

Proposition of an Ontology of Diseases and Their Signs from Medical Ontologies Integration

Adama Sow, Abdoulaye Guissé, Oumar Niang

Abstract—To assist medical diagnosis, we propose a federation of several existing and open medical ontologies and terminologies. The goal is to merge the strengths of all these resources to provide clinicians the access to a variety of shared knowledges that can facilitate identification and association of human diseases and all of their available characteristic signs such as symptoms and clinical signs. This work results to an integration model loaded from target known ontologies of the biportal platform such as DOID, MESH, and SNOMED for diseases selection, SYMP, and CSSO for all existing signs.

Keywords—Medical decision, medical ontologies, ontologies integration, linked data, knowledge ingeneering, e-health system.

I. INTRODUCTION

FOLLOWING a consultation and before prescribing a treatment to a patient, a clinician is often led to establish a medical diagnosis. It collects relevant clinical data through interrogation and examination, selection and ordering of signs and symptoms to suggest hypotheses that may relate to several potentials diagnoses. His knowledge and experience allow him to decide and validate the good diagnosis. This task can be complex for the clinician especially when it requires the use of informations he has no access. He will have to cope with the constraint of time and rapidly exploit the masses of medical knowledge that are constantly exploding on an international scale.

This research work is placed in a more global context of development of a recommendation search engine of relevant medical diagnostics. The goal is to enable clinicians to quickly access medical knowledge. The engine implements an expert system, it is based on a knowledge base consisting of a database of diagnosis cases validated by various clinicians and an ontology, whose constitution is the subject of this article. It is an ontology grouping the various human diseases that exist, as well as their symptoms and clinical signs. It is built from a integration of various medical dictionaries of ontological and terminological types. This ontology is used for the semantic comprehension of the description of the state of health of the patient during consultation visit, but also for the recommendation of potentials diagnosis for his case.

The semantic engine (Fig. 1) for its part covers a structured process (Fig. 1) ranging from the support of information on the patient to the proposal of relevant diagnostics to the clinician. Indeed, he is called by the engine to transcribe in a text

Adama Sow, Abdoulaye Guissé and Oumar Niang are with the Information Processing and Intelligent Systems Lab (LTISI), Computer Science and Telecommunications Engineering Department, Polytechnic School of Thies, Sénégal (e-mail: asow@ept.sn, aguisse@ept.sn, oniang@ept.sn).

summary the data of the patient at the end of the consultation. These correspond to the informations on the patient's state of health such as his personal opinion, the uses examinations made by the clinician, the results of the in-depth analyzes, and the patient's medical folder notifying his precedents. The summary is then annotated with regard to our diseases ontology in order to identify symptoms and clinical signs, and possibly old patient diagnoses. The engine will automatically be able to recommend several possible diagnoses, from which the clinician decides and validates the most reliable diagnosis for his patient.

Thus, into Section II, we take a look of existing medical ontologies and their use in medical diagnostic systems. Then, we present, in Section III, our methodology of constitution of our diseases ontology. Finally, in Section IV, before the conclusion, we describe the implementation and the result of this constitution through description of the obtained ontology.

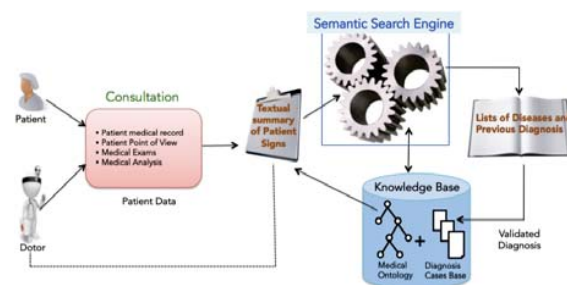


Fig. 1 Description of the medical diagnostic assisting process

II. USING MEDICAL ONTOLOGIES INTO DIAGNOSIS PROCESS

Medical ontologies are more and more used. They are designed with the specific purpose of establishing common medical vocabularies based on shared concepts that facilitate the interoperability of documents between stakeholders in the field and especially the development of knowledge [1]. Medical ontologies represent an evolution of medical thesaurus; they do not limit themselves to defining a terminology but they go further by clearly formalizing medical entities such as diseases and their characteristic signs.

The construction of medical ontologies [1]–[3] is generally supported by unstructured and heterogeneous documentary medical sources, the purpose of which is to arrive at a representation that facilitates the conceptualization and sharing of medical knowledge.

Medical ontologies have been proposed for medical diagnostic assist [2], [4]–[6]. However, it should be remembered that to facilitate one diagnosis, it is important for the clinician to have all informations about the condition of his patient. This is essentially the so-called non-nominal informations, such as the patient's opinion of his condition to identify his pains, examinations of uses that are usually performed, as well as the in-depth examinations (clinical and paraclinical) performed on the patient to identify his most complex and implicit signs.

These informations are generally recorded by the clinician into textual reporting, and their meaning comes in two types of signs [7] presented by the patient. It corresponds to the measurable clinical signs obtained through an examination (body temperature, weight, blood pressure, etc.), and to the apparent symptoms obtained by observation (age, eye color, respiratory problem, hemorrhage, vertigo, allucination, etc.).

In this case, the adapted medical ontologies are human diseases ontologies [1], [6]. The list is long and each ontology has its own specificities, but overall all diseases are covered and each refers to a concept grouping its various nominative terms and synonyms, its different definitions and textual axioms and its characteristics signs. Those include, among other things, symptoms, clinical signs, but also possibly the causative agent of the disease, the mode of transmission, and the localization in the human anatomical. However, signs are listed in a non-exhaustive way, and only the most common ones are available into ontologies of diseases because from one patient to another there are differences. There are, however, specific ontologies to the conceptualization of the signs, but they are not associated with diseases. Also, taxonomic (or hierarchical) links are defined from among disease concepts to classify them into disease categories. This is facilitated by the fact these ontologies are implemented in formal languages based on the principle of graphs, object-oriented design and descriptive logic, such as OWL (Ontology Web Language).

III. METHODOLOGY

The constitution of an ontology of diseases and signs consists of an integration of a set of ontologies around a structure unifying all human diseases as well as their characteristics signs. The diseases correspond to the possibles diagnosis. Signs are those that 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 sub-categories 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 the definitions in order to control the most shared semantics of the disease. Most of the known signs of each disease should be formally listed from those available in the target medical ontologies.

A. Targeted Medical Ontologies

We analyze here medical ontologies made available to the public via BioPortal platforms. We chose:

- the DOID [8], [12], the MESH [14], [13], and the SNOMED [11] as disease ontologies,
- the SYMP [12], [9], and the CSSO [9] as ontologies of signs.

The ontology DOID (Disease Ontology) serves us as a reference ontology. It proposes a hierarchy of 10389 human diseases and disease categories. With Fig. 2, we can see that 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 ("skin disease") and to the category "DOID_934" viral infectious diseases ("viral infectious disease"). 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.

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 is one of the reference thesaurus in the biomedical field. It is known for the multitudes of synonymous terms proposed as denominations of a disease. The terms are in English and are produced by the NLM (U.S. National Library of Medicine), but translation into other languages is also provide in several countries particularly in French with the work of Inserm (French National Institute of Health and Medical Research). 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*). Those terms correspond to different hepatitis A nominations around the medical world.

MESH is also known for its role as a medical dictionary of reference through these explicit definitions in human language. The DOID proposes a definition (*obo:IAO_0000115*) in a semi-formalized language that goes a little further in the description of the disease. It is also 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 symponies. This list of features is very variant from one disease to another in the DOID, they are not always all taken into account.

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, it was built from the meta-thesaurus UMLS (Unified Medical

```

<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 Description of Hepatitis A disease into DOID

Language System). It is very referenced as it covers all fields of medicine. SNOMED proposes a categorization of the different characteristics of a disease. It offers a rich and varied panorama of categories of terms as described in Table I.

All of these features currently exist as a list of terms in SNOMED. They are, except of course the categories "Living Organisms" and "Topographic Elements", signs that can be tracked in a sick patient. But for a doctor the priority lies in the observation of symptoms ("Morphological elements") and the identification and measurement of clinical signs ("biological functions"). All other categories of information are complementary to facilitate decision-making.

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, SYPM proposes a hierarchical structure complete of all the clinical signs and symptoms, which are also classified in categories of signs. SYPM affixes to each sign a definition referring to how it manifests itself in the patient. The second also brandishes the same goal as the SYPM but it is a little less accomplished. Only the third of SYPM signs are taken into account in CSSO. However, the latter brings a plus, a terminology for each sign. For example, for the sign *Fatigue* of Hepatitis A (Fig. 2), we have the following synonyms: *Lassitude*, *Tiredness*, *Weariness*. However, none of these two ontologies makes the difference between a clinical sign and a symptom, it will be necessary to make the mapping with the categorization of the SNOMED signs.

B. Data Model

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 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 (*SemanticCharacteristics 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 (Table I), 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. 4) 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*).

IV. RESULTS

This 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 [10], open query interfaces for browsing RDF graphs. Here we use BioPortal's. In total we have five (5) SPARQL query patterns that recovery:

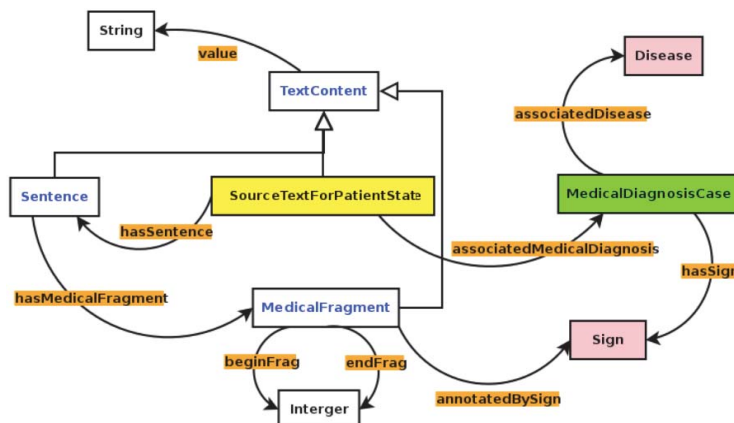


Fig. 4 Overall data structure to diagnostics recommendation system

TABLE II
DESCRIPTION OF THE ACTUAL STATE OF OUR ONTOLOGY OF DISEASES AND SIGNS

Element Types	Ontological object	Target ontologies	Number of elements
Diagnosis			
Maladies	Diseases Class	DOID	6442
Categories	SetOfDiseases Class	DOID	3947
Synonyms Terms	AnnotationProperty (prefLabel, altLabel, hiddenLabel)	DOID, MESH	27586
Signs			
Symptoms and Clinical Signs	Symptom Class and CincinalSign Class - subClassOf Sign Class	SYMP	942
Other Signs	PhysicalAgent Class, ChemicalAgent Class, TopographicalLocate Class, MedicalProcedure Class:subClassOf Class Sign	DOID, SNOMED	6020
Synonyms Terms	AnnotationProperty (prefLabel, altLabel)	CSSO, SNOMED	1346

```
<Disease rdf:about="&ourOnto;Hepatitis_A">
  <categorized_in rdf:resource="&ourOnto;SET_37"/>
  <categorized_in rdf:resource="&ourOnto;SET_934"/>
  <definition xml:lang="en">
    Inflammation of the liver in humans caused by a
    member of the hepatovirus genus, human hepatitis A virus.
    It can be transmitted through fecal contamination of food or water.
  </definition>
  <named rdf:resource="&ourOnto;Terms_12549"></named>
  <characterized_by rdf:resource="&ourOnto;SC_12549"></characterized_by>
  <hasReference=http://purl.obolibrary.org/obo/DOID_12549/></hasReference>
  <hasReference=http://purl.bioontology.org/ontology/MESH/D006506/></hasReference>
</Disease>
```

Fig. 5 Recovery of human diseases

```
<NominativesTerm rdf:about="&ourOnto;Terms_12549">
  <skos:prefLabel xml:lang="en">Hepatitis A</skos:prefLabel>
  <skos:altLabel xml:lang="en">Viral hepatitis A</skos:altLabel>
  <skos:altLabel xml:lang="en">Viral hepatitis type A</skos:altLabel>
  <skos:altLabel xml:lang="en">Hepatitis Infectious</skos:altLabel>
  <skos:altLabel xml:lang="en">Hepatitis Infectious</skos:altLabel>
  <skos:altLabel xml:lang="en">Infectious Hepatitis</skos:altLabel>
  <skos:altLabel xml:lang="en">Infectious Hepatitis</skos:altLabel>
  <hasReference=http://purl.obolibrary.org/obo/DOID_12549/></hasReference>
  <hasReference=http://purl.bioontology.org/ontology/MESH/D006506/></hasReference>
</NominativesTerm>
```

Fig. 7 Recovery of nominative terms of diseases

```
<SetOfDiseases rdf:about="&ourOnto;SET_37">
  <rdfs:label xml:lang="en">skin disease</rdfs:label>
  <rdfs:subClassOf rdf:resource="&ourOnto;SET_16"/>
  <description xml:lang="en">An integumentary system disease that is located in the skin.</description>
  <hasReference=http://purl.obolibrary.org/obo/DOID_12549/></hasReference>
</SetOfDiseases>
```

Fig. 6 Recovery of diseases categories

choose or validate the most reliable to his liking. These types of systems do not replace the doctor.

V. CONCLUSION

Into this article, the problem concerns the setting up of a medical diagnosis support system based on open and shared ontology resources. We are talking here about the constitution of a central ontology federating a set of target medical ontologies and terminologies, which respond to the need for information in order to facilitate the doctor’s task in identifying potential diagnoses, among which he will have the latitude of

We therefore proposed an integration methodology around an RDF graph model that facilitates data retrieval. In the end, we have an ontology of diseases and signs whose complete loading is still in progress, but will be able to serve as a base of knowledge in the recommendation engine that we aim at. The work in the prespective would be to validate this ontology by the actors of the field but this will be done only to evaluate its relevance and consistency in its role for the engine, which is to affix a semantic to the signs present in a patient, and to propose relevant diseases as diagnostics.

```

<SemanticsCharacteristic rdf:about="&ourOnto;SC_12549">
  <results_in rdf:resource="&ourOnto;Symp_inflammation"/><!--Symptôme le plus général-->
  <located_in rdf:resource="&ourOnto;TLSign_liver"/>
  <caused_by rdf:resource="&ourOnto;PASign_Hepatitis-A-virus"/>
  <transmitted_by>direct contact with an infected person</transmitted_by>
  <transmitted_by>ingestion of contaminated food or water</transmitted_by>
  <has_symptom rdf:resource="&ourOnto;Symp_fever"/>
  <has_symptom rdf:resource="&ourOnto;Symp_fatigue"/>
  <has_symptom rdf:resource="&ourOnto;Symp_loss-of-appetite"/>
  <has_symptom rdf:resource="&ourOnto;Symp_nausea"/>
  <has_symptom rdf:resource="&ourOnto;Symp_vomiting"/>
  <has_symptom rdf:resource="&ourOnto;Symp_abdominal-pain"/>
  <has_symptom rdf:resource="&ourOnto;Symp_clay-colored-bowel-movements"/>
  <has_symptom rdf:resource="&ourOnto;Symp_joint-pain"/>
  <has_symptom rdf:resource="&ourOnto;Symp_jaundice"/>
  <hasReference>http://purl.obolibrary.org/obo/D01D_12549</hasReference>
</SemanticsCharacteristic>

```

Fig. 8 Recovery of characteristic signs of diseases

```

<SYMPTOM rdf:about="&ourOnto;Symp_fatigue">
  <skos:prefLabel xml:lang="en">Fatigue</skos:prefLabel>
  <skos:altLabel xml:lang="en">Lassitude</skos:altLabel>
  <skos:altLabel xml:lang="en">Tiredness</skos:altLabel>
  <skos:altLabel xml:lang="en">Weariness</skos:altLabel>
  <hasReference>http://purl.jp/bio/11/csso/CSSO_000119</hasReference>
  <hasReference>http://purl.obolibrary.org/obo/SYMP_0019177</hasReference>
  <hasReference>http://purl.bioontology.org/ontology/SNOMEDCT/84229001</hasReference>
</SYMPTOM>

```

Fig. 9 Recovery of nominative terms of disease's signs's

REFERENCES

- [1] 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. IC 2012.
- [2] Anbarasi M. S., Naveen P., Selvaganapathi S., Mohamed Nowsath Ali I., Ontology Based Medical Diagnosis Decision Support System, In International Journal of Engineering Research and Technology (IJERT), 2013.
- [3] Robert Hoehndorf, Michel Dumontier and Georgios V. Gkoutos, Evaluation of research in biomedical ontologies, In Bioinformatics. vol 14, No 6. 696-712, 2012.
- [4] José A., Reye S-Ortiz, Ana L. Jimenez, Jaime Cater, Cesar A. Malendés, Ontology-based Knowledge Representation for Supporting Medical Decisions, Recherche in Computer Science, 2013.
- [5] Mohammed, Osama and Benlamri, Rachid and Fong, Simon, Building a Diseases Symptoms Ontology for Medical Diagnosis: An Integrative Approach, Décembre 2012, IEEE International Conference on Future Generation Communication Technology (FGCT 2012).
- [6] Hoehndorf, Robert and Schofield, Paul N. and Gkoutos, Georgios V., The role of ontologies in biological and biomedical research: a functional perspective, In Briefings in Bioinformatics Journal, 2015
- [7] Alexander P. Cox, Patrick L. Ray, Mark Jensen, Alexander D. Diehl, Defining 'sign' and 'symptom', IWOOD 2014 Workshop, In ICBO, Houston, TX, USA, October 6-7, 2014.
- [8] Lynn, Schriml and Arze, Cesar and al., Disease Ontology: A backbone for disease semantic integration, Nucleic acids research, 2011.
- [9] Whetzel PL, Noy NF, Shah NH, Alexander PR, al., BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications, Nucleic Acids Res. 2011 Jul, 39(Web Server issue):W541-5, Epub 2011 Jun 14.
- [10] Salvadores M, Horridge M, Alexander PR, Fergerson RW, Musen MA, and Noy NF, Using SPARQL to Query BioPortal Ontologies and Metadata, International Semantic Web Conference (Boston US) LNCS 7650, pp 180195, 2012.
- [11] Cornet, Ronald and De Keize, Nicolette, Forty years of SNOMED: a literature review, In First European Conference on SNOMED CT, 1-3 October 2006, Copenhagen, Denmark.
- [12] Smith, Barry and Ashburner, Michael and al., The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, In Nature Biotechnology, 2007, Volume 25.
- [13] Olivier Bodenreider, The Unified Medical Language System (UMLS): integrating biomedical terminology, in Nucleic Acids Research, 2004, Vol. 32.
- [14] Karen Patrias and Dan Wendling, Citing Medicine, 2nd edition, In National Library of Medicine, National Institutes of Health Bethesda (MD): National Library of Medicine (US); 2007.