Towards Meningitis Ontology for the Annotation of Text Corpora

W. R. Cédric Béré[1,4(✉)], Gaoussou Camara[2], Sadouanouan Malo[3], Moussa Lo[4], and Stanislas Ouaro[1]

[1] LAMI, Université Ouaga 1 Pr Joseph Ki-Zerbo, 03 BP 7021, Ouaga, Burkina Faso
cedric.bere@univ-ouaga.bf, ouaro@yahoo.fr
[2] EIR-IMTICE, Université Alioune Diop de Bambey, BP 30, Bambey, Senegal
gaoussou.camara@uadb.edu.sn
[3] Université Polytechnique de Bobo-Dioulasso, 01 BP 1091, Bobo, Burkina Faso
sadouanouan@yahoo.fr
[4] LANI, Université Gaston Berger, BP 234, Saint-Louis, Senegal
cedric.bere@gmail.com, moussa.lo@ugb.edu.sn

Abstract. Ontologies are formal models that describe the semantics of domain knowledge in order to facilitate automatic processing and provide reasoning possibilities. In the context of this work, we aim at designing an ontology for assisting data and event extraction from text corpora in social medias. These items extracted will be used as input for meningitis spreading simulation within a surveillance system. In this paper, we describe the methodology we propose for building such an ontology. Our approach is based on the reuse of some NeOn scenarios and the use of an annotation tool for text corpus of the biomedical domain. This tool takes as input a text corpus on meningitis and allows us to identify relevant ontological resources. The paper is ending by presenting examples of conceptual models resulting from the application of the first step of our methodology.

Keywords: Ontology · Meningitis · Annotation · Text corpus · Neon

1 Introduction

Meningitis is an infectious disease with disastrous consequences in many countries, particularly in African countries. It has viral or bacterial origin [14]. Detection of the first cases is crucial for taking appropriate measures to prevent a possible epidemic, hence the need of its surveillance [8]. This requires rapid collection and integration of data and events related to the risk factors of its spread. However, developing countries, particularly those in West Africa, are faced with a lack of a real-time data collection system. This situation delays considerably risk analysis and decision-making. It is in this context that we have initiated a project to set up a simulation platform dedicated to the control and prevention of the spread of meningitis epidemics in Burkina Faso. This platform will

C. M. F. Kebe et al. (Eds.): InterSol 2017/CNRIA 2017, LNICST 204, pp. 322–331, 2018.
https://doi.org/10.1007/978-3-319-72965-7_30

be part of the surveillance and awareness system on meningitis of the Nouna[1] Health Research Center in Burkina Faso. In this platform, the early detection of an epidemic risk is made by a qualitative simulation [12]. The models to implement take as input, qualitative data and events contained in text corpora from social medias. The use of an ontology makes possible to identify the relevant corpora and to extract the data and events related to the domain [2]. In their work, Camara et al. [4] have shown that the use of ontologies are suitable for improving epidemiological surveillance capabilities. For instance, domain ontologies provide a common vocabulary for the semantic annotation of resources in the surveillance domain.

In this paper, our purpose is to define a methodology for building an ontology of meningitis domain that will used in extracting data and events from text messages published in social medias. In addition to this, the ontology will allow data integration, domain knowledge sharing and efficient communication between actors within the epidemiological surveillance system.

In the remainder of this paper, we first present the related works in Sect. 2. Then, in Sect. 3, we present the Neon methodology which is the basis of the methodology we have proposed. In Sect. 4 we describe the methodology we have designed for the building a meningitis ontology. Our Sect. 5 presents examples of conceptual models of meningitis knowledge resulting from the application of the first step of our methodology. Finally, we conclude and draw the perspectives of our work.

2 What Are Ontologies and Ontology Engineering?

On one hand, the notion of ontology meets many definitions that evolve over time [3,9,20]. According to the purpose of our work, the definition from Studer seems more complete and more suitable: *An ontology is a formal, explicit specification of a shared conceptualization.*

On the other hand, ontology engineering is the discipline that studies principles, methods and tools for the building and management of the evolution of ontologies during their life cycles [7]. Several methods and good practices for the development of ontologies have been proposed in the literature [13,16,19]. One method could be considered as better than the other according to the ontology purpose, the actors involved in the process, etc. These methodologies could also be compared according to the type of the engineering process: from scratch, by integration or fusion with other ontologies, collaborative building, evolution of existing ontologies, etc.

In the context of our work, the choice of the ontology engineering method takes into account the fact that the semantics of the vocabulary used is closely related to a domain (contextualized vocabulary). We take also into account the fact that we build a modular application ontology [1]. To meet these criteria, Neon appears as the most exhaustive method which is flexible enough to allow us mixing its scenarios and enrich the methodology by integrating new steps.

[1] http://www.crsn-nouna.bf/.

3 The NeOn Methodology Overview

The methodological approach we propose in this paper is inspired from Neon project[2] [21]. This choice is motivated by the fact that we aim at building a modular ontology that requires knowledge from the domain experts involved in the project. In addition to this, we intend to reuse existing ontological resources. Neon is well adapted to our context because it is a flexible methodology compared to others. The Neon methodology proposes 9 scenarios that one can choose and combine to develop an ontology. The Neon methodology also provides a glossary of activities[3], two life-cycle models and a set of guidelines or methodological recommendations. The 9 scenarios are:

- Scenario 1: construction of ontologies from scratch (from specification to implementation);
- Scenario 2: reuse and re-engineering of non-ontological resources;
- Scenario 3: reuse of ontological resources;
- Scenario 4: reuse and re-engineering of ontological resources;
- Scenario 5: reuse and fusion of ontological resources;
- Scenario 6: reuse, fusion and re-engineering of ontological resources;
- Scenario 7: reuse of ontology design patterns and patterns;
- Scenario 8: restructuring of ontological resources;
- Scenario 9: localization of ontological resources.

In the remainder of the paragraph we detail some scenarios of the Neon methodology that we use in the methodology for building the meningitis ontology.

Scenario 1 of Neon: The activities in this scenario consist in specifying the purpose, scope, intent, and applications or uses that will be made of the ontology. It is also necessary to determine in this scenario the technical constraints in terms of reusability and modularity. All these specifications will be recorded in a specification document which will be used for the validation of the ontology by the domain experts. A first conceptual model of knowledge should be provided at the end of the activities of this scenario.

Scenario 3 of Neon: The activities of this scenario lead to the total or partial reuse of existing ontological resources. This is due to the abundance of ontologies and ontology repositories that lead us to explore these existing ontologies for reuse. There are three main activities: research, selection and evaluation of candidate ontologies.

Scenario 5 of Neon: The activities of this scenario consist in the selection of ontological resources of the same domain (if they exist) to be reused in order to create a new ontological resource. This scenario is therefore conditioned by the existence of these ontological resources of the actual domain.

[2] http://www.neon-project.org/.
[3] http://mayor2.dia.fi.upm.es/oeg-upm/files/pdf/NeOnGlossary.pdf.

Scenario 7 of Neon: The activities of this scenario involve reusing design patterns and best practices in ontology design. This approach ensures better ontology design and implementation.

Scenarios can be combined and are flexible. However, Scenario 1 is a mandatory scenario regardless of the combination to be made.

4 Meningitis Domain Ontology Building Methodology

4.1 Main Steps of the Methodology

We propose here a methodology to build the meningitis domain ontology. Our methodology is based on the NeOn approach. The ontology we propose to build is an extension of the Infectious Disease Ontology (IDO)[4] [6].

The choice of the Neon methodology is firstly motivated by the flexibility of the use of its scenarios. And secondly, because NeOn is a methodology that has been adopted successfully in many projects including a previous work of our research team for the building of the schistosomiasis domain ontology [5].

As mentioned in the previous paragraph, we will not apply all the nine scenarios presented in the NeOn methodology. We consider the scenarios 1, 3, 5 and 7 as illustrated in Fig. 2. The approach is thus customized in five (5) main steps as follows:

1. **Step 1:** this step consists of an implementation of Neon Scenario 1. This Scenario is a mandatory.
2. **Step 2:** annotation of text corpora. This step allows us identifying relevant ontologies from biomedical repositories. This annotation phase will be detailed in the next section.
3. **Step 3:** this step is the implementation of Neon Scenario 3. It consists of the reuse of the relevant ontologies suggested by the annotator at step 2.
4. **Step 4:** this step involves the implementation of Neon Scenario 5. At this stage, we use ontology alignment and/fusion techniques for the enrichment of the ontology from step 3.
5. **Step 5:** this step is the implementation of Scenario 7 of Neon. It consists of the use of ontological design pattern.

The step 2 of our methodology involves two main activities: the constitution of a text corpus and the semantic annotation. The constitution of the corpus is essential and will consist in the construction of a relevant corpus. This corpus must cover as large as possible the biological, clinical and epidemiological perspectives of the meningitis disease.

For the construction of the corpus we used two essentially sources. The first source [10] is realized by a group of multidisciplinary experts under the initiative of the INSERM. This group of experts analyzed the available global literature

[4] http://infectiousdiseaseontology.org/.

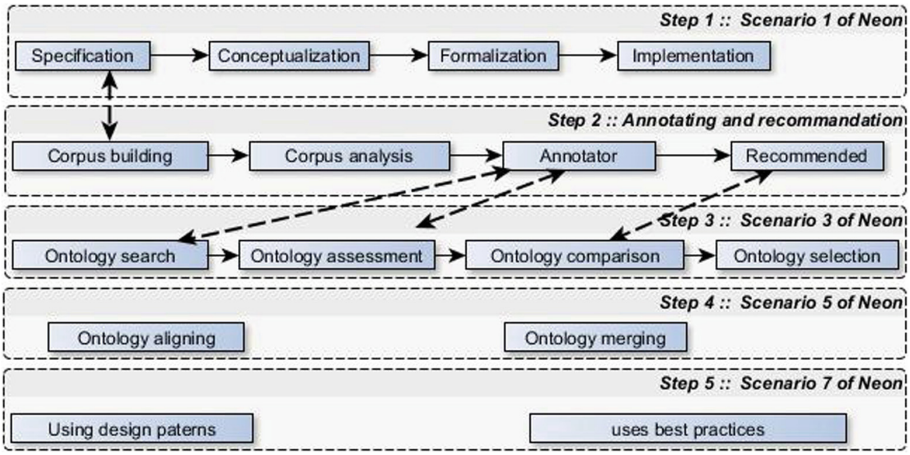

Fig. 1. Methodology of building of the meningitis ontology

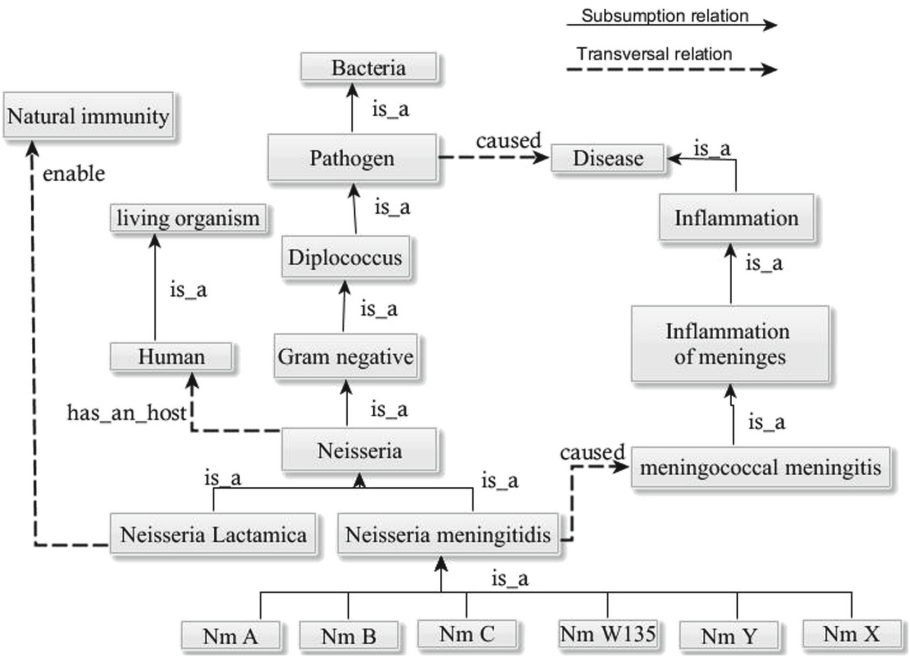

Fig. 2. Modeling the biology of meningococcus.

on meningitis (about 500 articles) and took into account almost all aspects of the disease. This corpus has several advantage:

- the corpus presents in a single document a review of the scientific knowledge on meningitis: the biological aspects of the bacterium, the treatment and prevention means and the epidemiological strategies of the disease;
- recommendation in terms of research direction on meningitis are recorded in the same document. Some of which are currently applied.

The second source is a set of documentary resources on meningitis listed in the INSERM iPubli-Inserm documentary database.

The validation of the definitive corpus is made by experts in meningitis domain. The text corpus validated by all the domain experts is finally used for the semantic annotation step. Ontology extraction will be performed on the multitude of biomedical ontologies available in Bioportal repository.

4.2 Using the Bioportal NCBO Annotator

The approach we have adopted for the building of a meningitis ontology includes the reuse of existing ontological resources. We used the NCBO annotator to exploit the large number of ontological resources existing in the biomedical domain. The annotator operates on the deposit of bioportal. This annotator allows us to retrieve relevant ontologies to a specific domain from a text corpus or from a set of key words or terms derived from a text corpus [11].

An ontology is considered as relevant according to 3 criteria:

- ***coverage:*** the ontology must cover the dataset as broad as possible;
- ***connectivity:*** the ontology must contain sufficient terms/concepts linked to other existing ontologies of the same domain;
- ***size:*** the number of concepts must be great enough.

According to the principle of OBO Foundry [18] (Open Biological and Biomedical Ontology Foundry), the reuse of the ontologies extracted from the bioportal repository will guarantee the reusability of our ontology in the biomedical domain.

5 Applying Step 1 of Our Methodology

In this section, we will implement the following activities defined in the scenario to provide the conceptual models of the different modules of the meningitis ontology: specification, acquisition and modeling.

- ***Specification:*** the building of the meningitis ontology is guided by the purpose of using it for identifying relevant messages from social medias and extract data and events related to risk factors for the spread of the disease.

- ***Acquisition:*** this task is done by interviews with meningitis domain experts and by using non ontological resources such as text corpora as said in previous sections. Text corpora helps for knowledge acquisition and ontological resources discovery from biomedical ontology repository.
- ***Modeling:*** the modeling phase provides as results conceptual models of knowledge that will be formalized by the ontology. These conceptual models support validation of knowledge of the ontology by domain experts.

The two conceptual models presented in Subsects. 5.1 and 5.2 represent modules extracted from the global meningitis ontology model.

5.1 Modeling the Biology of the Pathogen

We present here a scenario which allows us to model the biology of the pathogen of meningitis and more particularly that of meningococcus. Meningococcal meningitis is an inflammation of the meninges. The organism responsible for meningococcal meningitis is Neisseria meningitidis or meningococcus which is an encapsulated Gram-negative diplococcus bacterium belonging to the genus Neisseria. This bacterium has for its sole host a human being. Meningococci are classified according to serological typing. There are 13 distinct meningococcal groups including 6 serogroups (A, B, C, W-135, X, Y) witch are the major fatal causes of meningitis. They are represented Respectively in our model by Nm A, Nm B, Nm C, Nm W-135, Nm X, Nm Y.

5.2 Modeling Meningococcal Transmission, Acquisition and Infection

Transmission of the germ is achieved by close contact with infected nasopharyngeal secretions or by inhalation of droplets of cough (cough, sneeze) emitted by a carrier. After the germ has been acquired by a healthy individual, the bacterium is transported (germ transport process) from the respiratory tract to the nasopharynx through the outer membranes of the bacterium. The membrane proteins PorA and PorB are used to facilitate the movement of the bacteria to the nasopharyngeal mucosa. Then, the bacteria adhere to the mucosa. Colonization starts on the mucous membranes using a multifactorial process involving pili, twitching motility, lipooligosaccharide (LOS) and other more complex mechanisms (horizontal genetic exchange, molecular mimicry ...) [17]. Inflammation of the meninges occurs as soon as the bacterium crosses the blood-brain barrier (BBB) to invade the Cerebrospinal Fluid (CSF). Mechanisms for the penetration of BHM are complex and are still poorly known [15].

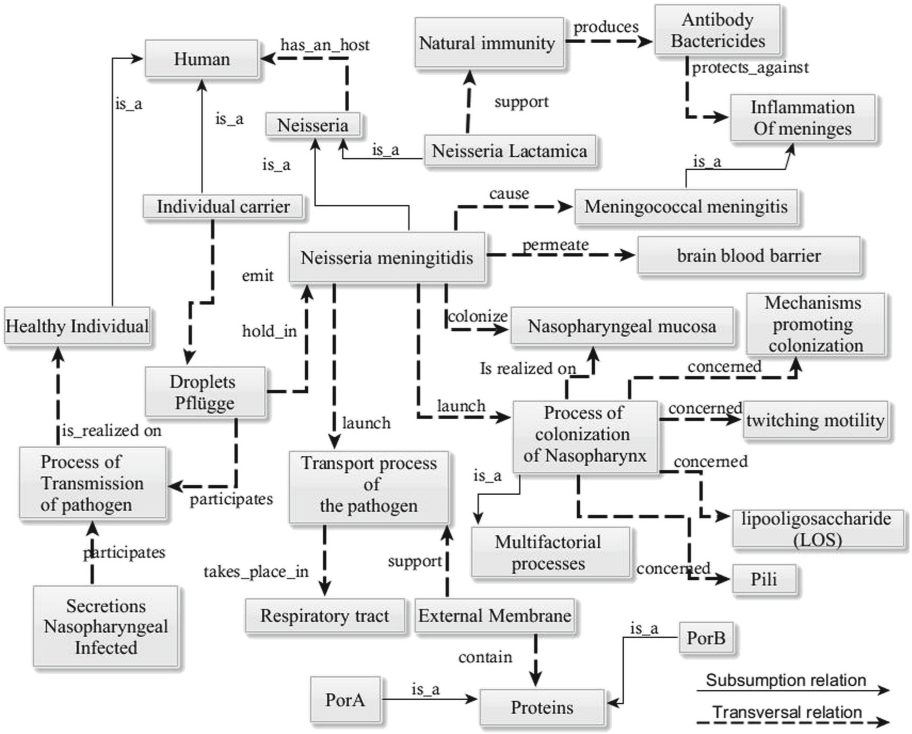

Fig. 3. Modeling meningococcal transmission, acquisition and infection

6 Conclusion and Perspectives

In this paper, we presented the methodology we have proposed for building a meningitis domain ontology. Our proposal is based on the Neon methodology. We also used a text corpus annotation tool for suggesting ontological resources relevant to meningitis domain and eligible for possible reuse in the meningitis ontology engineering.

In the following of this work, the existing ontological resources identified by the annotation tool will be reused partially or completely according to the level of coverage of knowledge related to meningitis. At the end of the construction of the ontology will come the phase of validation of this ontology by the experts of the meningitis domain. Then, the ontology is used for the annotation of text corpora from social medias that will be processed for extraction of data and events related to meningitis risk factors.

References

1. Abbès, S.B., Zargayouna, H., Nazarenko, A.: Une cartographie de domaine pour guider la construction d'une ressource sémantique. In: 23es Journées francophones d'Ingénierie des Connaissances, Paris, France, pp. 100–103, June 2012
2. Bendaoud, R.: Construction et enrichissement d'une ontologie à partir d'un corpus de textes, pp. 353–358. ARIA, March 2006
3. Borst, W.N.: Construction of Engineering Ontologies for Knowledge Sharing and Reuse. info:eu-repo/semantics/doctoralThesis, Universiteit Twente, Enschede, September 1997
4. Camara, G.: Conception d'un système de veille épidémiologique à base d'ontologies: application à la schistosomiase au Sénégal. Ph.D. thesis. Thèse de doctorat dirigée par Despres, Sylvie et Lô, Moussa Informatique Paris 6 2013 2013PA066619 (2013)
5. Camara, G., Despres, S., Lo, M.: IDOSCHISTO: une extension de l'ontologie noyau des maladies infectieuses (IDO-Core) pour la schistosomiase. In: Faron-Zucker, C. (ed.) IC - 25èmes Journées francophones d'Ingénierie des Connaissances, Clermont-Ferrand, France. Session 1: Construction, peuplement et exploitation d'ontologies, pp. 39–50, May 2014
6. Cowell, L.G., Smith, B.: Infectious Disease Ontology. In: Sintchenko, V. (ed.) Infectious Disease Informatics, pp. 373–395. Springer, New York (2010). https://doi.org/10.1007/978-1-4419-1327-2_19
7. Gómez-Pérez, A., Fernández-López, M., Corcho, O.: Ontological Engineering: With Examples From the Areas of Knowledge Management, e-Commerce and the Semantic Web. Springer, London (2006). https://doi.org/10.1007/b97353
8. Greene, M.: Epidemiological monitoring for emerging infectious diseases. Proc. SPIE **7666**, 76661B-1–76661B-5 (2010)
9. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. Int. J. Hum. Comput. Stud. **43**(5–6), 907–928 (1995)
10. Institut national de la santé et de la recherche médicale (Inserm): Méningites Bactériennes: Stratégies de traitement et de prévention. Les éditions Inserm (1996)
11. Jonquet, C., Musen, M.A., Shah, N.H.: Building a biomedical ontology recommender web service. J. Biomed. Semant. **1**(1), S1 (2010)
12. Kuipers, B.: Qualitative simulation. Artif. Intell. **29**(3), 289–338 (1986)
13. Mizoguchi, R., Ikeda, M.: Towards ontology engineering. J. Jpn. Soc. Artif. Intell. **13**, 9–10 (1998)
14. OMS: Contrôle des épidémies de méningite en Afrique: Guide de référence rapide à l'intention des autorités sanitaires et des soignants (2015)
15. Paireau, J.: Spatial epidemiology of meningococcal meningitis in Niger - influence of climatic, epidemiologic and socio-demographic factors on the spatio-temporal dynamics of epidemics. Theses, Université Pierre et Marie Curie - Paris VI, July 2014
16. Psyché, V., Mendes, O., Bourdeau, J.: Apport de l'ingénierie ontologique aux environnements de formation á distance. Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation (STICEF) **10**, 89–126 (2003)
17. Rouphael, N.G., Stephens, D.S.: Neisseria meningitidis: biology, microbiology, and epidemiology. Methods Mol. Biol. (Clifton, NJ) **799**, 1–20 (2012)
18. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.A., Scheuermann, R.H., Shah, N., Whetzel, P.L., Lewis, S.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat. Biotechnol. **25**, 1251–1255 (2007)

19. Staab, S., Studer, R., Schnurr, H.-P., Sure, Y.: Knowledge processes and ontologies. IEEE Intell. Syst. **16**(1), 26–34 (2001)
20. Studer, R., Richard Benjamins, V., Fensel, D.: Knowledge engineering: principles and methods. IEEE Trans. Data Knowl. Eng. **25**, 161–197 (1998)
21. Suárez-Figueroa, M.C., Gómez-Pérez, A., Fernández-López, M.: The NeOn methodology for ontology engineering. In: Suárez-Figueroa, M.C., Gómez-Pérez, A., Motta, E., Gangemi, A. (eds.) Ontology Engineering in a Networked World, pp. 9–34. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-24794-1_2