MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190172

The MeSH-Gram Neural Network Model: Extending Word Embedding Vectors with MeSH Concepts for Semantic Similarity

Saïd Abdeddaïm[a], Sylvestre Vimard[a], Lina F. Soualmia[a,b]

[a] *Normandie Univ., UNIROUEN, UNIHAVRE, INSA Rouen, LITIS, F-76000, Rouen, France,*
[b] *Sorbonne Univ., Inserm, LIMICS, F-75006 Paris, France*

Abstract

Eliciting semantic similarity between concepts remains a challenging task. Recent approaches founded on embedding vectors have gained in popularity as they have risen to efficiently capture semantic relationships. The underlying idea is that two words that have close meaning gather similar contexts. In this study, we propose a new neural network model, named MeSH-gram, which relies on a straightforward approach that extends the skip-gram neural network model by considering MeSH (Medical Subject Headings) descriptors instead of words. Trained on publicly available PubMed/MEDLINE corpus, MeSH-gram is evaluated on reference standards manually annotated for semantic similarity. MeSH-gram is first compared to skip-gram with vectors of size 300 and at several windows' contexts. A deeper comparison is performed with twenty existing models. All the obtained results with Spearman's rank correlations between human scores and computed similarities show that MeSH-gram (i) outperforms the skip-gram model and (ii) is comparable to the best methods that need more computation and external resources.

Keywords:

Medical subject headings, Neural network models, Unified medical language system.

Introduction

Eliciting semantic similarity and relatedness between concepts is a major issue in the biomedical domain. Different measures have been proposed the last decades [1]. Those measures quantify the degree to which two concepts are similar. They rely either on knowledge-based approaches using ontologies and terminologies, or on corpus-based approaches, which are founded on distributional statistics (e.g. literature-based drug discovery) [2-5]. Several clinical applications of importance rely on semantic similarity and relatedness [6], such as biomedical information extraction and retrieval, clinical decision support, or disease prediction. For instance, biomedical information extraction and retrieval are improved by including semantically related terms and concepts [7-10].

The recent approaches that have given better results in semantic similarity and relatedness measures are founded on word embedding vectors computed by neural networks. Indeed, such architectures implemented initially by word2vec [11], have gained in popularity in the biomedical domain as they risen to efficiently capture semantic similarity and relatedness relationships between words and concepts [12-17]. Word embeddings is based on neural network language modeling where words are mapped to fixed-dimension vectors of real numbers. The similarity between words can thus be measured by the (cosine) similarity between vectors that are constructed over a training corpus. All co-occurrences of a word and its neighbors (*i.e.* contexts) within a predefined window size are considered. The idea behind those representation learning approaches is that two words that have close meaning generally have similar contexts [18]. For example, the words *epilepsy* and *convulsion* will both have *brain* and *mind* as neighbors.

The word2vec developed by Mikolov et al. [11] is a neural network language model that learns word vectors that either maximizes the probability of a word given the surrounding context, referred to as the CBOW (Continuous Bag Of Words) approach, or to maximize the probability of the context given a word, referred to as the skip-gram approach.

In this study we propose a new method, named MeSH-gram, which relies on a straightforward approach: it computes the word vectors by only using the MeSH (Medical Subject Headings) descriptors that are already included in the PubMed/MEDLINE corpus. The MeSH-gram model extends the skip-gram neural network model used in word2vec [11] and fastText tools [19]. The fastText is a successful re-implementation of word2vec which is designed to compute the vector of each word using its neighbors. The extension we propose in the MeSH-gram model replaces the neighbors by the MeSH descriptors of the abstract where each word occurs.

Related Works

Several semantic similarity and relatedness measures have been proposed the last decades [17]. Many of them have been implemented in the UMLS::Similarity package [20] available in the UMLS (Unified Medical Language System). They differ on the method used: path-based, content-based, UMLS-based, corpus-based, and more recently, methods based on word vectors and concepts vectors. Path-based measures [7] use the hierarchical structure of a taxonomy to measure similarity: concepts close to each other are more similar. For instance, Sajadi et al. [21, 22] developed a ranking algorithm based on Wikipedia graph metrics and used it to compare biomedical concepts. Content-based information measures [23, 24] quantify the amount of information a concept provides: the more specific concepts have a greater amount of information content. Other approaches [25, 26] use the entire UMLS Metathesaurus® [27] in order to compare the context in the definition of the concept to quantify its relatedness.

Several methods are vector-based: the concepts are represented by vectors, and the relatedness is usually estimated using the cosine similarity between them. In [26], the authors proposed to compute gloss vectors based on second order co–occurrences trained on WordNet. In [28], the authors computed the cosine of two Latent Semantic Indexing concept vectors based on Pointwise Mutual Information association measure matrix. Recent vector-based methods use neural networks in order to compute concept vectors. The word2vec tool [11] was trained on different corpora: OSHUMED by Sajadi et al. [22]; PubMed/MEDLINE by Chui et al. [13]; PubMed Central by Muneeb et al. [12], Chiu et al. [13], and Pakhomov et al. [14]; and CLINICAL-ALL by [14]. Following the approach used by De Vine et al. [29] on OSHUMED, Yu et al. [15] trained word2vec on PubMed/MEDLINE transformed into UMLS concepts using the MetaMap indexing tool [30].

Other recent methods rely on word vectors. In their previous work, Yu et al. [31] retrofitted word vectors obtained by word2vec with hierarchical information from the MeSH thesaurus. Recently, Henry et al. [16] compared different ways to combine word vectors in order to compute multi-word term vectors. The compared multi-word term aggregation method consists in the summation (averaging) of component word vectors, creating concept vectors using the MetaMap indexing tool [30], and creating multi-word term vectors using the compoundify tool based on the UMLS Specialist Lexicon as glossary [27]. More recently, Henry et al. [17] used association measures for estimating semantic similarity and relatedness between biomedical concepts on PubMed/MEDLINE transformed into UMLS concepts. The best performance results were obtained by [15-17]. Their respective approach relies either on MetaMap in order to transform the text corpus into UMLS concepts, or on additional external resources such as the Specialist Lexicon.

The MeSH-gram model we propose in this study relies on a straightforward approach: it computes the word vectors by only using the MeSH descriptors that are already included in the PubMed/MEDLINE corpus. The extension we propose in the MeSH-gram model replaces the neighbors by the MeSH descriptors of the abstract where each word occurs.

In order to evaluate MeSH-gram, we use publicly available manually annotated corpora: two subsets from Mayo Clinic (MiniMayoSRS) of the MayoSRS (Mayo Semantic Relatedness Set) developed by Pakhomov et al. [32], and two from UMNSRS (The University of Minnesota Semantic Relatedness Set) developed by Pakhomov et al. [33]. MeSH-gram results are first compared to skip-gram and are then compared to twenty existing solutions reported in [17], including the best ones [15-17]. The MeSH-gram model has several advantages: (i) it avoids considering uninformative and too frequent words; (ii) there are less MeSH descriptors than possible context words; and (iii) MeSH descriptors are manually assigned and curated, which assures the best quality of indexing.

Methods

Neural network language models learn word vectors by either maximizing the probability of a word given the context, referred to as the CBOW approach, or by maximizing the probability of the context given a word, referred to as the skip-gram approach.

Skip-gram Word Embedding Model

Given $w_1\ w_2\ ...\ w_n$ a text line of words w_i, the skip-gram model maximizes the following average log probability:

$$\frac{1}{2r}\sum_{i=1}^{2r}\sum_{-r\leq j\leq r, j\neq 0} log\ p\left(w_{t+j}\middle|w_t\right)$$

where w_t is the target word, w_{t+j} is the context, and r is the context window radius. The context words surrounding the target term are determined by the context window radius r.

The probability of a context word w_c given a target word w_t, is computed by:

$$p\left(w_c\middle|w_t\right)=\frac{exp\left(V_{w_c}^T V_{w_t}\right)}{\sum_{w=1}^{N} exp\left(V_w^T V_{w_t}\right)}$$

where N is the vocabulary size, and V_w represents the vector of the word w.

MeSH-gram word embedding model

The MeSH-gram word-embedding model proposed in this paper extends the skip-gram neural network model used in word2vec [11] and fastText [19] tools: it uses MeSH descriptors that are already included in the PubMed/MEDLINE corpus to compute the word vectors.

Given $w_1\ w_2\ ...\ w_n$ the words of a PubMed/MEDLINE abstract, and $m_1\ m_2\ ...\ m_k$ the MeSH descriptors associated to this abstract, the MeSH-gram model maximizes the following average log probability:

$$\frac{1}{k}\sum_{i=1}^{k} log\ p\left(m_i\middle|w_t\right)$$

where w_t is the target word and m_i is a MeSH descriptor.

The probability of a context MeSH descriptor m_c given a target word w_t, is computed by:

$$p\left(m_c\middle|w_t\right)=\frac{exp\left(V_{m_c}^T V_{w_t}\right)}{\sum_{m=1}^{M} exp\left(V_m^T V_{w_t}\right)}$$

where M is the number of MeSH descriptors, V_m represents the vector of the MeSH descriptor m, and V_w the vector of the word w.

We have adapted fastText [19] in order to feed the neural network with pairs of (word, MeSH descriptor). For each abstract included in PubMed/MEDLINE, every word occurrence in the abstract text is associated to each MeSH descriptor, which means that each word vector reflects all the MeSH descriptors seen by its word occurrences in all the PubMed/MEDLINE abstracts.

Vector Representation and Similarity Computation

Using our MeSH-gram model and skip-gram model for comparison, we built word vectors of dimension 300. For the skip-gram model, we computed the vectors considering several window sizes W of 2, 5, 10 and 25.

In order to quantify the relatedness of a pair of words, the cosine distance between the distributional context vectors of each word is used. In the case of a multi-word term, the vector is generated by computing the average of the component word vectors that compose the term. As an example, for the term *epilepsy attack*, the vector $V_{epilepsy\_attack}$ will be computed as $V_{epilepsy\_attack}=(V_{epilepsy}+V_{attack})/2$ where $V_{epilepsy}$ and V_{attack} represent the vector of each word *epilepsy* and *attack* respectively. Rather than combining word vectors after construction, multi-word term vectors may be constructed directly from a preprocessed training corpus in which multi-word terms have been identified [17]. Otherwise, this will involve huge cost in preprocessing and storage requirements.

Training Corpus

We used the PubMed/MEDLINE[1] corpus that contains the abstracts of each article and the associated MeSH descriptors. The corpus was parsed with *pubmed_parser*[2], a python XML parser for PubMed dataset. Each abstract was tokenized using *polyglot*[3].

As fastText needs all the data integrated into one file, we have concatenated all the tokenized PubMed/MEDLINE abstracts. Each line of the resulting file consists of an abstract with its MeSH descriptors. We have adapted fastText in order to feed the neural network with all pairs of (word, MeSH descriptor) of each file line.

Gold Standard

In order to compare the MeSH-gram word embedding model proposed in this study with other methods, we used two evaluation benchmarks: MiniMayoSRS [32] and UMNSRS [33]. MiniMayoSRS consists of 29 clinical term pairs. Two thirty pairs (66.67%) contain a multi-word term. The relatedness of each word pair is rated by medical coders and also by physicians. UMNSRS consists of 566 and 586 pairs of medical terms, for measuring similarity and relatedness respectively. The degree of association between terms in each dataset was rated by four medical residents from the University of Minnesota medical school. As suggested by Pakhomov et al. [33], we use a subset of the ratings consisting of 401 pairs for the similarity set and 430 pairs for the relatedness set. Twenty (4.99%) and seventeen (3.95%) of the term pairs contain multi-word terms for the similarity and relatedness subsets respectively. All these clinical terms correspond to UMLS concepts included in the Metathesaurus®.

The correlations between the generated relatedness scores and the human-assigned scores are calculated using Spearman's rank.

Results

Skip-gram Model versus MeSH-gram Model

The results of the experiments are in Table 1 in which a comparison is performed between the results obtained with skip-gram model and those obtained by the MeSH-gram model using our modified version of fastText according to the four gold standards: MiniMayoSRS rated by physicians (MiniMayoSRS phys.), MiniMayoSRS rated by medical coders (MiniMayoSRS cod.), UMNSRS for similarity (UMNSRS Sim.) and UMNSRS for relatedness (UMNSRS Rel.). The number of calculated term pairs (*n*) is lower than the number of pairs in the UMNSRS gold standards (*Sim.* and *Rel.*) because the embedding vectors could not be computed for low frequency terms and contexts on PubMed/MEDLINE using fastText. Using descriptors rather than words as context allows to computed more term pairs.

MeSH-gram Model compared to Previous Works

Table 2 gathers the results obtained by the MeSH-gram model we developed and twenty previous works' results. It allows a comparison between all the models and on the same gold standards (MiniMayoSRS and UMNSRS). Table 2 complete the Table 12 given by Henry et al. [17].

Table 1 – Spearman's rank correlations between human scores and computed similarities

	MiniMayo		**UMNSRS**	
	Phys. n=29	**Cod. n=29**	**Sim. n=380**	**Rel. n=397**
Skip-gram				
W=2	0.740	0.757	0.679	0.529
W=5	0.763	0.779	0.704	0.576
W=10	0.776	0.789	0.716	0.589
W=25	0.766	0.781	0.718	0.608
MeSH-gram	**0.811**	**0.855**	**0.724*** *n=387	**0.643**** **n=407

Note: W: window size; n: number of pairs

Discussion

As one can see in Table 1 for the skip-gram model, the more the window is extended, the more the results are improved on the UMNSRS gold standard. The best results of skip-gram are obtained with a window size *W*=10 for the MiniMayoSRS set. This suggests that word vectors are a better solution when we consider an important number of context words in the abstract. The best results are obtained with the MeSH-gram model that considers MeSH descriptors as context for each term, suggesting that MeSH descriptors catch the semantics of all the abstracts associated with it. We can conclude that taking MeSH descriptors instead of context words gives better results than considering a large window size: (i) bigger window size does not lead necessary to better results and (ii) MeSH descriptors are fewer than context words (50 context words for window size *W*=25) leading also a reduced computation time.

For example, the pair *synthroid* and *hypothyroidism* is misclassified by the skip-gram model but better treated by the MeSH-gram model. They are considered as very related in UMNSRS as *Synthroid®* (*levothyroxime sodium*) is used to treat *hypothyroidism*. This term pair has a human-assigned score of 1473, which corresponds to the seventh most related mono-term pair in UMNRRS. While the skip-gram model (with *W*=2) ranks this pair at the 160th position with a similarity score of 0.40, the MeSH-gram model puts it at the 64th position with a similarity score of 0.58, which means that the MeSH-gram model better captures the relatedness for this example. The apparent reason for this could be that *synthroid* occurs only 107 times in PubMed/MEDLINE corpus, which is insufficient to construct reliable vectors using context words in the skip-gram model, while the MeSH-gram model computes its vectors using the more informative MeSH descriptors. The absolute difference between the skip-gram rank (160) and the MeSH-gram rank (64) is 96 for this term pair, however it is lower in average (36.5 with a standard deviation of 33.5) when we consider the scores obtained by the two methods for all the UMNSRS pairs.

On the contrary, the pair of the terms *weakness* and *emaciation* (low frequency term equals to 1433) are better scored by the skip-gram model (84th position) than the MeSH-gram model (165th position), while they are considered closely related by human scores (64th position).

From our observations, the better results obtained by the MeSH-gram model are not due to significant improvements for any specific category of term pairs (e.g. less frequent terms), but to the overall improvement (moderate increase or decrease)

[1] ftp://ftp.ncbi.nlm.nih.gov/pubmed/ [accessed Apr 1st, 2019]
[2] https://github.com/titipata/pubmed_parser [accessed Apr 1st, 2019]
[3] https://github.com/aboSamoor/polyglot [accessed Apr 1st, 2019]

of the similarity scores for each pair. We can only conclude from this that MeSH descriptors are globally more informative than context words.

The comparison with twenty methods displayed in Table 2 confirms that the MeSH-gram model gives comparable results with best previous work methods on the four gold standard datasets. While the methods (1) and (2) rely on the translation of PubMed/MEDLINE text data into ULMS concepts, and methods (4) and (5) require additional steps or resources such as compoundify tool (4) and MetaMapped MEDLINE corpus (5), the MeSH-gram model uses only the raw text corpus as input. The best previous works' results are obtained by the method (2) and then the method (3). However, the method (2) is not recommended by the authors themselves as it uses concept expansion, which requires additional computation cost without significantly increasing the performances for any dataset [17]. MeSH-gram is comparable to method (3) with better results on three datasets. All those results allow us to conclude that UMLS information used by the methods (1) to (5) is already contained in the MeSH descriptors available in the PubMed/MEDLINE corpus and used by the MeSH-gram model. Using MeSH descriptors as context is a good solution for datasets founded on UMLS concepts. However, MeSH-gram should be evaluated on other types of similarities such as BioSimVerb and BioSimLex [34].

Table 2 – Spearman's rank correlations between human scores and computed similarities using MeSH-gram and previous works' methods.

	MiniMayo		UMNSRS	
	Phys.	**Cod.**	**Sim.**	**Rel.**
MeSH-gram	0.81 (n=29)	**0.86** (n=29)	***0.72*** (n=387)	***0.64*** (n=407)
(1) Henry et al. [17]; recommended	**0.84** (n=29)	0.81 (n=29)	0.69 (n=392)	***0.64*** (n=418)
(2) Henry et al. [17]; not recommended	*0.85 (n=29)*	*0.84 (n=29)*	*0.73 (n=392)*	*0.66 (n=418)*
(3) Henry et al. [16]; CBOW words	***0.82*** (n=29)	0.82 (n=29)	0.69 (n=374)	0.61 (n=396)
(4) Henry et al. [16]; CBOW compounds	0.80 (n=29)	0.78 (n=28)	0.70 (n=373)	**0.65** (n=393)
(5) Henry et al. [16]; CBOW concepts	0.77 (n=29)	***0.83*** (n=29)	**0.73** (n=388)	0.60 (n=413)
(6) Yu et al. [15]; narrow +other relations	--	--	0.69 (n=526) 0.68 (n=418)	0.62 (n=543) 0.63 (n=427)
(7) Yu et al. [15]; no lexicons	--	--	0.64 (n=526) 0.63 (n=418)	0.59 (n=543) 0.59 (n=427)
(8) Yu et al. [31]	0.70 (n=25)	0.67 (n=25)	--	--
(9) Sajadi et al. [22]; HITS similarity	0.67 (n=29)	0.72 (n=29)	0.58 (n=566)	0.51 (n=587)
(10) Sajadi et al. [22]; (word2vec OSHUMED+UMLS)	--	--	0.39 (n=566)	0.39 (n=587)
(11) Sajadi et al. [22] (word2vec on OSHUMED)	--	--	0.26 (n=566)	0.29 (n=587)
(12) Chui et al. [13]	--	--	0.65 (n=n/a)	0.60 (n=n/a)
(13) Pakhomov et al. [14]	--	--	0.62 (n=449)	0.58 (n=458)
(14) Muneeb et al. [15]	--	--	0.52 (n=462)	0.45 (n=465)
(15) Workman et al. [22]	0.67 (n=29) 0.69 (n=25)	--	--	--
(16) Patawardhan and Pedersen [26]	0.59 (n=29)	0.58 (n=29)	0.58 (n=387)	0.45 (n=412)
(17) Lin [24]	0.42 (n=26)	0.53 (n=26)	0.49 (n=340)	0.29 (n=360)
(18) Resnik [23]	0.34 (n=26)	0.46 (n=26)	0.49 (n=340)	0.26 (n=360)
(19) Rada et al. [7]	0.35 (n=26)	0.44 (n=26)	0.53 (n=340)	0.29 (n=360)
(20) Lesk [25]	0.52 (n=29)	0.57 (n=29)	0.50 (n=387)	0.33 (n=412)

Note: n: number of pairs (inspired by [17]). The best result is highlighted in bold, the second best in italic and bold.

Conclusions

In this paper, we proposed a new method, MeSH-gram, to create distributional word vectors using MeSH descriptors as word context. We evaluated our results on four standard evaluation datasets, MiniMayoSRS Physicians, MiniMayoSRS Coders, UMNSRS tagged for relatedness, and UMNSRS tagged for similarity, and compared it against skip-gram model as a baseline and previous methods. All the obtained results of Spearman's rank correlations between human scores and computed similarities show that MeSH-gram (i) outperforms the skip-gram model and (ii) is comparable to the best recent methods, methods that need more computation and additional external resources.

We are trying different ways to combine the skip-gram model and the MeSH-gram model in order to improve the results. We also plan in our future works to include in MeSH-gram the MeSH qualifiers affiliated to the descriptors in order to have a more precise semantic meaning (e.g. the association *cancer/complications*, where *cancer* is a MeSH descriptor and *complications* is a MeSH qualifier, is more precise than *cancer* alone). A second step is to use fastText subwords and the evaluation of MeSH-gram for other kinds of similarities such as BioSimVerb and BioSimLex. MeSH-gram may also be used in other languages than English, for instance in French bibliographic corpora such as CISMeF [35], as well as in annotated electronic health records.

References

[1] C. Lofi, Measuring semantic similarity and relatedness with distributional and knowledge-based approaches, *Information and Media Technologies* **10** (3)(2015), 493-501.

[2] M. Yetisgen-Yildiz and W. Pratt, Using statistical and knowledge-based approaches for literature-based discovery. *J Biomed Inform* **39**(6)(2006), 600–611.

[3] P. Agarwal and D.B Searls, Can literature analysis identify innovation drivers in drug discovery? *Nat Rev Drug Discov* **8** (11)(2009), 865–878.
[4] F. Doshi-Velez et al. Graph-sparse LDA: a Topic Model with Structured Sparsity, In *Proc. 29th AAAI Conference on Artificial Intelligence* (2015), 2575–2581.
[5] M.R. Nelson, H. Tipney, J.L. Painter, J. Shen, P. Nicoletti, Y. Shen, et al., The support of human genetic evidence for approved drug indications, *Nat Genet* **47** (8)(2015), 856–860.
[6] B.T. McInnes and T. Pedersen, Evaluating semantic similarity and relatedness over the semantic grouping of clinical term pairs, *J Biomed Inform* **54** (2015), 329–336.
[7] R. Rada, H. Mili, E. Bicknell, and M. Blettner, Development and application of a metric on semantic nets, *IEEE SMC* **19** (1)(1989), 17–30.
[8] T. Cohen and D. Widdows, Empirical distributional semantics: methods and biomedical applications, *J Biomed Inform* **42** (2)(2009), 390–405.
[9] A. Henriksson, H. Moen, M. Skeppstedt, V. Daudaravičius, and M. Duneld, Synonym extraction and abbreviation expansion with ensembles of semantic spaces, *J Biomed Semantics* **5** (1)(2014), 6.
[10] C. Kurtz, C.F. Beaulieu, S. Napel, and D.L. Rubin, A hierarchical knowledge-based approach for retrieving similar medical images described with semantic annotations, *J Biomed Inform* **49** (2014), 227–244.
[11] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, Distributed Representations of Words and Phrases and their Compositionality, In *Advances in Neural Information Processing Systems* (2013), 3111–3119.
[12] T.H. Muneeb, S.K. Sahu, and A. Anand, Evaluating distributed word representations for capturing semantics of biomedical concepts, In *ACL-International Joint Conference on Natural Language Processing* (2015), 158.
[13] B. Chiu, G. Crichton, A. Korhonen, and S. Pyysalo, How to train good word embeddings for biomedical NLP, In *Proc. of the 15th Workshop on BioNLP* (2016), 166–174.
[14] S.V. Pakhomov, G. Finley, R. McEwan, Y. Wang, and G.B. Melton, Corpus domain effects on distributional semantic modeling of medical terms, *Bioinformatics* **32** (23) (2016), 3635–3644.
[15] Z. Yu, B.C. Wallace, T. Johnson, and T. Cohen, Retrofitting concept vector representations of medical concepts to improve estimates of semantic similarity and relatedness, *Stud Health Technol Inform* **245** (2017), 657–661.
[16] S. Henry, C. Cuffy, and B.T. McInnes, Vector representations of multi-word terms for semantic relatedness, *J Biomed Inform* **77** (2018), 111–119.
[17] S. Henry, A. McQuilkin, and B.T. McInnes, Association measures for estimating semantic similarity and relatedness between biomedical concepts, *Artif Intell Med* **93** (2019), 1-10.
[18] Z.S. Harris, Distributional structure, *Word* **10** (2–3) (1954), 146–162.
[19] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, Enriching word vectors with subword information *TACL* **5** (2017), 135-146.
[20] B.T. McInnes, T. Pedersen, and S.V. Pakhomov, UMLS-Interface and UMLS-Similarity: open source software for measuring paths and semantic similarity, In *AMIA Annual Symposium Proc* (2009), 431.
[21] A. Sajadi, Graph-based domain-specific semantic relatedness from Wikipedia, In *Advances in Artificial Intelligence* (2014), 381–386.
[22] A. Sajadi, E. Milios, V. Kešelj, and J.C. Janssen, Domain-specific semantic relatedness from Wikipedia structure: a case study in biomedical text, In *CLILing* (2015), 347–360.
[23] P. Resnik, Using information content to evaluate semantic similarity in a taxonomy, In *14th IJCAI Proc-Vol 1* (1995), 448–453.
[24] D. Lin, Using syntactic dependency as local context to resolve word sense ambiguity, In *8th Conference on European Chapter of the ACL Proc* (1997), 64–71.
[25] M. Lesk, Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone, In *5th Annual ICIS Documentation Proc* (1986), 24–26.
[26] S. Patwardhan and T. Pedersen, Using WordNet-based context vectors to estimate the semantic relatedness of concepts, In *Workshop on Making Sense of Sense: Bringing Psycholinguistics and Comput Linguistics Together Proc* (2006).
[27] O. Bodenreider, The unified medical language system (UMLS): integrating biomedical terminology, *Nucleic acids research* **32** (suppl_1)(2004), D267-D270.
[28] T.E. Workman, G. Rosemblat, M. Fiszman, and T.C. Rindflesch, A literature-based assessment of concept pairs as a measure of semantic relatedness, In *AMIA Annual Symposium Proceedings* (2013), 1512.
[29] L. De Vine, G. Zuccon, B. Koopman, L. Sitbon, and P. Bruza, Medical semantic similarity with a neural language model, In *Proc. of the 23rd ACM CIKM* (2014), 1819-1822.
[30] A.R. Aronson and F.M. Lang, An overview of MetaMap: historical perspective and recent advances, *JAMIA* **17** (3)(2010), 229-236.
[31] Z. Yu, T. Cohen, B. Wallace, E. Bernstam, and T. Johnson, Retrofitting word vectors of MeSH terms to improve semantic similarity measures. In *7th LOUHI Proc* (2016), 43-51.
[32] S.V Pakhomov, T. Pedersen, B. McInnes, G.B. Melton, A. Ruggieri, and C.G. Chute, Towards a framework for developing semantic relatedness reference standards, *Journal of Biomedical Informatics* **44** (2) (2011), 251–265.
[33] S. Pakhomov, B. McInnes, T. Adam, Y. Liu, T. Pedersen, and G. Melton, Semantic similarity and relatedness between clinical terms: An experimental study, *In AMIA Annual Symposium Proc* (2010), 572–576.
[34] B. Chiu, S. Pyysalo, I. Vulić, and A. Korhonen, Bio-SimVerb and Bio-SimLex: wide-coverage evaluation sets of word similarity in biomedicine, *BMC bioinformatics* **19** (1)(2018), 33.
[35] L.F. Soualmia and S.J. Darmoni, Combining different standards and different approaches for health information retrieval in a quality-controlled gateway, *Int J Med Inform* **74** (2-4) (2005), 141-150.

Address for correspondence

Lina F. Soualmia, lina.soualmia@litislab.eu. Normandie Université, CURIB (LITIS), 25 Rue Lucien Tesnière, 76130 Mont-Saint-Aignan, FRANCE. +33 232 955 173