Knowledge Extraction for the Web of Things (KE4WoT)

[WWW 2018 Challenge Summary]

Amelie Gyrard*, Manas Gaur*, Swati Padhee*, Amit Sheth*
*Kno.e시스, Wright State University, USA
{amelie,swati,manas,amit}@knoesis.org

Mihaela Juganaru-Mathieu△
△MINES Saint-Etienne, Institut Henri Fayol, France
mathieu@emse.fr

The “MyOntoSens” ontology, based on SSN V1 is being standardized as a Technical Specification (TS) within the SmartBAN (Body Area Networks) Technical Committee of the ETSI standardization body [8]. This ontology is relevant to build health applications based on smartphone devices. Smart Appliances REference (SAREF)7 is a European standard supported by ETSI M2M and SmartM2M. It mainly covers the smart building applicative domain. The SAREF ontology has been designed re-using SSN and oneM2M [7]. Schema.org is a well-known schema catalog to structure data on Web pages to describe the location, person, etc. [4]. The IoT Schema.org extension8 is planned; nothing concrete has been developed yet, but discussions are ongoing. The Haystack9 project aims to standardize semantic data models and web services. The Haystack Tagging Ontology10 which employs SSN V1 ontology has been developed [2].

The purpose of this challenge is to automatically extract the redundant knowledge already designed within existing standardizations, WoT applications, and different communities. Most of the existing knowledge extraction techniques are frequently applied to text found within documents or social networks. The main novelty of this challenge is to apply web-based knowledge extraction techniques on models.

2 CHALLENGE DESCRIPTION

The challenge is open to a large audience and complementary web communities: (1) Knowledge extraction experts to detect common patterns within ontologies - Task 1.1. (2) WoT/IoT and healthcare communities willing to discover and study already designed models - Task 1.1. (3) Ontology matching experts to align existing ontologies - Task 1.2. (4) Natural Language Processing researchers - Tasks 2.1 and 2.2. (5) Semantic Web researchers. (6) Any developers and/or data scientists willing to implement innovative methods.

2.1 Task 1: Exploiting the WoT Knowledge Base

The LOV4IoT ontology catalog11 [5] is referencing almost 400 WoT research projects in various domains. For both Task 1.1 and Task 1.2, we provide a set of 6 web services and dump files to easily query ontologies from the LOV4IoT ontology catalog (e.g., healthcare, IoT, agriculture, sensor networks, WoT and smart city ontologies).

1https://webofthings.org/
2http://wiki.knoesis.org/index.php/KE4WoTChallengeWWW2018
3https://www.w3.org/TR/vocab-ssn/
4http://iot.linkeddata.es/def/wot/index-en.html
5http://w3c.github.io/wot/wot-ucr.html#domain-healthcare_and_medical
6http://www.onem2m.org/technical/onem2m-ontologies

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW ’18 Companion, April 23–27, 2018, Lyon, France
© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.
https://doi.org/10.1145/3184558.3192305
2.2 Task 2: Named Entity Recognition and Question Answering in Healthcare Unstructured Text

Named entity recognition (NER) is considered as an essential natural language processing task [1] for entity extraction. The state-of-the-art NER systems lack focus on structured healthcare text due to the complexity to capture context in medical domain. Good quality NER is crucial for accurate Question Answering (Q/A) systems and hence we have designed the following two sequential tasks.

Task 2.1: Extracting named entities using ontologies. In this challenge, we encourage the idea of NER from unstructured healthcare text obtained from tweets. We categorize the named entities as follows: (1) Disease Entity: It is the name of the disease that is stated explicitly in the text. (2) Severity Entity (a severe form of disease entity): It is a disease entity of etiological origin from a relatively mild disease entity. (3) Trigger Entity: It refers to the cause (entity/substance/environmental condition) of the disease. For example, pollen, weather, cough can cause asthma. (4) Location Entity: It refers to the location affected in human anatomy. For instance, bones, muscles, nose, lungs, etc. (5) Procedure/Treatment/Device: These are entities that define a procedure, treatment or device used by the patient or clinician to treat the disease entity stated in the text. For example, an inhaler is a device to treat asthma. (6) Control: It is a dichotomous concept whose value is “yes” when the tweet talks about disease control, reduction in severity or reduced frequency of asthmatic attacks. This category is created for supporting the question answering task.

Impact: This task motivates the participants to develop a quality named entity extraction technique for unstructured healthcare text. It can be leveraged by a question answering system for healthcare domain.

Task 2.2: Question Answering (Q/A) System. In this task, the participants will be required to develop a Q/A system for healthcare tweet questions using their developed NER module developed in Task 2.1. They are encouraged to utilize existing domain knowledge sources (e.g., SNOMED, DBpedia, etc.) to enhance the efficiency of their model. We provide 25 test questions to evaluate their performance.14

Impact: Q/A systems have been an apogee of research in linked open data [9]. This task encourages research in Q/A for a less explored, complex and critical healthcare domain.

An illustrative example of this task: Tweet: Patients with severely uncontrolled asthma derive the most benefit from dupilumab. Question: Does dupilumab control asthma? Answer: Yes

3 LESSONS LEARNT

The KE4WoT challenge has attracted significant interest from the research community15. Interested participants gave feedback regarding time constraint to submit papers. We plan to pursue this research challenge idea and extend it.

4 ACKNOWLEDGMENTS

This work is partially funded by a bilateral research convention with ENGIE Research & Development, the ANR 14-CE42-0029 OpenSensingCity project, Hazards SEES NSF Award EAR 1520870, and KHealth NIH 1 R01 HD087132-01. Thanks to Dr. Antoine Zimmermann for his feedback.

REFERENCES


12http://lov4iot.appspot.com/?p=OntologyExtractionKE4WoTChallengeWWW2018
13http://lov4iot.appspot.com/?p=OntologyAlignmentKE4WoTChallengeWWW2018
15https://goo.gl/WKcMV