A Semantic Technologies Toolkit for Bridging Early Diagnosis and Treatment in Brain Diseases

Report from the Ongoing EU-funded Research Project ALAMEDA

Christoniki Maga-Nteve[1], Efstratios Kontopoulos[2], Nikos Tsolakis[1], Panagiotis Mitzias[2], Ioannis Katakis[3], Evangelos Mathioudis[3], Konstantinos Avgerinakis[2], Georgios Meditskos[4], Anastasios Karakostas[1], Stefanos Vrochidis[1] and Ioannis Kompatsiaris[1]

[1] Information Technologies Institute, Centre of Research & Technology Hellas, Greece
{chmaga, tsolakin, akarakos, vrochidis}@iti.gr
[2] Catalink Limited, Nicosia, Cyprus
{e.kontopoulos, pmitzias, koafgeri}@catalink.eu
[3] Department of Computer Science, School of Sciences and Engineering, University of Nicosia, 2417, Nicosia, Cyprus
katakis.i@unic.ac.cy, mathioudis.e@live.unic.ac.cy
[4] School of Informatics, Aristotle University of Thessaloniki, Greece
gmeditsk@csd.auth.gr

Abstract. Semantic Web technologies are increasingly being deployed in various e-health scenarios, prominently due to their inherent capacity to harmonize heterogeneous information from diverse sources and devices, as well as their capability to provide meaningful interpretations and higher-level insights. This paper reports on ongoing work in the recently started EU-funded project ALAMEDA towards a semantic toolkit for bridging the gap between early diagnosis and treatment in a variety of brain diseases. The toolkit comprises (a) a semantic model serving as the underlying knowledge base for the toolkit; (b) a flexible semantic data fusion framework; (c) a conversational agent for interacting with human users and other components of the ALAMEDA system.

Keywords: Ontologies, Semantic Data Fusion, Conversational Agent, Brain Disease, e-health.

1 Introduction

Semantic Web technologies are rapidly gaining popularity in the domain of e-health applications, where these technologies substantially facilitate the harmonization of data coming from multiple sources and devices, as well as its meaningful interpretation, providing, thus, context awareness and access to rich higher-level interpretations and insights. This paper reports on ongoing work within the context of the ALAMEDA EU-funded project (`https://alamedaproject.eu/`) aimed at the development of a sophisticated semantic toolkit for bridging the gap between early diagnosis and treatment in a variety of brain diseases: Parkinson's Disease, Multiple Sclerosis, and stroke. The key component of the toolkit is the ALAMEDA semantic model, which consists of a set of interconnected ontologies, for semantically representing all pertinent concepts and entities. Operating on-top of the semantic model, two additional components of the toolkit are also presented: (a) the semantic data fusion framework for populating the ontology with instance data, and (b) the conversational agent that utilizes the ontology as a common vocabulary facilitating interaction with human users and other components of the ALAMEDA system.

2 Conceptual Semantic Model

The overarching goal of our semantic model is to represent information that is made available via the questionnaires and the monitoring modules in the ALAMEDA system, as well as to establish semantic interoperability between the system components.

2.1 Existing Resources

Some of the most commonly used healthcare models for exchanging healthcare information are FHIR-HL7 (`http://hl7.org/fhir/`) and ICD-10 (`https://bioportal.bioontology.org/ontologies/ICD10`). Moreover, the Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT) [1] is a standardized, multilingual vocabulary of clinical terminology for the storage, retrieval, and exchange of electronic health data and for the representation of medical concepts. There also exist disease specific ontologies like PDON, a Parkinson's disease ontology [2], and MSO, a multiple sclerosis ontology [3], and the Dem@care ontologies (`https://dem-care.eu/ontologies/`) for representing knowledge relevant to dementia.

Compared to the aforementioned ontologies that can cover a subset of the respective domains, our proposed ontology seeks to respond to multiple aspects, comprised of modules for representing various needs, and can be easily adjustable and reusable.

2.2 Ontology Design

We relied on the *NeOn methodology* [4] for designing and developing the ALAMEDA ontology. The first phase involves the definition of the ontology requirements and the retrieval of the *Ontology Requirements Specification Document* (*ORSD*). At this point, the contribution of the domain experts was crucial, as they define the use cases and propose optimal matching to ontology requirements. These requirements correspond to a set of *Competency Questions* (*CQs*) [5], which specify what knowledge must be entailed in the ontology. The second phase involves the development of the ontology at a primary level, where the existing ontologies that will be (re)used are defined. The final phase contains the implementation and enrichment of the ALAMEDA ontology.

2.3 Ontology Modules

The ALAMEDA model contains six modules and a main ontology:

- *Model* is the parent of all the hierarchical relations in ALAMEDA.
- *Home* provides information about the behavioural interpretation and reported difficulties in the home environment.
- *Lab* indicates the types of information relevant to the tests, assessments, patient's clinical and experimental records in the lab environment.
- *Person* refers to human users' sociodemographic data and represents information about persons, diseases, gender, educational levels and languages.
- *Event* provides information relevant to the entities and activities that take place in the context of the ALAMEDA use cases. Its design is based on the Event Model F [6] and the Event Ontology[1].
- *Sensors* describes information concerning the type and properties of the sensors used in the ALAMEDA system, which may be fixed on wearable. Its design was strongly influenced by SSN/SOSA [7] and the Smart Home Ontology [8].
- *Time* represents the temporal dimension, namely, the time, duration, and information of the tasks/events taking place within the ALAMEDA context.

3 Semantic Data Fusion

The integration of the inputs from the various heterogeneous sources into the ALAMEDA semantic model will be handled by *CASPAR* (*Structured Data Semantic Exploitation Framework*)[2], our domain-agnostic semantic data fusion framework. CASPAR is based on the ontology population principles presented in previous works of ours [9, 10]. In a nutshell, the tool deploys a set of interconnected mechanisms for ingesting data into a semantic model that incorporate:

- *automated acquisition* of structured data from APIs, databases, messaging buses,
- *mapping* of input data fields to semantic entities (concepts, relationships, etc.),
- *semantic fusion* and population of knowledge into a semantic repository,
- *semantic enrichment* of existing knowledge from Linked Open Data sources,
- *rule-based semantic reasoning* to unveil underlying or generate new knowledge.

CASPAR defines mappings between input data fields and respective ontology concepts for the fusion and population of knowledge through a flexible methodology using a Domain-Specific Language (DSL) based on JSON syntax. The building blocks of a mapping are templates, individuals, and properties:

- A *template* serves as the mechanism for focusing on specific parts of the input. Since large pieces of input can be handled by CASPAR, defining several templates within a mapping that target specific parts allows easier maintenance of the mapping itself.

---

[1] `http://motools.sourceforge.net/event/event.html`

[2] `https://caspar.catalink.eu/`

- A template contains a set of *individuals*, which declare the nodes that need to be created or updated in the Knowledge Graph (KG). From the perspective of using an ontology as the KG schema, an individual is an instance of one or more classes.
- A *property* indicates a desired edge that needs to be created in the KG, connecting a node with another node or with a literal value. Properties in mappings are defined by a set of *predicates*, meaning the relationship types of the ontology, and *objects*, which indicate the value that will be given to the property.

4 Conversational Agent

Nowadays, more and more chatbot platforms are emerging with the aim to provide personalized health services. Through a chat with the patient the chatbot gathers information related to the symptoms and the person's condition and then provides a report to the clinician, assisting this way in better managing the patient's health condition.

4.1 Chatbots in Healthcare

There exist several chatbot applications in healthcare. Puffbot is a conversational agent that helps people with asthma [11]. EVA is another chatbot that helps people to self-manage their diabetes, by educating them, interacting with them, and giving them recommendations [12]. HOLMeS, on the other hand, serves as a medical recommendation system designed to autonomously handle discussions with patients and chat and act like a human physician helping patients in choosing their disease prevention pathway [13].

4.2 A Chatbot for Brain Diseases in the ALAMEDA Project

The *ALAMEDA chatbot* will gather lifestyle data generated from static and wearable sensors and will identify changes in the users' lifestyle. Unusual measurements will trigger the agent to ask the patient questions, and categorize these measurements based on the ALAMEDA semantic model (see Section 2). We have identified the following requirements for the ALAMEDA chatbot:

- *Non-intrusive*: The conversational agent should not interfere with the patients' daily activities. It should be up to the user to define how much information she wants to share and when is the appropriate time to be inquired for information.
- *Adaptive and personalized*: The agent should be unique, tailored for its user (patient or caregiver), and should be able to adapt based on their needs. Based on a sentiment analysis component [14], the agent will be able to sense the user's dissatisfaction.
- *Information disambiguation*: In the cases of missing, erroneous, or conflicting input, the conversational agent should be able to resolve the issue by sending an appropriate enquiry to the user [15].

In order for all the intelligent ALAMEDA components to communicate, a common semantic dictionary is necessary. In the case of the agent, such a resource is necessary whenever communication with human participants required. The ALAMEDA semantic model (see Section 2) serves as the common language between artificial agents and humans (patients and caregivers) (see Fig. 1).

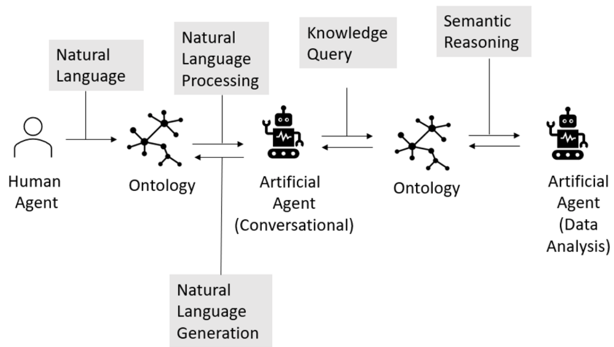

Fig. 1. The ALAMEDA ontology as a common vocabulary between human and artificial agents.

In order to illustrate the orchestration of the above components, let us present the following use-case scenario:

1. **(Intention)** User wishes to report to the chatbot that he feels his heart pacing fast.
2. **(Natural Language)** User types "*I feel my heart pacing fast*" in the chatbot app.
3. **(Natural Language Understanding)** The conversational agent identifies the intent that the user wants to declare an event of type: `increased_heart_rate`.
4. **(Knowledge Query)** As a knowledge management process, an `increased_heart_rate` event declared by the user initiates a knowledge query. The system has to confirm the event with the sensors.
5. **(Semantic Reasoning)** The Semantic Reasoning component retrieves the data of the patient (stored in their mobile device) to check if there are *recent* sensor data annotated with the `increased_heart_rate` event entity from the event ontology.
6. **(Natural Language Generation)** If not, the chatbot re-assures the user that everything looks normal, and suggests relaxing. The heart rate will be monitored again in 5 minutes and reported back to the patient.

5 Conclusions and Next Steps

This paper reported on ongoing work within the ALAMEDA EU-funded project involving the deployment of a toolkit based on semantic technologies that will constitute the backbone of the overall e-health system for brain diseases. We presented the current state of the ALAMEDA semantic model, the semantic data fusion framework, and the conversational agent. Work on these components is still at an early stage; thus, we will be able to report on more advanced iterations of those in future work of ours.

Acknowledgment

ALAMEDA project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No GA101017558.

References

1. Donnelly, K. (2006). SNOMED-CT: The advanced terminology and coding system for eHealth. Studies in health technology and informatics, 121, 279.
2. Younesi, E. et al. (2015). PDON: Parkinson's disease ontology for representation and modeling of the Parkinson's disease knowledge domain. Theoretical Biology and Medical Modelling, 12(1), 1-17.
3. Malhotra, A. et al. (2015). Knowledge retrieval from pubmed abstracts and electronic medical records with the multiple sclerosis ontology. PloS one, 10(2), e0116718.
4. Suárez-Figueroa, M. C., Gómez-Pérez, A., & Fernández-López, M. (2012). The NeOn methodology for ontology engineering. In Ontology engineering in a networked world (pp. 9-34). Springer, Berlin, Heidelberg.
5. Bezerra, C., Freitas, F., & Santana, F. (2013, November). Evaluating ontologies with competency questions. In 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) (Vol. 3, pp. 284-285). IEEE.
6. Scherp, A., Franz, T., Saathoff, C., & Staab, S. (2012). A core ontology on events for representing occurrences in the real world. Multimedia Tools and Applications, 58(2), 293-331.
7. Haller, A. et al. (2019). The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation. Semantic Web, 10(1), 9-32.
8. Sezer, O. B., Can, S. Z., & Dogdu, E. (2015, June). Development of a smart home ontology and the implementation of a semantic sensor network simulator: An Internet of Things approach. In 2015 Int. Conf. on Collaboration Technologies & Systems (pp. 12-18). IEEE.
9. Kontopoulos, E., Mitzias, P., Riga, M., & Kompatsiaris, I. (2017, October). A Domain-Agnostic Tool for Scalable Ontology Population and Enrichment from Diverse Linked Data Sources. In DAMDID/RCDL (pp. 184-190).
10. Riga, M., Mitzias, P., Kontopoulos, E., & Kompatsiaris, I. (2017, October). PROPheT–Ontology Population and Semantic Enrichment from Linked Data Sources. In Int. Conf. on Data Analytics and Management in Data Intensive Domains (pp. 157-168). Springer, Cham.
11. Teixeira, M. S., Maran, V., & Dragoni, M. (2021, March). The interplay of a conversational ontology and AI planning for health dialogue management. In Proceedings of the 36th Annual ACM Symposium on Applied Computing (pp. 611-619).
12. Anastasiadou, M., Alexiadis, A., Polychronidou, E., Votis, K., & Tzovaras, D. (2020, October). A prototype educational virtual assistant for diabetes management. In 2020 IEEE 20th Int. Conf. on Bioinformatics and Bioengineering (BIBE) (pp. 999-1004). IEEE.
13. Amato, F. et al. (2017, November). Chatbots Meet eHealth: Automatizing Healthcare. In WAIAH@AI*IA (pp. 40-49).
14. Agathangelou, P., & Katakis, I. (2019, April). A hybrid deep learning network for modelling opinionated content. In Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing (pp. 1051-1053).
15. Panagiotou, N. et al. (2016, September). Intelligent urban data monitoring for smart cities. In Joint European conference on machine learning and knowledge discovery in databases (pp. 177-192). Springer, Cham.