MuEVo, a breast cancer Consumer Health Vocabulary built out of web forums

Solène Eholié, Mike Donald Tapi Nzali, Sandra Bringay, Clement Jonquet

MuEVo, a breast cancer Consumer Health Vocabulary built out of web forums

Solène Eholié[1], Mike-Donald Tapi-Nzali[1,2], Sandra Bringay[1], and Clement Jonquet[1,3]

[1] Laboratory of Informatics, Robotics and Microelectronics of Montpellier (LIRMM)
University of Montpellier & CNRS, France
[2] Institut Montpelliérain Alexander Grothendieck (IMAG)
University of Montpellier, France
[3] Center for BioMedical Informatics Research (BMIR)
Stanford University, USA
{prenom.nom}@lirmm.fr

Abstract. Semantically analyze patient-generated text from a biomedical perspective is challenging because of the vocabulary gap between patients and health professionals. The medical expertise and vocabulary is well formalized in standards terminologies and ontologies, which enable semantic analysis of expert-generated text; however resources which formalize the vocabulary of health consumers (patients and their family, laypersons in general) remain scarce. The situation is even worse if one is interested in another language than English. In previous studies, we attempted to produce a French preliminary Consumer Health Vocabulary (CHV) by mining the language used within online public forums & Facebook groups about breast cancer. In this work, we show our effort to concretely align the vocabulary produced to standard terminologies and to represent its content (terms & mappings) using semantic web languages such as RDF, SKOS and PROV. We used a sample of 173 relations built around 64 expert concepts which have been automatically (89%) or manually (11%) aligned to standard biomedical terminologies, in our case: MeSH, MedDRA and SNOMEDint. The resulting vocabulary, called MuEVo (Multi-Expertise Vocabulary) and the mappings are publicly available in the SIFR BioPortal French biomedical ontology repository.

Keywords: Semantic Web, Biomedical Ontologies, Standard Terminologies, Consumer Health Vocabulary, SKOS, Web Forums, Breast Cancer, BioPortal.

1 Introduction

Over the last years, the Web has taken a great importance in the way health consumers access health information. In France, according to a 2013 TNS Sofres survey,[4] one in two people has already used the Web to search for medical information or discuss health topics. Health consumers, i.e., patients, are generally laypersons who do not have the technical or scientific expertise and hence expressions and vocabulary [9]. Often, they seek a first diagnosis, more precise information about a given disease or testimonies of other people facing the same health issues. According to Google's blog,[5]

[4] http://www.patientsandweb.com/?p=90
[5] googleblog.blogspot.fr/2016/06/im-feeling-yucky-searching-for-symptoms.html (June 2016)

one percent, so millions, of searches on Google online search engine are symptom-related. However, when laypersons express their queries or discuss in social media, they use a vocabulary different from the one of health-care professionals. With the explosion of Web 2.0 and social medias, doctors have definitively realized the enormous potential of data generated by patients [1].

In fact, expert language (terms, wording and expressions) is quite well captured and formalized thanks to miscellaneous semantic resources such as standard terminologies such as MeSH (Medical Subject Headings) or any other terminology from the Unified Medical Language System (UMLS) or ontologies such as the one in the OBO Foundry or the NCBO BioPortal [12]. However, capturing the language actually used by patients and formalizing it in to a Consumer Health Vocabulary (CHV) remains a research issue. Laypersons use abbreviations, misspellings, neologisms or existing words that are diverted from their standard professional use. Hence, the classic biomedical natural-languages-processing-resources do not apply easily to analyze patient-generated text. Many researchers have been working to reduce this vocabulary gap between laypersons and health care professionals by identifying CHV constituents and/or mapping them to their equivalents in the standard biomedical semantic resources [21,4]. However, this effort does not often result in reusable open access resources. Indeed, one of the only freely available CHV is the (English) Open-Access and Collaborative CHV (itself included in the UMLS Metathesaurus) that was developed by Univ. of Utah and recently updated by mining social network data [3].

Despite of the lack of publicly available CHV, patients express themselves online more and more every day, for instance on social medias such as web forums. According to a 2011 *Health On the Net Foundation* survey [16], the Web has become the second source of information for patients after consultations with a doctor. 24% of the population uses the Web to find health information at least once per day (and up to 6 times per day) and 25% at least several times per week. While maintaining anonymity, social media allow them to freely discuss with other users, and also with health professionals. They discuss about their medical results and their treatment options, but they also receive moral support. In France, online forums such as Doctissimo.fr (general health) or Lesimpatientes.com (breast cancer) are very successful. Therefore, online social medias are very relevant data sources to help to build CHV. Semantically representing CHVs' content and using them inside forum applications will enhance the patient's access to information by connecting the formal medical expertise to the actual content of the forums, inside the forums. It would also enable to process semantically patient-generated text. For instance, topics discussed will be more easily mined in order to identify what are the principal concerns of the patients [13]; forum providers would be able to connect their users to reference data resources that are indexed with standard medical terminologies but that are targeted for patients e.g., MedLinePlus.

In this paper, using patient-generated text from breast cancer health forums, we will present elements of responses and concrete results to this research question. Building on the previous results presented in [17] where we focused on the methodology to extract a preliminary CHV out of forum patient posts, we present is this paper a concrete machine-readable formalization of the extracted vocabulary, the provenance information and the alignment to standard terminologies, using the semantic Web languages:

RDF, SKOS and PROV. We focused on breast cancer and French language, but our model is generalizable to other domain or language. As a result we have produced a CHV, called MuEVo (Multi-Expertise Vocabulary) of 64 concepts and 173 lay-expert relations defined in SKOS that has been automatically (89%) or manually (11%) aligned to standard biomedical terminologies, in our case: MeSH, MedDRA and SNOMEDint. Although the size of the current vocabulary is quite small, this is the result of an automatic process that will be reproduced on other datasets to augment it. This paper mainly focuses on the representation of this CHV which is independent from the size.

The rest of the article is organized as follows: we first present the related work and background, second, we propose a SKOS/RDF formalization of a CHV, third we expose the methods we used to map this resource to standard biomedical terminologies. Finally, we present experimental results of the mapping step before concluding.

2 Background and related work

2.1 Characterization of a Consumer Health Vocabulary

A Consumer Health Vocabulary (CHV) is a set of terms preferred and used by laypersons to describe medical concepts like symptoms or diseases for examples. See [18] for a more conceptual characterization of CHV constituents. The research effort to acquire CHV terms and map them to expert terms has been important this last decade. In fact, CHVs are a key element to reduce the communication gap between laypersons and health care professionals. The existence of the gap has been enlightened namely by [9] while studying details of the queries performed by users of the US National Library of Medicine online services. They observed that visitors used a lot of misspelled forms and abbreviations. In their experiments the authors found 84% of query terms would not match directly to UMLS although a deeper analysis showed that 30% of those mismatches could manually be identified. Other works include: [15] that used email questions. [20] used Wikipedia. [3] that used Patientslikeme.com forum data. Beyond this lexical gap, there is, in part, a problem of comprehension of the medical jargon [15] or use of popular expressions. For instance, a doctor will talk about a *malignant neoplasm* with a colleague, to describe the *cancer* of a patient, whereas the patient will exchange on forums about his/her *crab*. From this observation, the community raised the need to build lexical resources to mediate between the two worlds. Such a resource should include misspellings, incomplete terms, specific synonyms. Attempts to bridge the gap have been motivated by essentially three reasons: (i) to enable search of professional content by laypersons and vice versa [8], (ii) to vulgarize professional content in a consumer-friendly terminology [22], (iii) to analyze semantically the content produced by laypersons [13]. From the related work, two challenges clearly appear:

- One is to *identify* the constituents of CHV (that is, to build a controlled vocabulary). For instance, we can cite [21] that mined MedlinePlus queries logs to extract frequent n-grams not in UMLS with frequency greater than 50. Then they validated 753 terms out of 7967 reviewed.
- The other is to explicitly *map* the constituents to expert vocabulary with the use of lexical or semantic tools. For instance, we can cite [7] that mapped consumer

health concepts extracted from health-focused bulletin boards to UMLS concepts (both manually and using MetaMap).

These are the challenges we attempted to tackle in previous work [17] where we proposed a semi-automatic hybrid approach in which the mapping task serves as validation step in the CHV acquisition process. The final output was a set of pairs of lay-expert terms. In this method, the candidates terms were extracted automatically from text generated by patients in social media. The validation step was semi-automatic. The mapping was performed based on, first, lexical information such as Carry stemming algorithm [14] to detect abbreviations and Levenshtein distance for spelling errors and, second, external resources such as Google index and Wikipedia structure. Few mappings were obtained based on the corpus only: 22 given by a custom Jaccard measure. Out of 1900 candidates, we reported a total number of 122 lay-expert pairs. However, the pairs extracted might not be exact synonyms.

2.2 Standard formalization of Consumer Health Vocabulary content

Although identification and mapping are crucial, to the best of our knowledge, there has been no proposition on how to represent the content (terms & mappings) of a CHV using a standard format to facilitate semantic interoperability and reuse. Even if this has been sometime pointed by the community [24], existing CHVs are either unavailable for public use or stored in a non-standardized format. In addition of OAC-CHV previously mentioned (available in XLS format) or our French preliminary CHV [17] (available in plain text), we can mention the Personal Health Terminology, developed by Intelligent Medical Object, the Mayo Consumer Vocabulary, developed by Mayo Clinic or the MedlinePlus Health Topics dictionary, developed by US National Library of Medicine.

Recently, some ontology designers have taken the initiative to directly populate their ontologies with layperson synonyms, which is probably the best practice (i.e., represent them as any other synonyms). For instance, in [19], the authors present how they have involved layusers in describing more than 6000 synonyms for the Human Phenotype Ontology.[6] In our study, this will not be possible as we are not the developers of the targeted ontologies, however, once available in standard format, the lay terms could be considered as candidate synonyms by the developers (e.g., French INSERM organization which is in charge of the French version of MeSH).

One candidate format for representing CHV as an independent resource is the Simple Knowledge Organization System (SKOS) language. This W3C recommendation is widely used in the semantic Web community. SKOS is a language to develop thesauri, taxonomies or controlled vocabularies [11]. It allows easy knowledge representation in a machine-readable format based on RDF graphs; plus it offers multiple standard properties to represent mappings between concepts *(skos:exactMatch*, *skos:closeMatch*, etc.). In biomedicine, OBO and OWL are also common standard used for knowledge formalization with more details than SKOS/RDF. However, when developing CHVs, researchers are usually only interested by the lexical/terminological level i.e., the labels and their relations only. Therefore, SKOS seems the most suitable standard to use and

[6] This recent results have not been used in our work.

this is the choice made in this paper to formalize both the CHV terms and the mappings to standard terminologies.

3 SKOS formalization model

3.1 Data used

We propose hereafter a model to formalize a CHV into a SKOS vocabulary and then a protocol to align it to standard terminologies. We have experimented with our preliminary CHV [17] and used terminologies and web services offered by the SIFR BioPortal, a repository of French biomedical terminologies (http://bioportal.lirmm.fr) [5].

Our CHV, MuEVo, is built out of 173 lay-expert relations extracted from public French breast cancer forums. These relations have been obtained via a mapping between a patient-generated corpus made of posts from Cancerdusein.org and public Facebook groups[7] and a seed expert vocabulary offered by French National Cancer Institute (INCa - www.e-cancer.fr/dictionnaire) [2]. As provenance information, each relation has a type, a discovering method, and a weight which represents the confidence of the relation (cf. table 1).

Table 1. Examples of lay-expert relations from our preliminary CHV [17]

Lay term	Expert term	Relation type	Method	Weight
nez	pharynx	association	wikipedia	10.0
abaltion	ablation	misspelling	aspell	100.0
onco	oncologue	abbreviation	carry	50.0
traitement hormonal	hormonothérapie	association	wikipedia	100.0

3.2 Representation of CHV terms

The knowledge unit in SKOS is *skos:Concept*; it is a RDF resource which formalizes an idea, a reality. It can have at most one preferred label (*skos:prefLabel*) used to denote the concept. Others terms can be associated to the concepts as valid variants (*skos:altLabel*) or deprecated/hidden variants (*skos:hiddenLabel*). SKOS's model alone is not enough to capture the provenance metadata describing each lay-expert relation: type, method and weight. To capture that, we used the PROV Ontology, a W3C recommandation to formalize provenance information. The complete model is described in figure 1 and exemplified in the listing hereafter.

```
<skos:Concept rdf:about=" http: // purl . lirmm. fr / ontology / MuEVo/vpm52">
  <rdf:type rdf:resource =" http: // www.w3.org/ns/prov#Entity"/>
  <skos:inScheme rdf:resource =" http: // purl . lirmm. fr / ontology / MuEVo"/>
  <skos:prefLabel xml:lang="fr">oncologue</skos:prefLabel>
  <skos:altLabel xml:lang="fr">onco</skos:altLabel>
  <prov:wasDerivedFrom
      rdf:resource=" http: // purl . lirmm. fr / ontology / MuEVo/provEntity86"/>
```

[7] *Cancer du sein*, *Octobre rose 2014*, *Cancer du sein - breast cancer, brustkrebs*

```
  <skos:broadMatch rdf:resource =" http: // chu-rouen.fr/cismef/SNOMED_int.#J-06120"/>
</skos:Concept>
<prov:Entity rdf:about =" http: // purl .lirmm. fr /ontology/MuEVo/provEntity86">
  <!-- carry|onco|oncologue|Abbr|50.0 -->
  < rdfs:label >onco</ rdfs:label >
  < isocat:abbreviation rdf:resource =" http: // purl .lirmm. fr /ontology/MuEVo/vpm52"/>
  < isocat:weight >50.0</ isocat:weight >
  <prov:wasGeneratedBy rdf:resource =" http: // purl .lirmm. fr /ontology/MuEVo/carry"/>
</ prov:Entity >
```

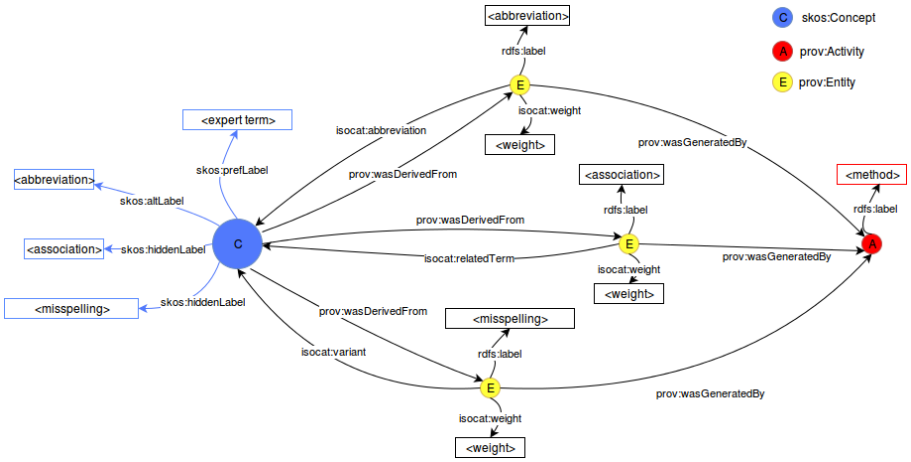

Fig. 1. Model to formalize lay-expert relations using SKOS+PROV

Each *skos:Concept* (in blue) is a formal representation for all the relations found for a given expert term. Relations between the expert term of the concept (represented with *skos:prefLabel*) to a lay term is formalized via the use of SKOS alternate labels (*skos:altLabel* or *skos:hiddenLabel*). The metadata describing the provenance of the relation is represented using a *prov:Entity* (in yellow). This entity is linked to the concept thanks to a ISOcat property[8] selected according to relations in table 2. The weight of the relation is also represented in the *prov:Entity*, with the property *isocat:weight*. Additionally, each *prov:Entity* representing the relation of a lay term relation with the expert term of a concept is linked to the corresponding *skos:Concept* with a *prov:wasDerivedFrom* property. Finally, each mapping method is represented by a *prov:Activity* (in red). Methods are simply described by a label e.g., carry, wikipedia.

3.3 Representation of the CHV mappings

Now, we would like to align *MuEVo* with some standard biomedical terminologies such as the ones that can be found in the NCBO BioPortal [12], a repository of biomedical ontologies and terminologies. As part of the SIFR project, our local appliance of

[8] www.isocat.org - ISOCat identifies properties with ids DC-XX, but for readability with have used here the corresponding name.

Table 2. Selection of SKOS labels and the ISOcat property per type of relation in the CHV.

Relation type	SKOS label	ISOcat property
abbreviation	*skos:altLabel*	*isocat:abbreviation* (DC-331)
misspelling	*skos:hiddenLabel*	*isocat:variant* (DC-330)
association	*skos:hiddenLabel*	*isocat:relatedTerm* (DC-438)

BioPortal gives access to some of the French versions formally mapped to the original English ones [5]. Via the portal, a user can share an ontology and align it to the ones already available in the repository or any other resources thanks to SKOS mappings. We thus uploaded MuEVo in the SIFR BioPortal and then explicitly linked it to the standard biomedical terminologies also available there. Such formal mappings will enable anyone to benefit from the expanded structured knowledge available in standard terminologies when using MuEVo. It can be useful, for example, to semantically index forum content with MuEVo. In our experiments, we used the SIFR BioPortal ontology recommender (working exactly as the NCBO Recommender originally described in [6]) to identify the most appropriate target terminologies and identified MeSH, SNOMEDint and MedDRA which are the set of terminologies that offer the better coverage of the expert terms. The alignment follows two phases:

The direct mapping phase consists in searching each expert term of MuEVo's expert vocabulary via BioPortal REST API search service.[9] The search is restricted to the targeted terminologies. If we find the exact same term (or its plural form) as preferred or alternative label of a concept in one of the targeted terminologies, we set an equivalence mapping, *skos:exactMatch*, between the current MuEVo concept and the one of the targeted terminology. In figure 2, the expert term *vpm:abdomen* matches the preferred label of a concept in a standard terminology, thus, a mapping with the standard concept *std:Abdomen* is created is mapped. The expert term *cancer* is an alternative label for the standard concept *std:Tumeurs* so a *skos:exactMatch* mapping is also created.

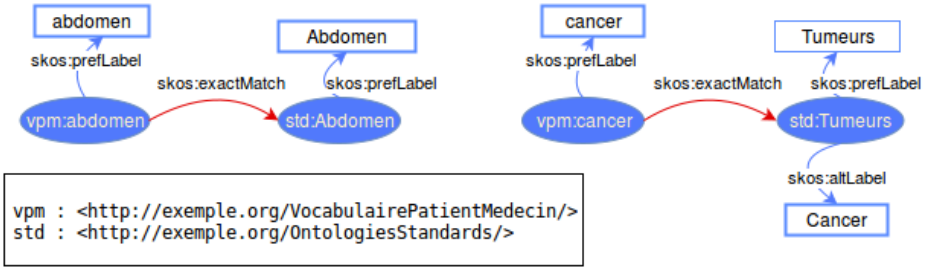

Fig. 2. Examples of direct mappings (namespaces are taken as examples)

The indirect mapping phase is necessary for the terms that are not preferred terms or synonyms of any concept in the targeted terminologies. In this case, we hy-

[9] http://data.bioportal.lirmm.fr/documentation

pothesis that there exists more general resources that can serve as intermediate between INCa list and the targeted standard terminologies. Thus, for a given MuEVo concept C_{MuEVo}, we will make use of an external resource, such as Wiktionary (www.wiktionary.org) [10] in our case, to find semantically related terms noted t_{expert} (*synonyms, hypernyms (broader terms), hyponyms (narrower terms)*) which are themselves used as labels in the targeted terminologies. We automatized the search in Wiktionary using the Java Wiktionary Library API [23] and customized the extraction of semantic relations for French language. We adopted the following workflow:

1. Given a concept C_{MuEVo} of MuEVo, we search the expert term t_{expert} in Wiktionary using C_{MuEVo}'s expert term. If a corresponding page exits, we fetch all terms *t* synonyms, hypernyms and hyponyms;
2. For each term *t*, we try to directly map it as described before;
3. In case of success, we define the following mappings between C_{MuEVo} and the concept C_{target} returned by BioPortal's search API:
   - if *t* is a synonym of t_{expert}: C_{MuEVo} *skos:exactMatch* C_{target};
   - if *t* is a hypernym of t_{expert}: C_{MuEVo} *skos:broadMatch* C_{target};
   - if *t* is a hyponym of t_{expert}: C_{MuEVO} *skos:narrowMatch* C_{target}.

   For example (figure 3), the expert term *cure* (cure) has synonym *traitement* (treatment); *oncologue* (oncologist) has *médecin spécialiste* (specialized physician) as hypernym and a hyponym of *atome* (atom) is *ion* (ion).
4. If no success, we do a manual mapping.

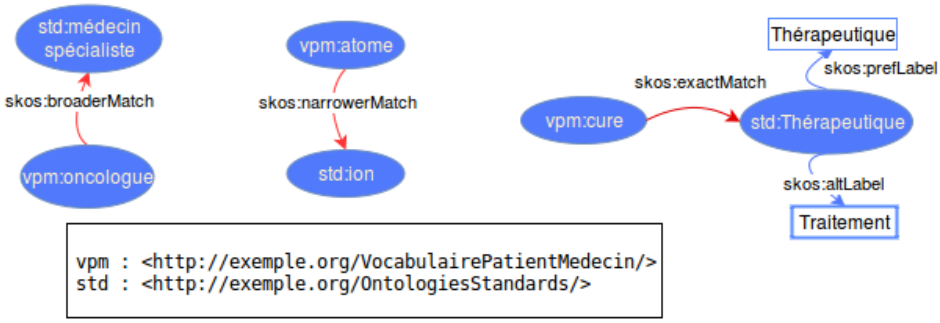

Fig. 3. Examples of indirect mappings (namespaces are taken as examples)

4 MuEVo terms and mappings results

A first version of MuEVo (version 1.3) as described in this paper is available within the SIFR BioPortal: http://bioportal.lirmm.fr/ontologies/MUEVO as well as the mappings created to MeSH, SNOMEDint and MedDRA. It contains 64 *skos:Concept*s which is the result of a fully automatic approach that will be reused in the future on other datasets to enhance the vocabulary. One feature of the portal is to dereference URIs (for *skos:Concept*s in the purl.lirmm.fr namespace) to the corresponding web page in

the web application, which makes the use of the vocabulary more comprehensible. Table 3 sums up the results of the automatic mapping of the 64 processed expert terms. Although the current version is pretty small, the mapping process has been automatized for scalability when the CHV will grow. The three targeted terminologies cover 84,38% of the expert terms: MeSH (70,31%), SNOMEDint (51,56%) and MedDRA (37,5%). 25% are only in MeSH, 7,81% only SNOMEDint and 4,69% only in MedDRA.[10] Among the ten missing terms, three have been successfully mapped thanks to hyponyms from Wiktionary. For the seven others[11], we performed a manual mapping. The validity of each mapping, both manual and automatic has been manually checked. The automatic alignment rates are good because: (i) the union of the three targeted ontologies is quite large, which increases the chance of finding a term using the SIFR BioPortal search web service, especially for the cancer domain that is well cover by MeSH or SNOMEDint; (ii) the seed expert vocabulary provided by INCa already used relevant biomedical terms that could have be found in standard terminologies.

Table 3. Mapping results obtained automatically for 64 input terms at direct mapping phase (1A, 1B) and 10 terms at indirect mapping phase (2).

	Number	**Examples**
1A : Singular	51	abdomen → Abdomen (MeSH)
1B : Plural	17	glucide → Glucides (MeSH)
1A+1B	54	
2 : Hyponyms	3	atome → ion (SNOMEDint)

Some obvious terms related to breast cancer may actually not be included in the vocabulary if they are not used by laypersons as synonyms within the forum data. For instance, the term 'sein' (breast in English) is not in MuEVo mainly because this is the term actually used by patients online, where as the term 'cancer' (cancer in English) is actually in MuEVo because we have identified the relation with the term 'crabe' (crab in English), familiar term that happens to be used by layusers. In addition, it is important to note that MuEVO is not an ontology as it's goal is not to capture medical information that has already been captured in existing standards terminologies and ontologies. MuEVo is a simple vocabulary which goals is to complete existing knowledge with relevant alignments. It essentially aims at building a bridge, in the context of breast cancer forums, between lay expressions and their expert semantic neighbors.

5 Conclusion & Perspectives

In this article, we presented a formalization of a lay-expert vocabulary with semantic Web languages as well as our method to map the expert terms to standard biomedical terminologies. This new resource, once extended in future work, could become a key component for multi-expertise (lay-expert) information retrieval or text mining

[10] 16 are only in MeSH, 5 only SNOMEDint and 3 only in MedDRA

[11] cure, guérison, médecin traitant, oncologue, organe, physicien, rémission

application. For example, health professionals could be interested in identifying recurrent symptoms mentioned in social media. Moreover, it will enable to mediate between these two expertise levels: from expert to lay, e.g., to vulgarize a medical production such as medical records or from lay to expert, e.g., to retrieve expert documents given a lay query. Once the two vocabularies interconnected, it will become possible to use the knowledge formalized in ontologies to semantically search patient data. Or it will be possible to automatically classify forum posts with general medical terms as it was done for instance with scientific literature (e.g., GoPubMed). This feature are very relevant for forum provider such as Doctissimo.com or Lesimpatientes.com. For MuEVo, being hosted in the SIFR BioPortal enables to benefit from all functionalities especially the SIFR/French Annotator that will now be able to use MuEVo to semantically index patient generated text.

The current version of MuEVo is limited to breast cancer but our CHV extraction method and representation model (terms & mappings) are generalizable to other domains or languages. In future work, we consider several perspectives for this work:

- to extract more lay-expert pairs of terms and obtain finer type of association. Especially by using ontology terms directly in the seed vocabulary;
- to process data from other sources (other topic, other domain) and adapt our method to deal with new issues such as disambiguation;
- to use other resources than Wiktionary in the indirect mapping phase, including other ontologies (that can be aligned to one target ontology);
- to use MuEVo to semantically index social media data and evaluate the results in terms of semantic search. For instance, to mesure recall increase when including posts with misspelled words in the query responses;
- to use MuEVo for classification tasks using the hierarchies of the ontologies to which MuEVo is aligned to.

6 Acknowledgements

This work is achieved within the SIFR project funded by the European H2020 Marie Curie actions (grant 701771) and the French National Research Agency (grant ANR-12-JS02-01001) as well as by University of Montpellier and the CNRS. We also acknowledge support of French IReSP (Institut de Recherche en Santé Publique).

References

1. Breton, D., Bringay, S., Marques, F., Poncelet, P., Roche, M.: Epimining: Using Web News for Influenza Surveillance. In: 3rd Workshop on Data Mining for Healthcare Management, DMHM'12. Kuala Lumpur, Malaysia (May 2012)
2. Delavigne, V.: Peut-on "traduire" les mots des experts? un dictionnaire pour les patients atteints de cancer. Dictionnaires et traduction pp. 233–263 (2012)
3. Doing-Harris, K.M., Zeng-Treitler, Q.: Computer-assisted update of a consumer health vocabulary through mining of social network data. Medical Internet research 13(2), e37 (2011)
4. Jiang, L., Yang, C.C., Li, J.: Discovering consumer health expressions from consumer-contributed content. In: Social Computing, Behavioral-Cultural Modeling and Prediction. pp. 164–174. Springer (2013)

5. Jonquet, C., Annane, A., Bouarech, K., Emonet, V., Melzi, S.: SIFR BioPortal : Un portail ouvert et générique d'ontologies et de terminologies biomédicales francaises au service de l'annotation sémantique. In: 16th Journées Francophones d'Informatique Médicale, JFIM'16. Genève, Suisse (July 2016)
6. Jonquet, C., Musen, M.A., Shah, N.H.: Building a Biomedical Ontology Recommender Web Service. Biomedical Semantics 1(S1) (June 2010)
7. Keselman, A., Smith, C.A., Divita, G., Kim, H., Browne, A.C., Leroy, G., Zeng-Treitler, Q.: Consumer health concepts that do not map to the umls: where do they fit? In: American Medical Informatics Association. vol. 15, pp. 496–505. Elsevier (2008)
8. Kogan, S., Zeng, Q., Ash, N., Greenes, R.A.: Problems and challenges in patient information retrieval: a descriptive study. In: AMIA Annual Symposium. p. 329. (2001)
9. McCray, A.T., Loane, R.F., Browne, A.C., Bangalore, A.K.: Terminology issues in user access to web-based medical information. In: AMIA Annual Symposium. p. 107. (1999)
10. Meyer, Gurevych: Wiktionary: A new rival for expert-built lexicons? exploring the possibilities of collaborative lexicography. In: Electronic Lexicography. pp. 259–291 (2012)
11. Miles, A., Matthews, B., Wilson, M., Brickley, D.: Skos core: simple knowledge organisation for the web. In: Int. Conference on Dublin Core and Metadata Applications. p. 3 (2005)
12. Noy, N.F., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N.B., Jonquet, C., Rubin, D.L., Storey, M.A., Chute, C.G., Musen, M.A.: BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Research 37, 170–173 (May 2009)
13. Opitz, T., Azé, J., Bringay, S., Joutard, C., Lavergne, C., Mollevi, C.: Breast cancer and quality of life: medical information extraction from health forums. In: Medical Informatics Europe. pp. 1070–1074 (2014)
14. Paternostre, M., Francq, P., Lamoral, J., Wartel, D., Saerens, M.: Carry, un algorithme de désuffixation pour le français. Rapport technique du projet Galilei (2002)
15. Patrick, T.B., Monga, H.K., Sievert, M.C., Hall, J.H., Longo, D.R.: Evaluation of controlled vocabulary resources for development of a consumer entry vocabulary for diabetes. Medical Internet research 3(3), 24 (2001)
16. Pletneva, N., Vargas, A., Boyer, C.: How Do General Public Search Online Health Information? Results of survey, Health On the Net Foundation, Geneva, Switzerland (July 2011)
17. Tapi Nzali, M.D., Bringay, S., Lavergne, C., Opitz, T., Azé, J., Mollevi, C.: Construction d'un vocabulaire patient/médecin dédié au cancer du sein à partir des médias sociaux. In: Journées Francophones d'Ingénierie des Connaissances. pp. 9–20 Rennes, France. (2015)
18. Tse, T., Soergel, D.: Exploring medical expressions used by consumers and the media: an emerging view of consumer health vocabularies.In: AMIA Annual Symposium. pp. 674–8 (2003)
19. Vasilevsky, N., Engelstad, M., Foster, E., Mungall, C., Robinson, P., Kohler, S., Haendel, M.: Enhancing the human phenotype ontology for use by the layperson. In: 7th Int. Conference on Biomedical Ontologies, ICBO'16, poster Session. No. IT402, Corvallis, Oregon, USA (August 2016)
20. Vydiswaran, V.V., Mei, Q., Hanauer, D.A., Zheng, K.: Mining consumer health vocabulary from community-generated text. In: AMIA Annual Symposium. vol. 2014, p. 1150. (2014)
21. Zeng, Q., Tse, T., Divita, G., Keselman, A., Crowell, J., Browne, A., Goryachev, S., Ngo, L.: Term identification methods for consumer health vocabulary development. Medical Internet Research 9(1), e4 (2007)
22. Zeng, Q.T., Tse, T.: Exploring and developing consumer health vocabularies. In: American Medical Informatics Association. vol. 13, pp. 24–29. Elsevier (2006)
23. Zesch, T., Müller, C., Gurevych, I.: Extracting lexical semantic knowledge from wikipedia and wiktionary. In: LREC. vol. 8, pp. 1646–1652 (2008)
24. Zielstorff, R.D.: Controlled vocabularies for consumer health. Biomedical Informatics 36(4), 326–333 (2003)