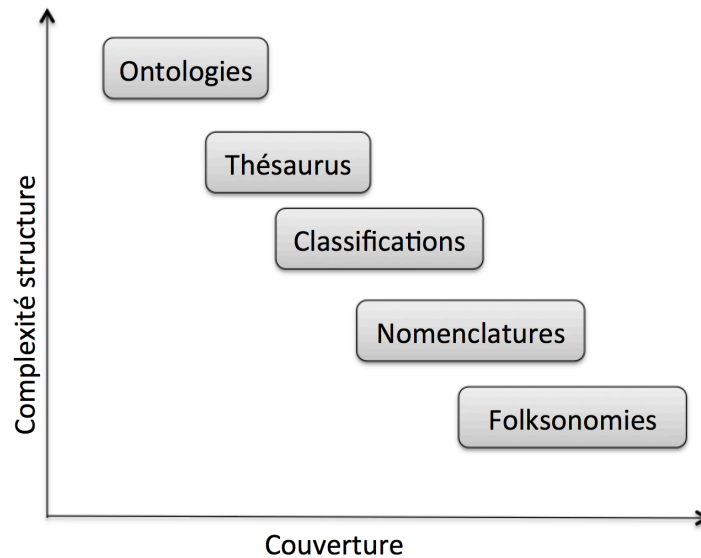


Figure 1: Classification des SOC selon la complexité de structure et la couverture

3. Le crowdsourcing pour la construction collaborative des connaissances

Le terme crowdsourcing est apparu en 2005 (Safire, 2009) et il a été adopté rapidement par la communauté scientifique. Dès son apparition, beaucoup de chercheurs se sont intéressés à étudier et à évaluer les bénéfices et les gains qu'ils pourraient obtenir de l'utilisation d'une telle approche. La plupart de premières études présentent dans la littérature mentionnent Amazon Mechanical Turk (AMT), avec CrowdFlower qui est apparu un peu plus tard.

Notre étude sur les travaux antérieurs qui ont expérimenté le crowdsourcing est organisée selon les thématiques de recherche suivantes:

- Sémantique
- Fouille d'opinions et analyse de sentiments
- Entités nommées
- Traduction automatique
- Recherche/Extraction d'information & questions réponses

3.1. Sémantique

Dans (Rumshisky et al., 2009), les auteurs ont expérimenté la construction d'une base de données contenant les sens flous d'un ensemble de verbes polysémiques en faisant appel à des annotateurs non-experts. Les verbes sont affichés aux annotateurs dans leur contexte d'apparition (phrases) et ces derniers annotent les différents sens du verbe.

Les résultats ont été ensuite évalués en les comparant à un ensemble d'annotations qui ont été effectuées par des lexicographes professionnels. La comparaison entre les deux ensembles présentait une valeur de F-score de 0.93.

Dans un travail plus récent (Rumshisky, 2011), l'auteur conclut que le regroupement (clustering) de 350 lignes de concordance par groupe de sens similaires n'est fiable que pour seulement environ 140 lignes de concordance. Dans (Rumshisky et al., 2012), les auteurs expliquent que l'un des problèmes majeurs est d'attirer des annotateurs avec un bon niveau de performance sur AMT pour effectuer des tâches linguistiques complexes. L'utilisation d'un guide de bonnes pratiques peut améliorer la qualité des annotations mais cela ne substitue pas le besoin d'effectuer des expérimentations préliminaires afin de calibrer les paramètres de la tâche d'annotation et de l'interface.

Dans (Munro et al., 2010), les auteurs ont étudié l'utilisation de AMT pour un ensemble d'expériences linguistiques allant de la sémantique à la psycholinguistique traitant :

- les verbes sémantiques,
- la segmentation de séquences audio,
- le modèle de langage,
- la grammaire du discours,
- le rôle thématique,
- le traitement de la métaphore par le cerveau et,
- le degré de concentration en lecture

Les auteurs mentionnent une valeur de Kappa de 0.9 entre les données annotées sur AMT et celles en laboratoires pour la tâche du modèle du langage. En se basant sur les résultats de leurs expériences, les auteurs

concluent que le crowdsourcing permet l'exécution systématique et à grande échelle des études avec des coûts pas trop élevés et de façon plus simple que celle effectuée dans des conditions de laboratoires.

3.2. Recherche et extraction d'information

(Grady and Lease, 2010) ont étudié l'impact de plusieurs facteurs et paramètres sur la tâche de jugement de pertinence en recherche d'information. Les auteurs ont étudié l'impact des quatre facteurs ci-dessous sur le coût, le délai ainsi que la qualité des évaluations :

1. Fournir aux annotateurs uniquement un titre pour la requête au lieu d'une description détaillée.
2. Différents étiquettes pour le titre de la tâche (spécialisée versus générique (oui/non)).
3. La rémunération de la tâche (0.01\$ ou 0.02\$).
4. Le bonus (0 ou 0.02\$).

Les évaluations ont été effectuées sur des documents issus du corpus TREC TIPSTER, qui contient des articles journalistiques. Au total, 5 lots d'évaluations ont été créés. Pour chacune de quatre thématiques du corpus, cinq documents ont été évalués et 10 évaluations ont été collectées pour chaque document. Au total, 200 tâches ont été soumises à AMT pour chaque lot d'évaluation (1000 tâches pour les 5 lots d'évaluations). La taille d'un document est entre 162 et 2129 mots. Pour chaque tâche, l'annotateur doit associer un seul lien de pertinence entre une requête et un document.

Dans le lot d'évaluation 5, 23 bonus ont été donnés pour un total de 0.46\$. La seule différence observée par rapport aux autres lots est la longueur de commentaires donnés par les annotateurs. La meilleure valeur d'exactitude était de 70.5% et atteinte par le lot 3, qui utilise des titres de requête (au lieu d'une description détaillée) et des réponses simples (yes/no).

Les documents évalués par les annotateurs du lot 5 ont été mieux évalués que le reste des documents (3.37 documents correctement annotés contre 2.10-2.20 pour les lots d'évaluations 1 à 3 et 1.85 pour le lot 4).

Au final d'après les auteurs, les résultats de leur expérimentation sont peu concluants.

3.3. Fouille d'opinions et analyse de sentiments

Dans (Hsueh et al., 2009), les auteurs ont utilisé AMT pour classifier des segments de texte extraits depuis des blogs politiques selon leurs polarités (positif, négatif, mixte ou neutre). Chaque segment de texte a été annoté par trois annotateurs. Les auteurs mentionnent une valeur de Kappa entre les annotateurs de 70.4%.

Une tâche d'annotation plus fine pour identifier le terme qui exprime une opinion/sentiment par rapport à un terme cible présent dans la phrase a été expérimentée par (Sayeed et al., 2011) sur CrowdFlower. Les annotateurs doivent classifier en positif, négatif ou neutre le mot d'opinion présélectionné en fonction de sa relation avec le terme cible présélectionné dans la même phrase. L'expérience comprend 200 tâches payées 0.04\$ la tâche. Chaque tâche est annotée par trois annotateurs différents pour un coût total de 60\$ et un délai de 24h pour toute l'annotation. Afin, d'identifier les annotateurs douteux, les auteurs ont utilisé 30 tâches comme données de référence. Ainsi, seulement 22 annotateurs ont réussi le test de contrôle de qualité de CrowdFlower. La suppression des annotateurs douteux (en se basant sur les données de référence du test) avait un effet positif sur les deux valeurs F-mesure et Kappa. La meilleure valeur de Kappa est de 0.65 obtenue lors de la suppression de 7 annotateurs douteux. Alors qu'une valeur de précision de 0.85 et une valeur d'exactitude de 0.88 ont été atteints après la suppression de 10 mauvais annotateurs. Cette expérience montre qu'il est possible d'utiliser le crowdsourcing pour des tâches d'annotation fine mais cela nécessite la mise en place d'une procédure de contrôle de qualité.

3.4. Les entités nommées

Dans (Finin et al., 2010), les auteurs décrivent leur expérience d'annotation en entités nommées des messages Twitter (organisation, personne, lieu, aucune). Au total 251 tâches ont été envoyées sur AMT et chaque tâche a été annotée par deux annotateurs. Le délai global pour chaque tâche est d'environ 15 heures et le coût est de 27.61\$, qui

correspond à 0.0275\$ par tweet. Au total 42 annotateurs ont participé à l'annotation et la plupart ont été localisés aux Etats-Unis, Inde et Australie. À la fin de la procédure d'annotation, l'accord inter-annotateur a été calculé en utilisant un algorithme proche de PageRank (Page et al., 1999). La performance des annotateurs de AMT a été jugée inférieure à celle des annotateurs experts mais il est possible d'atteindre des performances équivalentes en combinant les annotations de plusieurs annotateurs.

3.5. La traduction automatique

Dans (Zaidan and Callison-Burch, 2011), les auteurs ont expérimenté la tâche de traduction de l'urdu vers l'anglais. Le prix a été fixé à 0.10\$ par traduction (environ 0.005\$ par mot). Pour chaque phrase les auteurs ont collecté quatre traductions proposées par différents traducteurs de AMT, ensuite ils ont créé un ensemble de tâches de validation croisée de traductions. Les tâches de validation croisée consistent à trier les traductions par ordre de qualité ou à post-éditer pour corriger certaines traductions. Les coûts globaux de l'annotation étaient de :

- coût de la traduction : 716.80 \$,
- coût de la post-édition : 447.50\$,
- coût du tri de traductions : 134.40\$

Au total, 7000 traductions ont été effectuées, 17000 post-édition et 35000 tri de traductions. En incluant le coût de la traduction professionnelle de la référence qui est de 1000\$ ainsi que les charges de Amazon, le coût global de cette expérience était de 2500\$.

Les auteurs ont étudié la possibilité de réduire le coût en éliminant la traduction professionnelle des données de référence et en augmentant le nombre de post-édition. Une première mesure obtenue montre un score BLEU de 34.86 alors qu'une deuxième mesure permettant de réduire les coûts tout en garantissant une bonne performance montre un score BLEU de 38.67. Ainsi la meilleure stratégie pour la tâche de traduction est la suivante :

- produire plusieurs traductions possibles pour chaque phrase,
- trier plusieurs traductions,

- éditer uniquement les meilleures traductions

4. Notre approche pour la construction collaborative de connaissances factuelles et affectives

Notre approche se base sur deux étapes principales : la collecte de documents pour la construction d'un corpus textuel multilingue (français, anglais et allemand) suivie de l'extraction de connaissances.

4.1. Collecte du corpus textuel

Dans le cadre du projet uComp¹, nous nous sommes intéressés à la constitution d'une base de connaissances factuelles et affectives relatives au domaine du changement climatique. Ainsi, nous avons constitué une liste fermée de mots clés qui décrivent au mieux le domaine de changement climatique. Nous avons utilisé ensuite cette liste pour collecter depuis le Web des articles journalistiques ainsi que des messages issus de médias sociaux dans les trois langues suivantes : anglais, français et allemand (Weichselbraun et al., 2013). Le Tableau 1 décrit les caractéristiques du corpus collecté dans les trois langues.

| | Anglais | Français | Allemand |
|----------------------------|------------|----------|----------|
| # Articles journalistiques | 1 275 000 | 720 000 | 650 000 |
| # messages médias sociaux | 20 000 000 | 410 000 | 565 000 |

Tableau 1: Caractéristiques du corpus collecté

4.2. Extraction des termes affectifs candidats

Dans un travail antérieur (Fraisse et Paroubek, 2014), nous avons proposé une méthode qui permet d'associer de façon automatique à chaque document affectif sa classe d'émotion (par exemple : joie, tristesse, amour, plaisir, peur). Ainsi, afin d'extraire la liste des termes affectifs candidats pour chaque classe d'émotion nous nous sommes basés sur une approche statistique, qui consiste à extraire, à partir du

¹ uComp : Embedded Human Computation for Knowledge Extraction and Evaluation
<http://www.ucomp.eu/>

corpus émotionnel collecté (décrit dans la Tableau 1), et pour chaque classe d'émotion, l'ensemble de mots qui lui est associé. Afin de mesurer l'association entre un mot m du corpus et une émotion e , nous nous sommes basés sur la mesure de l'information mutuelle introduite par (Fano, 1961), (Church and Hanks, 1990), qui pour chaque couple de variables aléatoires $(X; Y)$ calcule leur degré de dépendance au sens probabiliste. L'information mutuelle est donnée par la formule suivante :

$$IM(X, Y) = \log_2 \left(\frac{P(X, Y)}{P(X) \cdot P(Y)} \right)$$

Ainsi, dans notre cas, il s'agit de mesurer le degré de dépendance entre un mot m et une émotion e .

$$IM(m, e) = \log_2 \left(\frac{freq(m, e)}{freq(m) \cdot freq(e)} \right)$$

$freq(m, e)$ est le rapport entre le nombre de messages étiquetés e contenant le mot m ($|T_{m,e}|$) et le nombre total de messages ($|T|$).

$$freq(m, e) = \frac{|T_{m,e}|}{|T|}$$

$freq(m)$ est le rapport entre le nombre total de messages contenant le mot m ($|T_m|$) et le nombre total de messages.

$$freq(m) = \frac{|T_m|}{|T|}$$

$freq(e)$ est le rapport entre le nombre total de messages étiquetés e ($|T_e|$) et le nombre total de messages.

$$freq(e) = \frac{|T_e|}{|T|}$$

4.3. Validation des termes affectifs candidats

Afin de valider les termes extraits de façon automatique, nous avons mené deux types d'expériences de validation. La première se base sur une approche de crowdsourcing et la deuxième consistait à mettre en place un jeu à but.

4.3.1 Validation par le crowdsourcing

Nous avons mené une expérience de crowdsourcing pour valider la liste des termes affectifs français (9939 termes). Chaque tâche d'annotation est constituée de 8 entrées lexicales (termes). Pour chaque entrée, l'utilisateur doit répondre aux quatre questions décrites dans la Figure 2.

ignorance

Ce terme peut-il être utilisé pour exprimer une opinion, un sentiment ou une émotion ?

Oui

Non

Ce terme peut-il déclencher une opinion, un sentiment ou une émotion ?

Oui

Non

Ce terme a-t-il une connotation

Positive

Negative

Neutre

Quelle est l'opinion, le sentiment ou l'émotion que ce terme peut exprimer ou déclencher ?

AMOUR

APPAISEMENT

DERANGEMENT

ENNUI

PLAISIR

DEPLAISIR

TRISTESSE

PEUR

COLÈRE

SURPRISE POSITIVE

SURPRISE NEGATIVE

SATISFACTION

INSATISFACTION

VALORISATION

DEVALORISATION

ACCORD

DESACCORD

DÉGOÛT

AUCUNE (TERME NEUTRE)

Figure 2 : Description de la tâche d'annotation

Afin de contrôler la qualité des réponses des utilisateurs, nous avons défini une question test qui permet d'attribuer un niveau de confiance à chaque annotateur et d'identifier les ceux qui sont douteux. CrowdFlower reparti les utilisateurs en trois niveaux en fonction de leur maîtrise de la langue (Good, Great, Best). Dans un premier temps, nous avons choisi le meilleur niveau (Best) et nous avons fixé le prix de la

Vers des pratiques collaboratives pour les systèmes d'organisation de connaissances

tâche à 6 centimes. Cela nous a permis de valider uniquement 9,1% des termes. Ainsi, nous étions amené à modifier la configuration de la tâche en sélectionnant des annotateurs avec le premier niveau (Good). En une semaine tous les termes étaient validés.

4.3.2 Jeu à but : Climate Challenge

Nous avons mené une deuxième expérience de validation collaborative en mettant en place un jeu à but que nous avons appelé Climate Challenge² (jeu en cours d'expérimentation).

Dans un premier temps, la tâche de validation consiste à attribuer à un terme donné une valeur de polarité parmi les 5 valeurs suivantes : très négatif, négatif, neutre, positif, très positif (Figure 3).

Afin d'avoir un niveau de confiance assez élevé, le même terme est validé par plusieurs joueurs. Chaque joueur dispose d'un score qui indique son niveau dans le jeu: ce score est calculé en fonction du nombre total de termes validés par le joueur ainsi que le nombre de bonnes réponses.

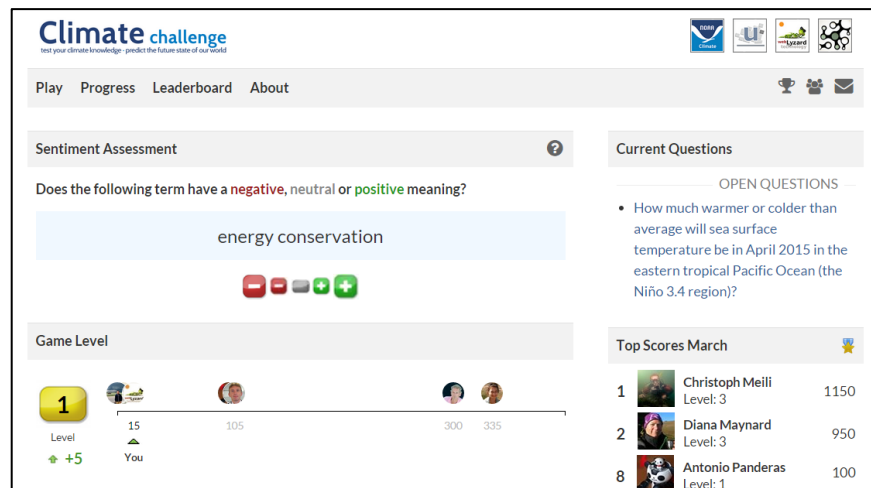


Figure 3: Interface du jeu à but Climate Challenge

4.4. Constitution collaborative d'une ontologie du domaine

Dans le cadre du projet uComp, Wohlgenannt et al. (2013), se sont basés aussi sur des approches statistiques afin d'extraire une première

² <https://www.ecoresearch.net/climate-challenge/>

liste de termes candidats à partir du corpus textuel collecté. En utilisant de ressources linguistiques externes comme WordNet ils ont établi des relations (synonymie, antonymie, etc.) entre les concepts collectés. Pour valider les concepts et les relations qu'ils ont construit ils ont mis en place un plugin Protégé³ qui permet de créer des tâches de validation unitaires et de les déployer directement sur des sites d'annotation collaborative comme CrowdFlower (Figure 4). Ainsi, les utilisateurs peuvent accomplir certaines tâches en validant et/ou proposant des termes ou des liens sémantiques entre les termes de l'ontologie (Hanika et al., 2014). Le résultat de cette approche a montré un score d'exactitude qui varie entre 89% et 99%, que le temps passé est équivalent à celui d'un expert du domaine et que le coût de production de l'ontologie est réduit de 75% à 81%.

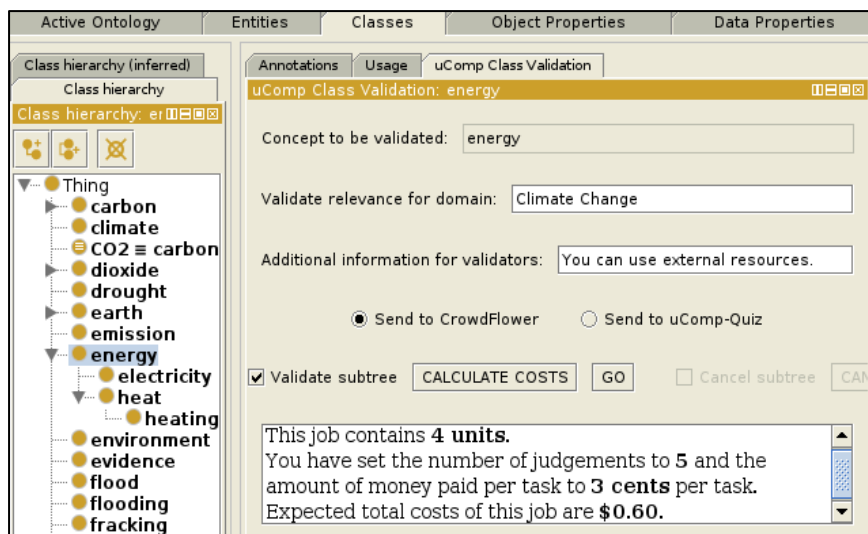


Figure 4: Interface du plugin Protégé pour la validation des termes de l'ontologie par les utilisateurs

5. Conclusion

Pour conclure, nous avons proposé une approche basée sur la participation collaborative de différents types d'utilisateurs (simple utilisateur, experts du domaine, ingénieur de connaissance, etc.). Cette

³ <http://protege.stanford.edu/>

Vers des pratiques collaboratives pour les systèmes d'organisation de connaissances

approche se base sur l'utilisation de jeu à but ainsi que le crowdsourcing. Nos deux expérimentations montrent que la qualité obtenue est équivalente à celle qui serait obtenue par des experts de la connaissance et les coûts et les délais sont réduits. Pour la suite de nos travaux, nous souhaitons évaluer la généralité de notre approche en l'appliquant à d'autres domaines et à d'autres types de documents. Nous envisageons aussi d'évaluer la performance de systèmes de recherche d'information qui se basent sur des SOC construits de façon collaborative.

Références

- Cleveland, D. B. and Cleveland, A. (2001) Introduction to indexing and abstracting, Englewood, CO: Greenwood Press.
- Fraisse, A. and Paroubek, P. (2014) Toward a Unifying Model for Opinion, Sentiment and Emotion Annotation and Information Extraction, *in proceedings of (LREC 2014) The 9th International Conference on Language Resources and Evaluation*, pages 3881-3886. May 26-31, 2014. Reykjavik, Iceland.
- Finin, T., Murnane, W., Karandikar, A., Keller, N., Martineau, J., and Dredze, M. (2010) Annotating named entities in twitter data with crowdsourcing, *in proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pages 80-88, Los Angeles. Association for Computational Linguistics.
- Grady, C. and Lease, M. (2010) Crowdsourcing document relevance assessment with mechanical turk, *in proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pages 172_179, Los Angeles. Association for Computational Linguistics.
- Hanika, F., Wohlgenannt, G. and Sabou, M. (2014) The uComp Protege Plugin for Crowdsourcing Ontology Validation, *in proceedings of the ISWC 2014 Posters and Demonstrations Track a track within the 13th International Semantic Web Conference*, Italy, October 21, 2014.
- Hodge, G. (2000) Systems of knowledge organization for digital libraries: beyond traditional authority files, Digital Library Federation, Council on Library and Information Resources.
- Hsueh, P.-Y., Melville, P., and Sindhvani, V. (2009) Data quality from crowdsourcing: A study of annotation selection criteria, *in proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing*, pages 27-35, Boulder, Colorado. Association for Computational Linguistics.
- Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006) HT06, tagging paper, taxonomy, Flickr, *in proceedings of the 17th Conference on Hypertext and Hypermedia* (pp. 31-40). New York: ACM.

- Munro, R., Bethard, S., Kuperman, V., Lai, V. T., Melnick, R., Potts, C., Schnoebelen, T., and Tily, H. (2010) Crowdsourcing and language studies: the new generation of linguistic data, *in proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pages 122-130, Los Angeles. Association for Computational Linguistics.
- Lancaster, F. W. (2003) Indexing and abstracting in theory and practice (3rd ed.), Urbana, IL: University of Illinois Urbana.
- Page, L., Brin, S., Motwani, R., and Winograd, T. (1999) The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab. Previous number = SIDL-WP-1999-0120.
- Rumshisky, A. (2011) Crowdsourcing word sense definition, *in proceedings of the 5th Linguistic Annotation Workshop*, pages 74-81, Portland, Oregon, USA. Association for Computational Linguistics.
- Rumshisky, A., Botchan, N., Kushkuley, S., and Pustejovsky, J. (2012), Word sense inventories by non-experts, *in proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey. European Language Resources Association (ELRA).
- Rumshisky, A., Moszkowicz, J. and Verhagen, M. (2009) The holy grail of sense definition: Creating a sense-disambiguated corpus from scratch, *in proceedings of 5th International Conference on Generative Approaches to the Lexicon*, Pisa, Italy.
- Safre, W. (2009) On language. *New York Times Magazine*.
- Sayeed, A., Rusk, B., Petrov, M., Nguyen, H., Meyer, T., and Weinberg, A. (2011) Crowdsourcing syntactic relatedness judgements for opinion mining in the study of information technology adoption, *in proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 69-77, Portland, OR, USA. Association for Computational Linguistics.
- Scharl, A., Sabou, M. and Föls, M. (2012) Climate Quiz: a Web Application for Eliciting and Validating Knowledge from Social Networks, *in proceedings of the 18th Brazilian symposium on Multimedia and the web, WebMedia '12*, pages 189–192. ACM.

- Weichselbraun, A., Scharl, A. and Lang, H.-P. (2013) Knowledge Capture from Multiple Online Sources with the Extensible Web Retrieval Toolkit (eWRT), *in proceedings of the Seventh International Conference on Knowledge Capture (K-CAP 2013)*. Banff, Canada.
- Wohlgemann, G., Belk, St., Schett, M. (2013) Computing Semantic Association: Comparing Spreading Activation and Spectral Association for Ontology Learning, *in Computing Semantic Association: Comparing Spreading Activation and Spectral Association for Ontology Learning*, Hrsg. Ramanna, S., Lingras, P., Sombaththeera, C., Krishna, A. (eds.), MIWAI, Lecture Notes in Computer Science (LNCS) 8271, 317-328. Krabi, Thailand: Springer.
- Zaidan, O. F. and Callison-Burch, C. (2011) Crowdsourcing translation: Professional quality from non-professionals, *in proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 1220-1229, Portland, Oregon, USA. Association for Computational Linguistics.