Merging Ontologies and Data from Electronic Health Records

Subjects: Computer Science, Information Systems

Contributor: Salvatore Calcagno , Andrea Calvagna , Emiliano Tramontana , Gabriella Verga

The Electronic Health Record (EHR) is a system for collecting and storing patient medical records as data that can be mechanically accessed, hence facilitating and assisting the medical decision-making process. EHRs exist in several formats, and each format lists thousands of keywords to classify patients data. The keywords are specific and are medical jargon; hence, data classification is very accurate. As the keywords constituting the formats of medical records express concepts by means of specific jargon without definitions or references, their proper use is left to clinicians and could be affected by their background, hence the interpretation of data could become slow or less accurate than that desired.

Electronic Health Records   HL7   ontology   data integration   data analysis

1. Introduction

The Electronic Health Record (EHR) can be seen as the natural evolution of the physical medical record, which consisted of hard-copy medical documents for a patient. Nowadays, electronic versions of medical records allow citizens to easily track the entire history of their healthcare life and share it with healthcare professionals. Moreover, doctors are better equipped to keep up with the large amount of data, more easily access them, and deliver a significant improvement in health services for patients. To facilitate the syntactic analysis of such data and some level of interoperability between different software systems documents follow a standard format: Health Level 7 (HL7) Clinical Document Architecture (CDA) [1][2]. Such a standard is a valuable tool for data sharing; however, its adoption has been slow due to the length and complexity of the defined formats. In fact, given the wide spectrum and complexity of the medical field, several templates have been defined within HL7, one for each type of clinical document (http://cdasearch.hl7.org, last accessed on 10 November 2023).

Documents produced according to HL7 present two critical issues that derive from its formats: (i) complexity in the organisation of data, which makes it difficult to read data, i.e., navigating the formats to extract data is a complex task, and this can be a significant impediment when developing software systems that extract data from an EHR [2]; (ii) though the terms are very specific, they present only the essential pieces of information, without any hints to other related terms or factors, such as, e.g., related diseases, symptoms, possible causes, etc. Moreover, pieces of data are scattered in the document, without evident relationships among them. There is a need for a way to organise data in an EHR to make it possible to discover and suggest potential diseases, causes, etc., related to the medical values and conditions.

To organise knowledge in a domain, some previous approaches have attempted to build ontologies automatically [3]; however, in the medical field it is paramount to have a highly reliable ontology curated by experts. A significant number of well-established ontologies exist in the biomedical realm (http://obofoundry.org, last accessed on 10 November 2023). Although ontologies have been used to guide the diagnosis process, they have not been related to EHR data [4].

2. Merging Ontologies and Data from Electronic Health Records

In recent years, the high complexity of clinical documents has been largely recognised [5], and the importance of a correct interpretation of data for an appropriate diagnosis has been also highlighted [6][7][8]. Moreover, there exist some gaps in the representation of symptoms and clinical data in the structure of EHR clinical documents, because the most significant descriptions of conditions remain in unprocessed free text [9][10][11]. Indeed, the importance of sharing and using EHRs has also been shown [12].

JSON is a standard text-based format for representing structured data based on JavaScript object syntax. It is commonly used for transmitting data in web applications (e.g., sending some data from the server to the client, so it can be displayed on a web page, or vice versa). Many researchers have highlighted that JSON is a notation that is gradually replacing XML due to its relative simplicity, intuitiveness, compactness, and the ability to directly map the native data types of popular programming languages [13][14].

A first prototype conversion tool of clinical documents to a JSON format was introduced by Rinner et al. [15]. The goal of their work was to allow easier access to health data stored in the Austrian information system for EHR (ELGA) and convert CDA documents into Fast Healthcare Interoperability Resources (FHIR) resources, via XML. The limitation of such a proposal is that it implements a basic prototype with limited capabilities for converting from the XML document to an equivalent JSON file. Moreover, they do not interpret or enrich the processed data.

Rousseau et al. [14] aimed at providing a more complete clinical document to facilitate doctors' work by integrating Social Determinants of Health (SDoH) into EHRs, leveraging for this a specific ontology. Specifically, they combine the HL7 standard with the SNOMEDCT ontology (https://bioportal.bioontology.org/ontologies/SNOMEDCT, last accessed 10 January 2024), using JSON as a means to better connect data in medical records to external data sources. Indeed, the authors concluded that the research and development teams of health IT and healthcare services strive to make information in existing semantic frameworks more accessible and connected to EHRs. However, they left the integration of data from other ontologies as future work.

Ontologies are among the most effective and most widely adopted means to represent knowledge in the medical (or any other) domain [16][17][18][19]. Several authors (i.e., [20]) have been using them to facilitate medical decision-making, e.g., to suggest diagnoses and treatments, also taking advantage of probabilistic algorithms applied to the ontology data. However, successfully processing clinical data to navigate complex ontologies down to a possibly correct diagnosis is not at all an easy task to accomplish, because probabilistic algorithms applied to the rigid

complex structure of an ontology can fail when the clinical data are poorly specified or relate to little-known cases. Other authors [21][22][23][24] have been more concerned with expanding the domain of medical knowledge through an accurate review of the related concepts, e.g., the analysis of disease ontology terms or IDs (DOIDs), highlighting the importance of an enriched and complete ontology for the analysis of diseases.

A further obstacle is the complexity of the HL7 CDA standard itself, whose data format is very widespread but also difficult to manage and to apply in the actual daily routine [15]. Specifically, this is the first approach processing the HL7 XML formatted clinical documents in order to produce a single, both enriched and easier to understand (and process), JSON object. This allows EHRs to go beyond the mere need for storing and exchanging clinical records towards a new conception as data objects that can be really effective and practical tools supporting clinicians' everyday work.

In fact, the resulting medical document offers a clearer and more complete overall clinical picture of the patient since it is able to integrate the data in its medical records with data from multiple related medical ontologies. As a consequence, it facilitates the analysis of the patient by being able to provide the definitions of the terms and the relationships among all the basic pieces of clinical data, assisting clinicians in an in-depth analysis of the patient and the construction of a correct diagnosis. Moreover, having to manage a single JSON-formatted document object is of great added value not only because it facilitates its integration into multiple frameworks or software but also because it is more user friendly for automated processing and exchange.

References

1. Henderson, M.L.; Dayhoff, R.E.; Titton, C.P.; Casertano, A. Using IHE and HL7 conformance to specify consistent PACS interoperability for a large multi-center enterprise. J. Healthc. Inf. Manag. 2006, 20, 47.

2. Noumeir, R. Active learning of the HL7 medical standard. J. Digit. Imaging 2019, 32, 354–361.

3. Al-Aswadi, F.N.; Chan, H.Y.; Gan, K.H. Automatic ontology construction from text: A review from shallow to deep learning trend. Artif. Intell. Rev. 2020, 53, 3901–3928.

4. Dissanayake, P.I.; Colicchio, T.K.; Cimino, J.J. Using clinical reasoning ontologies to make smarter clinical decision support systems: A systematic review and data synthesis. J. Am. Med Inform. Assoc. 2020, 27, 159–174.

5. Shaker, N.; Patel, A.; Tozbikian, G.; Parwani, A. Anastomosing hemangioma: A case report of a benign tumor often misdiagnosed as a malignant epithelioid angiosarcoma. Urol. Case Rep. 2022, 42, 102023.

6. Chen, J.J.; Kardon, R.H. Avoiding clinical misinterpretation and artifacts of optical coherence tomography analysis of the optic nerve, retinal nerve fiber layer, and ganglion cell layer. J. Neuro-

Ophthalmol. 2016, 36, 417.

7. Beckmann, C.L.; Keuchel, D.; Soleman, W.O.I.A.; Nürnberg, S.; Böckmann, B. Semantic Integration of BPMN Models and FHIR Data to Enable Personalized Decision Support for Malignant Melanoma. Information 2023, 14, 649.

8. Torres-Silva, E.A.; Rúa, S.; Giraldo-Forero, A.F.; Durango, M.C.; Flórez-Arango, J.F.; Orozco-Duque, A. Classification of Severe Maternal Morbidity from Electronic Health Records Written in Spanish Using Natural Language Processing. Appl. Sci. 2023, 13, 10725.

9. Rosenbloom, S.T.; Denny, J.C.; Xu, H.; Lorenzi, N.; Stead, W.W.; Johnson, K.B. Data from clinical notes: A perspective on the tension between structure and flexible documentation. J. Am. Med. Inform. Assoc. 2011, 18, 181–186.

10. Rousseau, J.F.; Ip, I.K.; Raja, A.S.; Valtchinov, V.I.; Cochon, L.; Schuur, J.D.; Khorasani, R. Can automated retrieval of data from emergency department physician notes enhance the imaging order entry process? Appl. Clin. Inform. 2019, 10, 189–198.

11. Rousseau, J.F.; Ip, I.K.; Raja, A.S.; Schuur, J.D.; Khorasani, R. Can emergency department provider notes help to achieve more dynamic clinical decision support? J. Am. Coll. Emerg. Physicians Open 2020, 1, 1269–1277.

12. Abdelgalil, L.; Mejri, M. HealthBlock: A Framework for a Collaborative Sharing of Electronic Health Records Based on Blockchain. Future Internet 2023, 15, 87.

13. Lin, B.; Chen, Y.; Chen, X.; Yu, Y. Comparison between JSON and XML in Applications Based on AJAX. In Proceedings of the International Conference on Computer Science and Service System, Nanjing, China, 11–13 August 2012; pp. 1174–1177.

14. Rousseau, J.F.; Oliveira, E.; Tierney, W.M.; Khurshid, A. Methods for development and application of data standards in an ontology-driven information model for measuring, managing, and computing social determinants of health for individuals, households, and communities evaluated through an example of asthma. J. Biomed. Inform. 2022, 136, 104241.

15. Rinner, C.; Duftschmid, G. Bridging the gap between HL7 CDA and HL7 FHIR: A JSON based mapping. In Proceedings of the eHealth, Vienna, Austria, 24–25 May 2016; pp. 100–106.

16. Haendel, M.A.; Chute, C.G.; Robinson, P.N. Classification, ontology, and precision medicine. N. Engl. J. Med. 2018, 379, 1452–1462.

17. Grimm, S.; Hitzler, P.; Abecker, A. Knowledge representation and ontologies. In Semantic Web Services; Springer: Berlin/Heidelberg, Germany, 2007; pp. 51–105.

18. Gai, K.; Qiu, M.; Chen, L.C.; Liu, M. Electronic health record error prevention approach using ontology in big data. In Proceedings of the Conference on High Performance Computing and

Communications, Symposium on Cyberspace Safety and Security, and Conference on Embedded Software and Systems, New York, NY, USA, 24–26 August 2015; pp. 752–757.

19. Abdelreheim, M.; Soliman, T.H.A.; Klan, F. A Personalized Ontology Recommendation System to Effectively Support Ontology Development by Reuse. Future Internet 2023, 15, 331.

20. Mahmoodi, S.A.; Mirzaie, K.; Mahmoudi, S.M. A new algorithm to extract hidden rules of gastric cancer data based on ontology. SpringerPlus 2016, 5, 312.

21. Reyes-Ortíz, J.A.; Jiménez, A.L.; Cater, J.; Meléndez, C.A.; Márquez, P.B.; García, M. Ontology-based Knowledge Representation for Supporting Medical Decisions. Res. Comput. Sci. 2013, 68, 127–136.

22. Sow, A.; Guissé, A.; Niang, O. Enrichment of Medical Ontologies from Textual Clinical Reports: Towards Improving Linking Human Diseases and Signs. In Proceedings of the International Conference on Innovations and Interdisciplinary Solutions for Underserved Areas, Cairo, Egypt, 14–15 February 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 104–115.

23. Tramontana, E.; Verga, G. Ontology Enrichment with Text Extracted from Wikipedia. In Proceedings of the International Conference on Software Engineering and Information Management (ICSIM), online, 21–23 January 2022; pp. 113–117.

24. Tramontana, E.; Verga, G. Keeping Researchers Updated by Automatically Enriching an Ontology in the Medical Field. In Proceedings of the International Conference on Computer and Communications Management (ICCCM), Okayama, Japan, 29–31 July 2022; pp. 257–262.