Enrichment of Medical Ontologies from Textual Clinical Reports: Towards Improving Linking Human Diseases and Signs

Adama Sow, Abdoulaye Guissé(✉), and Oumar Niang

Information Processing and Intelligent Systems Lab (LTISI),
Computer Science and Telecommunications Engineering Department,
Polytechnic School of Thiès, Thies, Senegal
{asow,aguisse,oniang}@ept.sn

Abstract. Healing a sick patient requires a medical diagnosis before proposing appropriate treatment. With the explosion of medical knowledges, we are interested in their exploitation to help clinician in collecting informations during diagnostic process. This article focuses on the development of a data model targeting knowledges available in both formal and non-formal resources. Our goal is to merge the strengths of all these resources to provide access to a variety of shared knowledges facilitating the identification and association of human diseases and to all of their available relevant characteristic signs such as symptoms and clinical signs.

On one side, we propose an ontology produced from a merging of several existing and open medical ontologies and terminologies. On another side, we exploit real cases of patients whose diagnosis has already been confirmed by clinicians. They are transcribed in textual reports in natural language, and we show that their analysis improves the list of signs of each disease. This work results in a knowledges base loaded from the known target ontologies on the bioportal platform such as DOID, MESH and SNOMED for diseases and, SYMP and CSSO ontologies for all existing signs.

Keywords: Medical diagnosis · Medical ontologies · Ontologies integration · Knowledge engineering · E-health system

1 Introduction

Medical diagnosis, as described in the [2] book, is a patient-centered cognitive activity whose quintessential competence belongs to the clinician. It's a process that consists of a continuous collection of the medical informations that the clinician makes before integrating and interpreting it for the management of his patient's health problems. The diagnosis usually includes four iterative steps: (i)

G. Bassioni et al. (Eds.): InterSol 2019, LNICST 296, pp. 104–115, 2019.
https://doi.org/10.1007/978-3-030-34863-2_10

the acquisition of contextual informations that takes into account antecedents, first physical examinations, and advanced examinations or clinical analysis, (ii) the formulation of hypothesis of potential diagnosis into a list of one or more diseases, (iii) the consistency collected informations with each hypothesis, (iv) and finally the evaluation of each hypothesis to identify and confirm the most certain diagnosis, otherwise the entire process must be taken up by expanding the collection.

This first collection step is as important as it's complex for the clinician, especially when it necessitates quickly recourse to masses of medical knowledge that are constantly exploding on an international scale. It's into the perspective of assisting clinicians in the exploitation of this knowledge, that our research is located. Our goal is more generally to develop a search engine (Fig. 1) that guides access to relevant medical informations at each of diagnostic process step. This engine would make it possible to navigate a knowledge base consisting of a medical ontology and a cases database of clinical diagnosis that have already been validated by clinicians.

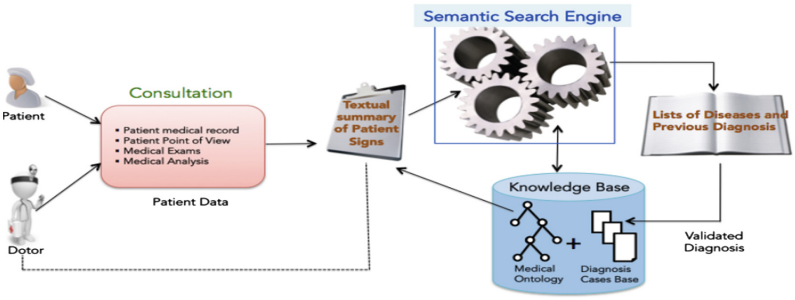

Fig. 1. Description of the medical diagnostic helping process

This article focuses on modeling this knowledge base (KB). It's about producting a data model targeting knowledge available in both formal and non-formal resources. The goal is to merge the strengths of all these resources to provide access to a variety of shared knowledge facilitating identification and association of human diseases to of their available relevant characteristic signs such as symptoms and symptoms and clinical signs.

The core of this KB is an ontology produced from a federation of several existing and open medical ontologies and terminologies. The obtained ontology covers a multitude of descriptive informations of human diseases but also describes the typology and the semantic of signs collected from a patient. Indeed, in existing ontologies we find on the one hand ontologies of diseases associated for each to a list of symptoms whose exhaustiveness is to be clarified, and on the other hand ontologies which conceptualize all the signs that can be appeared in a patient but without no link with diseases. These last ontologies include clinical signs whose values are obtained from in-depth examinations. Ontologies of diseases

aiming at a generic conception do not take into account clinical signs which are nevertheless known into sign ontologies.

We propose to enrich our ontology associating each disease with their clinical signs. This is made possible by exploiting clinical reports for real cases of patients. These reports are usually transcribed into textual format in natural language by clinician. This one systematically archive all data of any patient in his medical folder (MF). Although the MFs are confidential, we have been able to obtain, in collaboration with local hospitals, anonymous descriptive case reports. The analysis of their content makes it possible to identify all the symptoms observed on a patient, as well as the clinical signs that made it possible to confirm an accurate diagnosis. However, these last signs being specific to a given patient, they are associated with a disease by the case that carries them. Cases are stored in the knowledge base. Each disease of our ontology is described by all the signs observed and verified in all the patients carrying this same disease. The association of sickness and its signs is thus continuously nourished as there are new cases of diagnosis.

Thus, in the Sect. 2, we perform a state study on medical ontologies and their use in diagnostic systems. Then, in Sect. 3, we present a selection of reference medical ontologies from which we begin to build an ontology suitable for diagnosis. In the Sects. 4 and 5, we describe our modeling approach to build and enrich our ontology from these target ontologies but also from textual cliniacal reports. Finally, in the Sect. 6, before the conclusion, we show our implementation process to load the extracted data, corresponding to diseases and their characteristic signs, into our resulting ontology.

2 Related Works

In order to establish the diagnosis [2], it is important for the clinician to cross-check all informations on the patient's state of health. It's precisely of the patient's opinion about his condition to identify his pains, physical examinations made by the clinician during consultations, and in-depth examinations (clinics and paraclinical), allow the identification of the most complex and implicit signs.

In medical diagnostic support systems [7], this collection phase is a cognitive activity where the semantics of information is controlled through knowledge known in medical jargon. It's for this very purpose medical ontologies have been conceived [1,4]. These are common medical vocabularies based on shared concepts facilitate the interoperability of documents between stakeholders in the field and especially the development of knowledge. Medical ontologies represent an evolution of medical thesauri; they do not limit themselves to defining terminologies but it goes further by clearly modeling medical entities such as diseases, their characteristic signs, their known treatments, or the hospital processes of patient care.

We are interested here in the medical ontologies of human diseases. The list is long and each ontology has its own specificities. But overall most of the known diseases are covered and each refers to a concept grouping its various nominative terms and synonyms, its different definitions and textual axioms and its

characteristic signs. These include, among others, clinical signs and symptoms, but also possibly the causative agent of the disease, the mode of transmission, and localization in human anatomy. Also, taxonomic (or hierarchical) links are defined from among disease concepts to classify them into disease categories. This is facilitated by the fact that these ontologies are implemented in formal languages, such as OWL (Ontology Web Language), based on the principle of conceptual graphs, object-oriented concept and description logic.

Medical diagnostic support aids are expert systems where medical ontologies can be used as a knowledge base [5,7,8]. They are exploited globally for decision support: either to facilitate the comprehension of the terms present in the documents and the medical reports, or to allow the reasoning and the search for information in particular, in the diagnostic process, when it comes to identifying diseases associated with a given symptom or the characteristic symptoms of a specific disease. They have also been used to alert clinicians about the effects of chemicals on the treatment of certain diseases.

Thus, since the diagnostic process is based on the reasoning around diseases and their characteristic signs, the current difficulty, with regard to existing disease ontologies, lies in the fact that these signs are listed in a non-exhaustive manner [5] and not very formal [6]. Only the most common symptoms are stated in these ontologies and their presence varies from one patient to another. Moreover, the clinical signs take values at a patient are not even taken into account. There are, however, ontologies specific to the conceptualization of the signs [1,5] but they are not associated with diseases.

We are not aiming here to build an ontology from non-formal resources [3] but our goal is to merge the strengths of several existing ontologies in order to have an ontology sufficiently provided in terms of diseases and to associate with each of them all of its relevant signs and appearing in most of the patients who have been affected by the same diseases. This association has already been the subject of research. Indeed, [6], propose in their ontology project Disy, to give to the clinicians latitude to cite for each disease all its signs. The work of [5] offers an integration of ontologies in order to group together for each disease all of its signs present in these target resources. For our part, our proposal is similar to that of the latter authors in that we are also looking for a federation of ontologies of human diseases and signs in order to constitute a news that is adapted for medical diagnosis. However, despite this desire for federation, current ontologies are still not large enough to describe in detail the diseases with all of their characteristic signs. To overcome this, we try to focus on the analysis of real cases of patients who have already been diagnosed and whose clinicians have transcribed the entire process in textual reports. This analysis then makes it possible to list new signs, hitherto not yet taken into account in existing medical ontologies.

3 Ontologies for Medical Diagnosis

Constitution of our ontology consists of a federation of a set of ontologies around a structure unifying all human diseases as well as their characteristic signs. The

diseases correspond to the possible diagnosis. The signs are those can be identified on a patient in order to conclude on a specific diagnosis that can refer to one or more diseases. Diseases are organized in a hierarchical way. They and their derived forms are grouped into categories, which may themselves be subcategories of diseases. The diseases are lexicalized in order to have for each disease the set of the most known nominative terms and their synonyms. For each disease, it will be important to keep all definitions in order to have the most shared semantics. Most of the known signs of each disease are formally listed from those available in the target medical ontologies.

We analyze and exploite here medical ontologies made available to the public via the BioPortal platform. We chose DOID[1], MESH[2], SNOMED[3] as disease ontologies, as well as SYMP[4], and CSSO[5] as ontologies of signs.

DOID ontology (Disease Ontology) serves us as a reference ontology. It proposes a hierarchy of *10389* human diseases and disease categories. With the Fig. 2, we can see each disease has a unique identifier (*rdf:about*), and is classified in one or more categories (*rdfs:subClassOf*). The disease of *Hepatitis A* belongs to the category "*DOID_37*" of ("*skin diseases*") and to the category "*DOID_934*" of ("*viral infectious diseases*"). However, from one identifier to another, there is no description to say that a given identifier refers to a disease or a category of diseases. But, considering the hierarchical graph, all the leaf concepts correspond to the diseases and those who have threads constitute categories.

```
<owl:Class rdf:about="http://purl.obolibrary.org/obo/DOID_12549">
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">hepatitis A</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://purl.obolibrary.org/obo/DOID_37"/>
    <rdfs:subClassOf rdf:resource="http://purl.obolibrary.org/obo/DOID_934"/>
    <obo:IAO_0000115 rdf:datatype="http://www.w3.org/2001/XMLSchema#string">A viral infectious disease that
    results_in inflammation located_in liver, has_material_basis_in Hepatitis A virus,
    which is transmitted_by ingestion of contaminated food or water,
    or transmitted_by direct contact with an infected person.
    The infection has_symptom fever, has_symptom fatigue, has_symptom loss of appetite, has_symptom nausea,
    has_symptom vomiting, has_symptom abdominal pain, has_symptom clay-colored bowel movements,
    has_symptom joint pain, and has_symptom jaundice.</obo:IAO_0000115>
    <oboInOwl:hasAlternativeId rdf:datatype="http://www.w3.org/2001/XMLSchema#string">DOID:12547</oboInOwl:hasAlternativeId>
    <oboInOwl:id rdf:datatype="http://www.w3.org/2001/XMLSchema#string">DOID:12549</oboInOwl:id>
    <oboInOwl:hasDbXref rdf:datatype="http://www.w3.org/2001/XMLSchema#string">MESH:D006506</oboInOwl:hasDbXref>
    <oboInOwl:hasDbXref rdf:datatype="http://www.w3.org/2001/XMLSchema#string">NCI:C3096</oboInOwl:hasDbXref>
    <oboInOwl:hasDbXref rdf:datatype="http://www.w3.org/2001/XMLSchema#string">SNOMEDCT_US_2016_03_01:154347003</oboInOwl:hasDbXref>
    <oboInOwl:hasDbXref rdf:datatype="http://www.w3.org/2001/XMLSchema#string">SNOMEDCT_US_2016_03_01:40468003</oboInOwl:hasDbXref>
    <oboInOwl:hasDbXref rdf:datatype="http://www.w3.org/2001/XMLSchema#string">UMLS_CUI:C0019159</oboInOwl:hasDbXref>
    <oboInOwl:hasRelatedSynonym rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Viral hepatitis A</oboInOwl:hasRelatedSynonym>
    <oboInOwl:hasExactSynonym rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Viral hepatitis, type A (disorder)</oboInOwl:hasExactSynonym>
    <oboInOwl:hasOBONamespace rdf:datatype="http://www.w3.org/2001/XMLSchema#string">disease_ontology</oboInOwl:hasOBONamespace>
</owl:Class>
```

Fig. 2. *Hepatitis A* disease description in DOID

Each disease in DOID refers (*oboInOwl:hasDbXref*) to the same disease in other ontological bases such as that of the Medical Subject Headings (MESH) terminological resource. It's one of the reference thesauri in the biomedical field.

[1] http://purl.bioontology.org/ontology/DOID.
[2] https://www.nlm.nih.gov/mesh/.
[3] http://purl.bioontology.org/ontology/SNOMEDCT.
[4] http://purl.bioontology.org/ontology/SYMP.
[5] http://purl.bioontology.org/ontology/CSSO.

It's known for the multitudes of synonymous terms proposed as denominations of a disease. Each of the diseases has a preferential term (*prefLabel:hepatitis A*) which is the most used denomination, but also of several synonymous terms (*altLabel:Viral hepatitis A, Viral hepatitis type A, Hepatitis Infectious, Hepatitides Infectious, Infectious Hepatitis, Infectious Hepatitides*). These terms correspond to different hepatitis A nominations around the world. Definitions disease available in MESH will be conserved in our ontology result. Otherwise, the DOID proposes also one tag as definition (*obo:IAO_*) in a semi-formalized language goes a little further in the description of the disease. It's easy to decompose this description from groups of verbal words such as *results_in, located_in, caused_by (or has_material_basis_in), transmitted_by or has_symptom* which refer to the characteristic signs of a disease, corresponding respectively to the manifestation of the disease, to its location in the human anatomy, to the agent at the origin of the disease, to its modes of transmission, and to his symptoms. This list of features is very variant from one disease to another in the DOID, it's always informed.

To overcome this lack of information, we use SNOMED (also referenced with *oboInOwl:hasDbXref*) which is one of the most successful ontologies in the medical field. SNOMED proposes a categorization of the different characteristics of a disease. It offers a rich and varied panorama of seven signs categories: Physical agents, Living organisms, Morphological properties (Symptoms), Biological functions (Clinical Signs), Chemical compounds, Social conditions, Topographic properties.

It is with this in mind that we have to consider the SYPM and CSSO ontologies. The first one is developed in the same project as the DOID, and in the same way as this one for the diseases, SYMP proposes a hierarchical structure complete of all the clinical signs and symptoms, which are also classified in categories of signs. SYMP affixes to each sign a definition referring to how it manifests itself in the patient. The second also brandishes the same goal as the SYMP but it is a little less accomplished. Only the third of SYMP signs are taken into account in CSSO. However, the latter brings a plus, a terminology for each sign. However, none of these two ontologies makes the difference between a clinical sign and a symptom, it's necessary to make the mapping with the categorization of the SNOMED signs.

4 Data Model Stucture

The different data formats of the ontologies we have selected are implemented with the W3C standards of the Semantic Web around the RDF, RDFS and OWL languages. So to facilitate the recovery of targeted data on each of these resources, we propose a structure (Fig. 3) using the same technologies and which inherits from them the same conceptual formalisms.

The structure is disease-centric (*Disease Class*) with all informations classes necessary for understanding the disease as well as the recommendation of potential diagnosis. Each disease is identified (*categorized_in*) in one or more categories (*SetOfDiseases Class*). Each disease is associated (*named*) with a set of

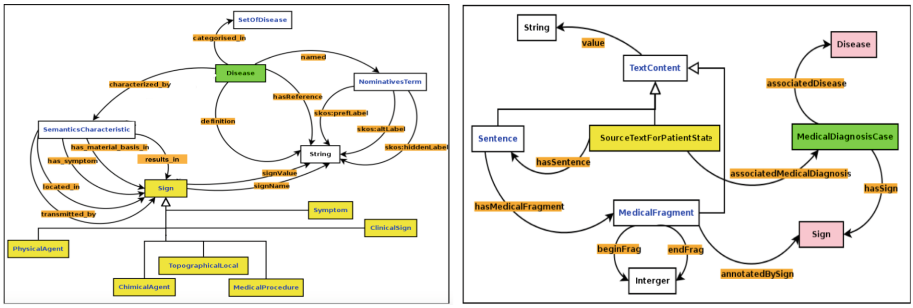

Fig. 3. Overall view data structure of our medical ontology combining diseases and signs

nominative terms (*NominativeTerms*) synonyms, from the preferred term (skos: prefLabel), to alternative terms (skos:altLabel, skos:hiddenLabel). Each disease is associated (characterized_by) with a set of semantic characteristics (Semantic-Characteristics Class) and through the relations *has_symptom, transmitted_by, located_in, caused_by, results_in* refer respectively to different types of signs such as Symptom or ClinicalSign, PhysicalAgent, TopographicalLocate, PhysicalAgent, ChimicalAgent or Symptom (morphological elements) or MedicalProcedure (Medical Procedure).

Each sign has a name and possibly a value, especially in the case of measurable clinical signs. Each of the signs classes, identifiable in SNOMED, group and list all the possible signs, but a given disease is associated only with the most common signs, the other signs are attached on a specific patient case for same diagnosis, and varie from one case to another. Moreover, in the overall data structure (Fig. 3) of the diagnostic recommendation engine, we can see that the patient is materialized by the textual description of his state of health (*SourceTextForPatientState*), and is associated with a medical diagnostic case (*MedicalDiagnosisCase*). The latter is linked (*associatedDisease*) to a disease based on a set of signs (*hasSign*). As a result the data structure implemented here, stores both an ontology of diseases and signs, but also the case base of validated diagnoses and especially links between diseases and their descriptive signs through these latter cases.

5 Contributions of Clinical Reports

After having established our federation of ontologies, we are discovering to what extent we can enrich this ontology from the analysis of medical reports of diagnostic cases that have already been validated by clinicians. We consider here a case of a patient diagnosed with Hepatitis A[6]. This disease is of viral ori-

[6] Example from http://www.immunologyclinic.com/.

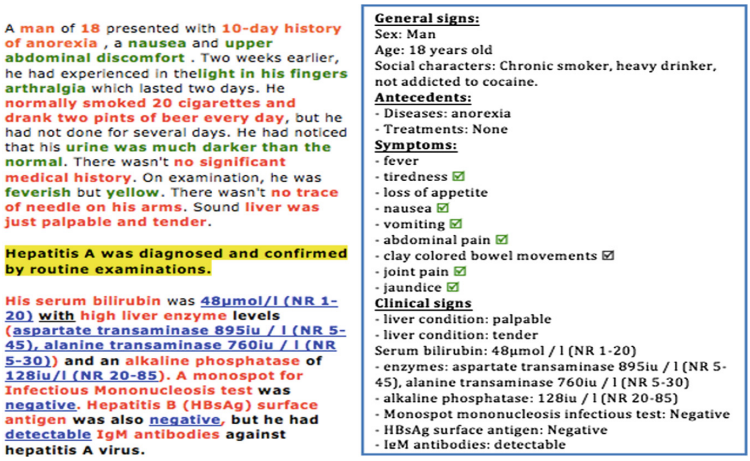

Fig. 4. Textual report of real case of Hepatitis A

gin designating inflammation of the liver. It's listed among sexually transmitted infections is in the top ten (10) of most dangerous diseases in Senegal[7].

The case, taken as an example (Fig. 4), is transcribed in a textual report which includes the symptomatic description of the state of health of a patient whose diagnosis is then confirmed after a set of in-depth examinations. These types of reports explode in the registers (digital or not) of clinicians and gives a visibility on the signs necessary for diagnosis confirmation. Indeed, this report crosses all the characteristic signs allowing to conclude on Hepatitis A disease, and we will discover that its analysis extends our ontology because it makes it possible to associate a given disease with the set of relevant signs.

First, several types of signs present in this report. For Hepatitis A, only 9 general symptoms appear in the ontology. These are fever, fatigue, loss of appetite, nausea, vomiting, abdominal pain, clay colored bowel movements, joint pain, and jaundice. And only 7/9 are therefore identifiable for this case and correspond to the first observable signs in the patient. Other signs, although listed in the ontology (from SYMP and CSSO) and not associated with Hepatitis A, correspond to the general signs indicating its sex, age, and excesses, but also to antecedents and clinicals signs. These result from in-depth examinations and refer to nominative terms and values. In the end, more than 16 signs are added to those who describe Hepatitis A in the ontology.

Extraction of the signs from this textual report poses two problems: the identification of the known general symptoms of the ontology for Hepatitis A, and specific signs to the patient and not listed for this same disease. In both cases, there is a problem of NLP especially since the text is transcribed in natural language. The terms referring to symptoms are not difficult to detect according to detailed lexicons (preferential terms and synonyms) affixed to each sign. For

[7] http://www.who.int/countries/sen/en/.

other signs, in addition to being named in the text, it's imperative to extract their values. They refer to specific named entities and the assignment of one of these signs to a value have to go through the identification of the relationship (verbal or adjectival) that binds them in the text. It is then necessary to use NLP tools to identify all text fragments that describe a sign or a sign value. This part is not detailed in this article and is the subject of an upcoming one.

The data structure (Fig. 3) shows that our ontology, while listing all the signs that may be present in a patient from the ontologies of target signs, only associates the most common signs to a given disease. Therefore, the specific signs described in the contents of a case are also stored in the ontology but their values can be recorded only in the case of type *"MedicalDiagnosisCase"*, which is associated with it with all the signs present in its content as well as their values, and the diseases (or diseases) to which it corresponds. Consequently, a disease will always be related to all these common symptoms via the ontology of diseases, and to a set of specific signs according to the number of real cases already diagnosed.

6 Results

6.1 Data Selection from Target Medical Ontologies

Data structure (Fig. 3) is loaded by querying the different target ontological resources with the SPARQL query language. These are directly executed on SPARQL EndPoint, open query interfaces for browsing RDF graphs. Here we use BioPortal's. In total we have five (5) SPARQL query patterns that recovery:

- all the diseases which constitute the leaves of the classes starting from the DOID, as well as their definitions starting from MESH;
- all disease categories from the DOID where we select their name, description, and parent categories;
- all nominative terms synonyms of diseases from the DOID, but especially from MESH, are the preferred label, as well as alternative labels for each disease;
- all the basic characteristic signs for each disease from semi-formalized descriptions of the DOID;
- all the nominal terms synonymous of signs: the preferential labels are extracted from SYMP, the alternative labels are extracted from the ontologies CSSO, and SNOMED.

Thus in the Table 1, we show statical description of the ontology resulting from this loading of data. Only the human diseases available in DOID are taken into account, as well as their respective categories. For each disease, about 4 registered nominative terms are listed, which facilitates the identification of diagnosis in the exploited textual clinical reports. As for the signs, each is associated with about 5 nominative terms on average but only the general signs directly index diseases in the ontology, the clinical signs are associated with

a disease only through the real cases where they take value. Knowing that a case as described in a textual clinical report always concludes on a disease and under the condition of the appearance of a precise list of symptoms and clinical signs.

Table 1. Description of resulting diseases and signs ontology

Element types	Ontological object	Target ontologies	Number of elements
Diagnosis			
Maladies	Diseases Class	DOID	6442
Categories	SetOfDiseases Class	DOID	3947
Synonyms diagnosis terms	AnnotationProperty (prefLabel, altLabel, hiddenLabel)	DOID, MESH	27586
Signs			
Symptoms and Clinical Signs	Symptom Class and CinicalSign Class - subClassOf Sign Class	SYMP	942
Other Signs	PhysicalAgent Class, ChemicalAgent Class, TopographicalLocate Class, MedicalProcedure Class:subClassOf Class Sign	DOID, SNOMED	6020
Synonyms signs terms	AnnotationProperty (prefLabel, altLabel)	CSSO, SNOMED	4710

6.2 Data Selection from Clinical Reports

Selection of data from the clinical reports involves their loading into the case base as described in the data structure. This process is based on the extraction of the signs present on each report. The cases are then formalized as RIF rules[8] that are compatibles with the ontology manipulation languages we use here such as RDF and OWL. Each case is described in two sides: the premise that refers to identified signs for the case and the conclusion corresponds to the diagnosed disease. Thus, for the extraction of signs, textual clinical reports are annotated with NLP tools such as NooJ[9] and Clamp[10] with regard to the ontology of diseases and signs. This work is the subject of another paper.

In the experimental setting of this work, we use on the one hand, ten (10) real cases of patients who have already been diagnosed. The examples chosen are different cases on tropical diseases[11] and allow us to visualize the contribution of cases. Indeed, on the Table 2 we can notice that for each disease, there is a precise number of general symptoms indicated by the ontology but the totality of them are not present in this patient. In addition, new symptoms identified in the ontology and not associated with the disease are emerging, as well as clinical

[8] https://www.w3.org/TR/rif-overview/.
[9] http://www.nooj-association.org/.
[10] https://clamp.uth.edu/.
[11] Examples in http://medecinetropicale.free.fr/.

Table 2. Symptoms and clinical signs into sample clinical reports

Disease	Symptoms linking with diseasee	Symptoms present in example case	New added symptoms	Clinical signs
Hepatitis A	9	7	7	9
Cholera	5	3	10	2
Rougeaole	6	4	11	3
Dengue	10	5	11	20
Tetanus	4	3	12	4
Malaria	6	4	8	24
Syphilis	5	2	12	14
Chikungunya	9	5	7	29
Typhoid fever	8	5	8	3
Meningitis	9	4	5	7

signs specific to each patient. For example, for the case that refers to Hepatitis A, of the 9 symptoms that appear in the ontology and associated with this disease, only 7 are identified, and in addition 7 other new symptoms are detected as well as the clinical signs.

On the other hand, we used a larger sample of 156 cases of the same Hepatitis A disease. The Table 3 shows symptom appetition rates and test intervals measuring clinical signs. This shows the importance of clinical reports, especially in the context of medical diagnostic assistance. It's possible to classify the characteristic features of each disease by order of appearance in most patients who have already been diagnosed. For clinical signs, the intervals indicate, as reports are added in the case base, what are the most frequent minimum and maximum values.

Table 3. Rate appearance of some symptoms and clinical signs on Hepatitis A clinical reports dataset

Symptoms	Fever	Fatigue	Loss-of-appetite	Vomiting	Abdo-pain	c-c-b-movs	Joint-pain
% appearance	39.1	64.1	20.5	50.0	12.8	11.5	19.2
Clinical signs	Liver-big	Liver-firm	Bilirubin	Aspartate	Alk-phos	Albumin	Protime
% or interval appearance	76.9	38,5	0.3 to 8	14 to 648	26 to 296	2.1 to 6.4	0 to 100

7 Conclusion

In this article, the problem is focused on the establishment of a medical diagnostic support system based on open and shared ontology resources. It's a question

here of the constitution of a central ontology federating a set of ontologies and medical terminologies targets, which answer the need for information in order to facilitate the task of the clinician in the identification of the potential diagnosis, among which he will have the latitude of choose or validate the most reliable knowingly. This type of system does not replace the clinician.

We have therefore proposed a federation methodology around a data structure of RDF graph type facilitating the recovery of human diseases and their most relevant characteristic signs, from targeted ontologies but also from an analysis of real cases of confirmed diagnosis. In the end, we have an ontology of diseases and signs should serve as a knowledge base in the search engine we aim for. The work in perspective would be to validate this ontology by the actors in the field but this will only be done to assess its relevance and consistency in its role for the engine, which is to identify present signs at a patient, and find relevant diseases as diagnosis.

References

1. Anbarasi, M., Naveen, P., Selvaganapathi, S., Nowsath, M.: Ontology based medical diagnosis decision support system. Int. J. Eng. Res. Technol. (2013)
2. Balogh, E.P., Miller, B.T., Ball, J.R.: Improving diagnosis in health care. In: National Academies of Sciences, Engineering, and Medicine. The National Academies Press, Washington, DC (2015)
3. Charlet, J., Declerck, G., Dhombres, F., Gayet, P., Miroux, P., Vandenbussche, P.: Construire une ontologie médicale pour la recherche d'information: problématiques terminologiques et de modélisation (2012)
4. Hoehndorf, R., Schofield, P., Gkoutos, G.: The role of ontologies in biological and biomedical research: a functional perspective. Brief. Bioinf. J. **16**, 1069–1080 (2015)
5. Mohammed, O., Benlamri, R., Fong, S.: Building a diseases symptoms ontology for medical diagnosis: an integrative approach. In: IEEE International Conference on Future Generation Commnication Technology (FGCT 2012), Dcembre 2012
6. Oberkampf, H., Zillner, S., Bauer, B.: Interpreting patient data using medical background knowledge. In: 3rd International Conference on Biomedical Ontology (ICB0), 21–25 July 2012, Austria (2012)
7. S-Ortiz, J.A.R., Jimenez, A.L., Cater, J., Malends, C.A.: Ontology-based knowledge representation for supporting medical decisions. Rech. Comput. Sci. **68**, 127–136 (2013)
8. Valencia-García, R., Lagos-Ortiz, K., Medina-Moreira, J., Paredes-Valverde, M.A., Espinoza-Morán, W.: An ontology-based decision support system for the diagnosis of plant diseases. J. Inf. Technol. Res. **10**(4), 45–55 (2017)