2021 International Conference on Platform Technology and Service (PlatCon)

Automatic Stroke Medical Ontology Augmentation with Standard Medical Terminology and Unstructured Textual Medical Knowledge

2021 International Conference on Platform Technology and Service (PlatCon) | 978-1-6654-1766-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/PlatCon53246.2021.9680753

Soonhyun Kwon
Department of KSB(Knowledge-converged Super Brain) Convergence Research
Electronics and Telecommunications Research Institute (ETRI)
Daejeon, Republic of Korea
kwonshzzang@etri.re.kr

Jaehak Yu
Department of KSB(Knowledge-converged Super Brain) Convergence Research
Electronics and Telecommunications Research Institute (ETRI)
Daejeon, Republic of Korea
dbzzang@etri.re.kr

Sejin Park
Research Team for Health & Safety Convergence
Korea Research Institute of Standards and Science (KRISS)
Daejeon, Republic of Korea
sjpark@kriss.re.kr

Jong-Arm Jun
Department of KSB(Knowledge-converged Super Brain) Convergence Research
Electronics and Telecommunications Research Institute (ETRI)
Daejeon, Republic of Korea
jajun@etri.re.kr

Cheol-Sig Pyo
Department of KSB(Knowledge-converged Super Brain) Convergence Research
Electronics and Telecommunications Research Institute (ETRI)
Daejeon, Republic of Korea
cspyo@etri.re.kr

Abstract— **The need for medical ontology to provide stroke medical knowledge is increasing as much research has recently been conducted to predict stroke diseases using AI technology quickly. Medical ontology serves as a medical explanation of predictions in conjunction with methods of analysis using machine learning and deep learning to analyze clinical data obtained from the medical field, medical imaging devices (MRI, CT, ultrasound, etc.). However, the existing medical ontology focused on is-a relationships in taxonomy to define the classification system for diseases, symptoms, and anatomical structures. This medical ontology is insufficient to explain complex organic relationships to disease-symptom-body-patients, a knowledge structure for predicting disease. Furthermore, although professional standard terms exist in medicine, electronic medical records (EMR), electronic health records (EHR) medical professional books, and medical papers that use common terms to express professional are mostly unstructured forms. To overcome this limitation, in this paper, we propose a stroke medical ontology automatic augmentation method via unstructured text medical knowledge using the lowest instance-level medical term ontology and top-level schema-level medical ontology for stroke disease prediction through standard medical terms. The proposed method extracts and stores data in resource description framework (RDF) form with unstructured textual medical knowledge (medical papers, medical professional books), health data, and syntactic morphology analysis of clinical data, with instance-level ontologies capable of linking top-level schema to standard medical terminology ontologies such as the international classification diseases (ICD), systematized nomenclature of medicine - clinical terms (SNOMED-CT), and foundational model of anatomy (FMA). We also use a medical data-knowledge mapping DB that stores the frequency of extracted data torches for the abstraction of extracted RDF data.**

Keywords— stroke medical ontology; ontology augmentation; standard medical terminology; unstructured textual medical knowledge; stroke

I. Introduction

Recently, many studies have been conducted to predict stroke diseases using AI technology quickly [1][2]. Many studies predict stroke diseases by analyzing clinical data obtained at medical sites and image and image data received from medical imaging devices (MRI, CT, ultrasound) [2][3]. Furthermore, deep learning, which shows good performance in various domains, has recently demonstrated excellent speech recognition or medical image and image analysis of patients for disease prediction in the medical field, and its number has been rapidly increasing [4][5].

However, existing data-dependent analysis methods, Machine Learning, and Deep Learning applied classification or regression methods to provide only numerical probability values or discrete results. These prediction results alone lack medical explanations that the patient is interested in. Also, by not providing patients with medical interpretation information based on the outcome of the disease prediction, it cannot offer seriousness to the current situation. It is necessary to provide medical interpretation of the predicted results, anatomical information that causes the disease, the area of the body to be affected by it, the expected symptoms, and the details of the condition. Therefore, the use of medical knowledge in AI-based disease prediction services is essential.

What matters in any field of knowledge is a clear understanding of the entity under consideration. The essential substance in medical knowledge is an understanding of the nature of the disease. The expression of relationships with other entities directly or indirectly related to the disease is paramount to explain the substance of the disease. The relationship between disease and the anatomy of the human body, the relationship between disease and symptoms, the relationship between the patient's holding disease (base disease), and the treatment of the disease is essential to explain the substance of the disease.

An ontology defines the relationship between these **entities and entities [6][7]. Ontology is a knowledge**

modeling method that provides a computer-understandable way of understanding the concepts and attributes of a particular domain and the relationship between them [7]. Due to this nature, the concepts of disease-symptom-body-patients in the medical field are suitable for expressing complex cross-referenced knowledge structures, and many medical ontologies are currently being developed and used.

Existing medical ontologies are divided into three primary groups [8]. The first form is a medical term ontology developed based on standard terms for consistency in medical terms. Examples include the ICD [9], SNOMED-CT [10], FMA [11], and the international nursing classification for nursing practice (ICNP) [12]. The second form is the Actor Profile ontology to identify positions and responsibilities in the healthcare organization or healthcare sector [13]. Finally, the third ontology is an ontology that defines terms and processes for managing workflows in clinical decision systems [14].

However, the existing medical ontology focuses on taxonomy relationships to define classification systems for diseases, symptoms, and anatomy. This medical ontology is insufficient to define complex organic relationships to disease- symptom-body-patients, a knowledge structure for predicting disease. Furthermore, although professional standard terms exist in medicine, electronic medical records (EMR), electronic health records (EHR) medical professional books, and medical papers that use common terms to express professional are mostly unstructured forms.

To overcome this limitation, in this paper, we propose a stroke medical ontology automatic augmentation method via unstructured text medical knowledge using the lowest instance-level medical term ontology and top-level schema-level medical ontology for stroke disease prediction through standard medical terms. The proposed method defines the top schema ontology, which represents stroke-related medical knowledge (body structure, body function, disease, lesion, patient), and the bottom medical terms ontology that encompasses the standard medical terms ICD, SNOMED-CT, FMA, and ICND. Stroke ontology is augmented by collecting stroke-related unstructured text medical knowledge (medical papers, medical literature) for semantic linkage between the top and bottom ontology and generating RDF [15] data through syntactic, morphological analysis, and knowledge mapping. Furthermore, for vertical knowledge abstraction of generated RDF-type medical knowledge, we proceed with the vocabulary expansion of ontology relationships (subclass relationships: owl:subclassOf [16]) and relationship abstractions (rdfs:subpropertyOf [17]). To this end, the frequency of medical data-knowledge mapping DB is used.

II. RELATED WORKS

Stroke ontologies, which allow us to systematically define stroke prediction, treatment, emergencies, have been developed to support stochastic statistical methods, machine learning, and deep learning. They monitor data on the status of patients, which is collected in real-time, this allows these systems to recognize anomalies and provide analysis to support medical staff in their decision making. They have also been used as an ex post facto method of suggesting medical decisions to medical staff as well as of defining clinical data, drug information, and specific stroke treatments.

For real-time prediction of mini-strokes, ontologies have been developed to define an emergency instruction manual in systems where computational stochastic risk indices are determined using bio-signal data collected from wearable sensors and mobile applications [18]. The proposed ontology was used to perform preemptive treatment by providing real-time information to emergency room medical staff about mini-stroke patients' symptoms. To support acupuncture plans for ischemic stroke patients, an ontology that provides acupuncture treatment was also developed, this ontology considers clinical stages, acupuncture treatments, stimulation points, stimulation intensity, and previous treatment records [19]. The stroke treatment ontology was developed to also provide optimized therapy to patients with hemiparesis, a sequela of stroke [20].

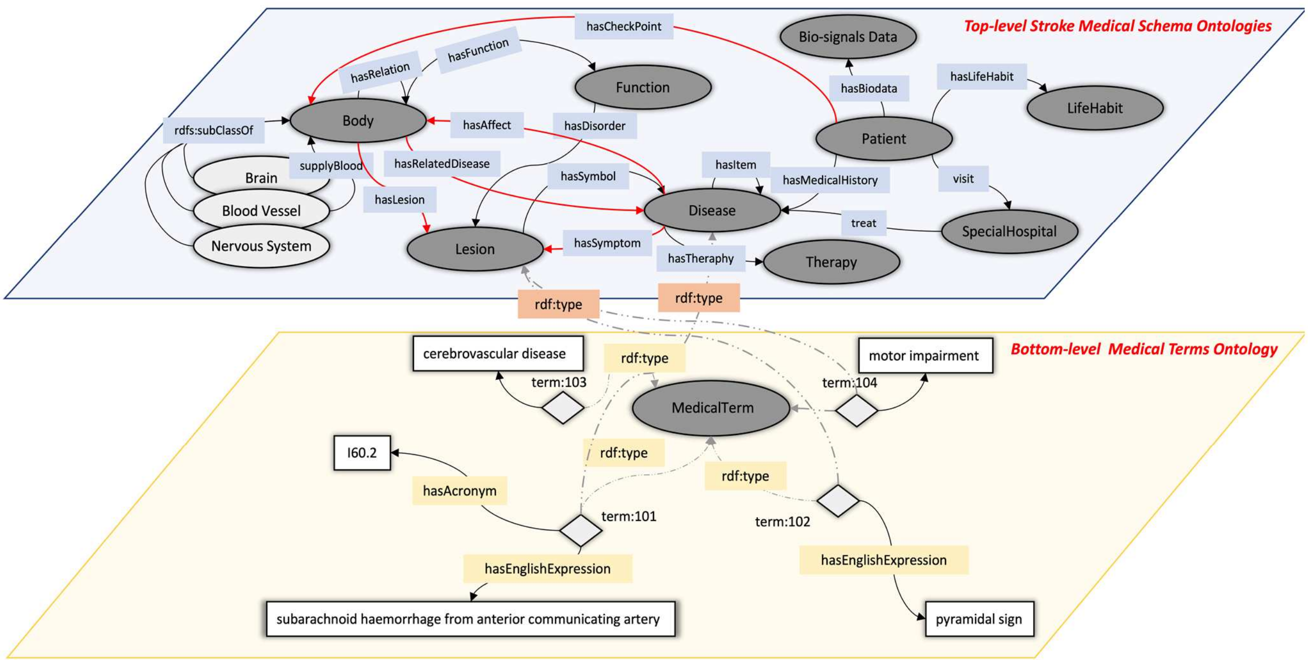

Fig. 1. Overview of the Stroke Medical Ontology

III. Development of The Stroke Medical Ontology Augmentation System

This section describes the basic structure of stroke medical ontology and the stroke medical ontology augmentation system structure.

A. Stroke Medical Ontology

Stroke medical ontology is largely divided into the top-level stroke medical schema ontology, which includes concepts and relationships related to stroke, and the bottom medical term ontology, which encompasses standard medical terms (Fig. 1).

The top-level stoke medical schema ontologies include top-level concepts of the human body, lesions, diseases, and treatment methods associated with stroke. It also includes patients' bio-signal data, lifestyle, and specialized hospitals the patients visited. Each concept extends to more complex concepts through the is-a relationship. We also infer the patient's checkpoint information based on the patient's disease and the body information. Based on the schema structure encompassing ICD, ICND, FMA, SNOMED-CT, the medical term ontology performs semantic linkage with schema-level stroke medical ontology for stroke definition through semantic annotation. By separating the schema layer from the instance layer, it has the advantage of being able to develop abstractly in future ontology-based inference rules and SPARQL protocol and RDF query language (SPARQL) query writing.

B. Storke Medical Ontology Augmentation System

The stroke medical ontology augmentation System loads top-level schema medical ontology and builds instance ontology through the lowest medical term ontology schema encompassing standard medical terms. It also extracts medical data from unstructured textual medical knowledge (medical papers, medical professional books, health data, clinical data, machine learning, and prediction results of deep learning). The extracted medical data automatically augments the medical ontology by linking the top schema-level medical ontology with the bottom instance-level ontology and generating ontology knowledge data that can be abstracted. It also constructs a Medical Knowledge Abstract for abstractions (concepts and relationships extensions) of ontologies generated using medical data-medical mapping DB (Fig. 2).

IV. Procedure of The Stroke Medical Ontology Augmentation

The stroke medical ontology augmentation system described in the previous section is performed by loading unstructured medical knowledge with the top stroke schema ontology and the bottom medical term ontology loaded (Fig. 3). Stroke medical ontology augmentation procedures are as follows:

- The schema level stroke medical ontology loader loads schema ontologies that define relationships between top-level concepts and concepts for disease definition in the medical field.
- The medical term annotator collects standard medical terms and loads them, for instance construction. The loaded standard medical terms construct the bottom-level medical ontology instances by creating instances concerning the bottom-level medical ontology schema structure.
- The unstructured textual medical knowledge collector collects knowledge from medical knowledge (medical papers, medical professional books), health data, clinical data, machine learning, and deep learning.
- The medical data extractor divides the collected unstructured text medical knowledge into syntactic

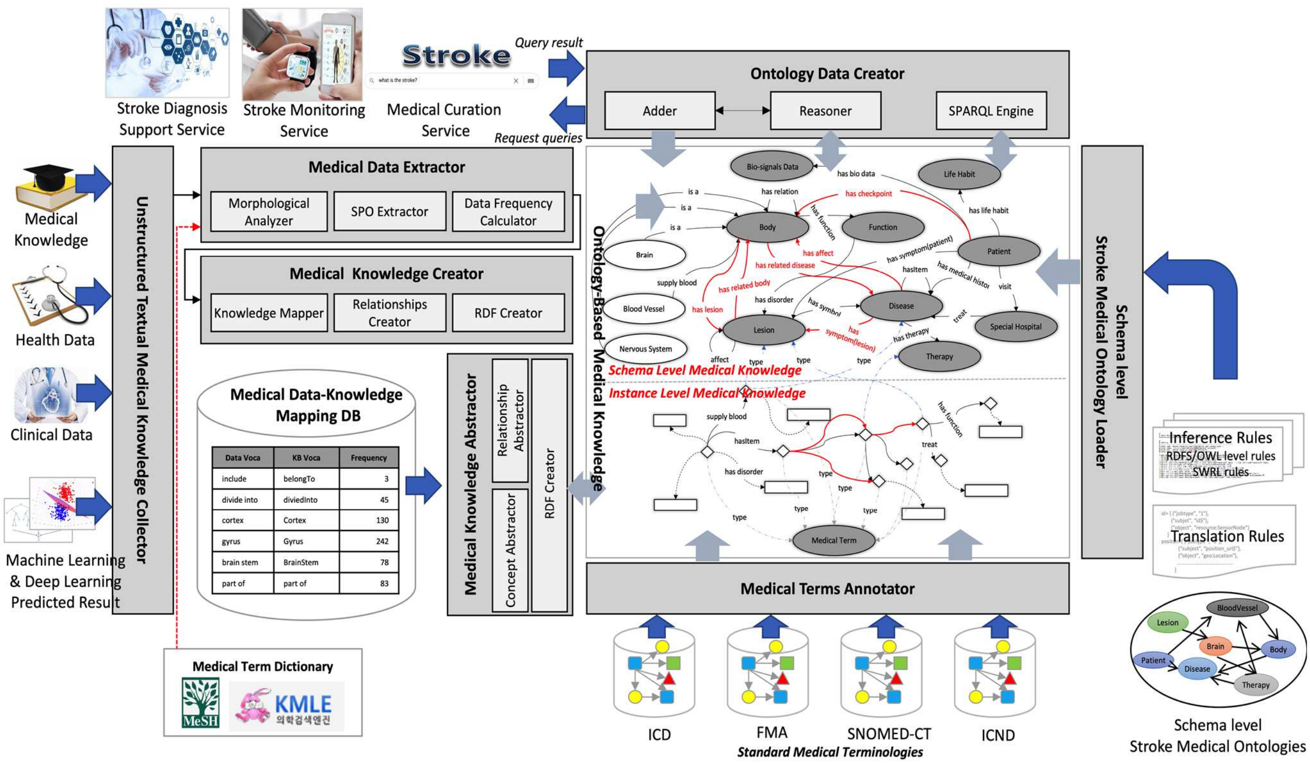

Fig. 2. The Structure of the Stroke Medical Ontology Augmentation System

units and generates a subject, a predicate, and an object in each phrase. The basis of each extracted phrase is replaced through medical term dictionary search to generate three RDF data candidates. Furthermore, it stores medical data-frequency information for a set of RDF data candidates in a medical data-to-knowledge mapping DB for medical ontology augmentation via a medical knowledge abstract.

- The medical knowledge data creator performs knowledge mapping with the lowest medical term ontology based on a set of RDF data candidates generated from the medical data extractor. At this point, three RDF data candidates that are not knowledge-mapping are discarded. It generates knowledge data, RDF, for the syntax that becomes knowledge mapping and stores it in a medical knowledge ontology. Furthermore, in the future, it stores medical data-medical knowledge data-frequency information in a medical data-to-knowledge mapping DB for processing of the medical knowledge abstract.

- Based on medical data-knowledge mapping DB, medical knowledge abstractors augment knowledge by linking top-level schema ontologies with bottom-instance ontologies by adding conceptual and object property information for each data with high frequency. It changes existing instance information to concept for concept abstraction, and adds conceptual subclass information (owl:subClassOf [16]) for conceptual hierarchy. It also defines the up-and-down relationship (rdfs:subPropertyOf [17]) to the top schema ontology based on the relationship stored in the bottom-level instance ontology for relation abstraction, and adds domain relationship (rdfs:domain [17]) and range relationship (rdfs:range [17]) information for each relationship.

V. Conclusion

We present a technique to automatically collect and analyze unstructured text medical knowledge using the bottom-level, instance-level medical term ontology and the top schema-level medical ontology for stroke disease prediction via standard medical terms. Furthermore, for vertical abstraction of generated medical knowledge, knowledge refinement work was carried out through the frequency of medical data-knowledge databases to expand concepts, properties, domain relationships, and range relationships. The procedures of the methods presented in this paper are validated by implementing the system through simple textual medical knowledge related to stroke.

In the future, we plan to extend the methods presented in this paper to validate medical knowledge ontology augmentation methods applicable to various disease predictions and stroke and methods presented in this paper through much unstructured medical knowledge.

Acknowledgment

This work was supported by the National Research Council of Science & Technology (NST) grant by the Korea government (MSIP) (No. CRC-15-05-ETRI).

References

[1] A. D. Dolatabadi, S. E. Z. Khadem, and B. M. Asl, “Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM,” Computer methods and programs in biomedicine, vol. 138, pp. 117-126, January, 2017.

[2] “A GA-based feature selection and parameter optimization of an ANN in diagnosing breast cancer,” Pattern Analysis and Applications, vol. 18, no. 4, pp. 861-870, May, 2015.

[3] S. M. McKinney, et al. “International evaluation of an AI system for breast cancer screening,” Nature, vol. 577, no. 7788, pp. 89-94, January, 2020.

[4] J. Yu, S. H. Kwon, C. B. Ho, R. K. Lee, N. S. Kim, C. S. Pyo, and S. Park, “Stroke Disease Prediction based on Deep Learning using the

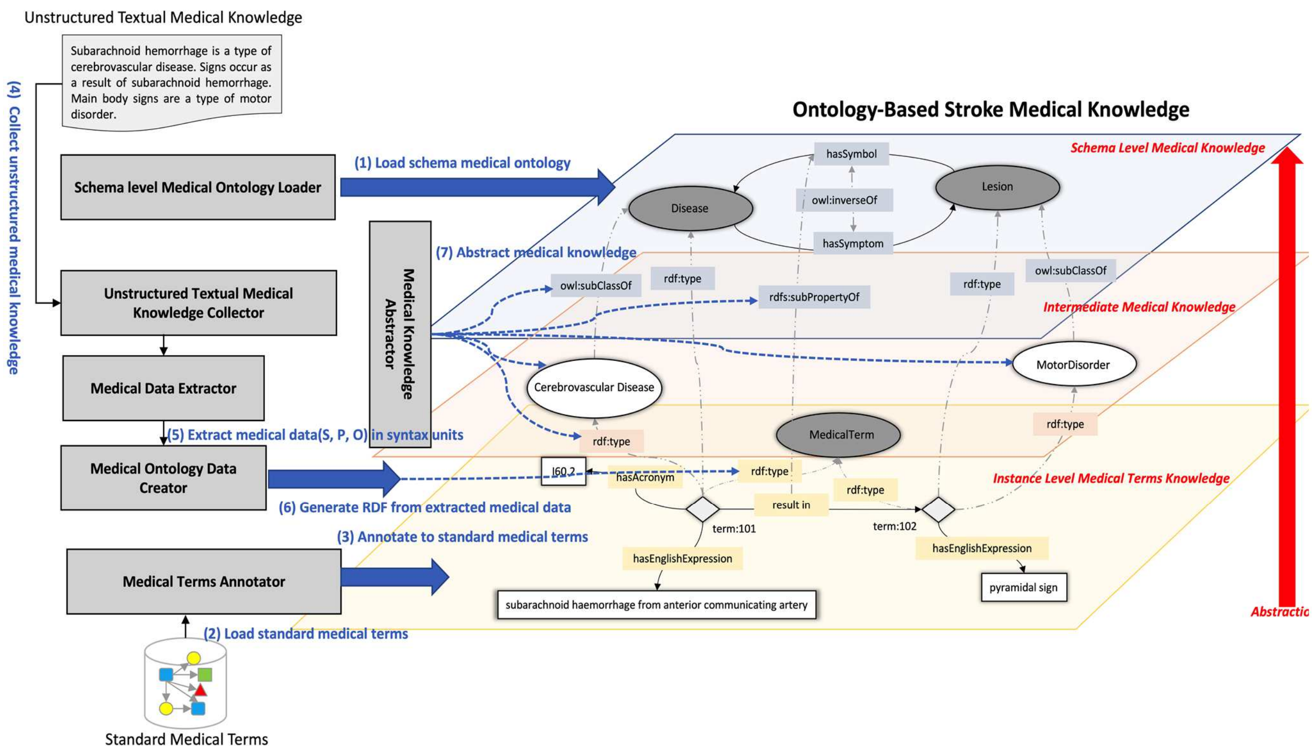

Fig. 3. Procedure Of the Stroke Medical Ontology Augmentation

Elderly Cohort DB," Journal of Digital Contents Society, vol. 21, no. 5, pp. 1191-1200, February, 2020.

[5] J. Yu, S. Park, S. H. Kwon, C. M. B. Ho, C. S. Pyo, and H. Lee, "AI-Based Stroke Disease Prediction System Using Real-Time Electromyography Signals,", vol. 10, no. 19, 6791, September, 2020.

[6] T. R. Gruber, "A translation approach to portable ontology specifications," Knowledge acquisition, vol. 5, no. pp. 199-220, June, 1993.

[7] T. R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing?," International journal of human-computer studies, vol. 43 no. 5-6, pp. 907-928, November, 1995.

[8] H. B. Elhadj, et al. "Do-Care: A dynamic ontology reasoning based healthcare monitoring system," Future Generation Computer Systems, vol. 118, pp. 417-431, May, 2021.

[9] J. M. Rodrigues, et al. "ICD-11 and SNOMED CT Common Ontology: circulatory system," MIE, pp. 1043-1047, January, 2014.

[10] L. Cui, et al. "Mining non-lattice subgraphs for detecting missing hierarchical relations and concepts in SNOMED CT," Journal of the American Medical Informatics Association, vol. 24, no. 4, pp. 788-798, July, 2017.

[11] R. Messaoudi, et al. "Ontologies for liver diseases representation: A systematic literature review," Journal of digital imaging, vol. 33, no. 3, pp. 563-573, June, 2020.

[12] A. Coenen, and T. Y. Kim. "Development of terminology subsets using ICNP®.," International journal of medical informatics, vol. 79, no. 7, pp. 530-538, July, 2010.

[13] A.Valls, K. Gibert, D. Sánchez, and M. Batet, "Using ontologies for structuring organizational knowledge in Home Care assistance," international journal of medical informatics, vol. 79, no. 5, pp. 370-387, May, 2010.

[14] P. I. Dissanayake, K. C. Tiago, and J. C. James, "Using clinical reasoning ontologies to make smarter clinical decision support systems: a systematic review and data synthesis," Journal of the American Medical Informatics Association, vol. 27, no. 1 pp. 159-174, January, 2020.

[15] World Wide Web Consortium(W3C), Resource Description Framework(RDF)[Online]. Available: https://www.w3.org/RDF(downloaded 2021, June. 12)

[16] World Wide Web Consortium(W3C), Web Ontology Language (OWL) [Online]. Available: https://www.w3.org/OWL (downloaded 2021, June. 15)

[17] World Wide Web Consortium(W3C), RDF Schema (RDFS) [Online]. Available: https://www.w3.org/2001/sw/wiki/RDFS (downloaded 2021, April. 11)

[18] H. Mcheick, et al. "Stroke prediction context-aware health care system," 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). IEEE, 2016.

[19] Y. Shen, C. Joël, and J. A. Armelle, "Decision Support System for Acupuncture Treatment of Ischemic Stroke," International Conference on Advanced Data Mining and Applications, Springer, Cham, 2020.

[20] C. Townsend, et al. "Ontology-based knowledge acquisition for neuromotor functional recovery in stroke," 2010 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW). IEEE, 2010.