Toward Answering Time-Related Questions from Adverse Event Reports Using Ontology-based Approaches

Cui Tao[†], Guoqian Jiang[†], Kim Clark[‡,§],
Deepak Sharma[†], and Christopher G. Chute[†]

[†]Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN
[‡]Department of Biomedical Engineering, University of Minnesota, Minneapolis, MN
[§]Boston Scientific Corporation, Maple Grove, MN

Abstract. Identifying the time patterns from medical device adverse event reports can help clinical researchers or medical devices manufacturers to understand cause of events and predict future events. Ontology-based approaches can provide a reliable and effective system for representing both events and temporal relations, annotating the useful information from narratives, and querying and reasoning over the data. This paper discuss our vision on such a system, as well as challenges we encountered during our preliminary studies.

1 Introduction

Potential temporal patterns may exist within reports of similar adverse events or similar devices. These patterns may include a similar sequence of events, durations of or between events, or a time/date during which the adverse event occurred. These temporal properties and relationships, however, are often buried within the text of the narrative, requiring an astute observer to detect patterns while reading several reports for the same failure mode. This method for assessing tens of thousands of adverse event reports is time consuming, expensive, and the potential exists for a missed pattern observation or error in interpreting event sequencing. In addition, because temporal relations may require inference if they are not explicitly expressed within the narrative, temporal reasoning is also needed in order to analyze the trends in time. An automated temporal analysis of reports for similar adverse events across similar products of multiple device manufacturers could lead to faster identification of trends, quicker identification of the origin of the adverse event, a more detailed understanding of the events leading up to the failure, and earlier prediction of a future failure based on similarities in event order and/or duration.

In this paper, we introduce our vision on an ontology-based system for answering time-related questions based on adverse event reports from the Food and Drug Administration (FDA) Manufacturer and User Facility Device Experience (MAUDE) database, which contain more than 1.64 millon files as of March of 2011 [4]. Figure 1 shows the system overview. The arrows denote the system

workflow. The hexagons indicate important steps for 1)processing the data: selecting the adverse reports from the MAUDE database, 2)annotating the reports with respect to domain ontologies, and 3)retrieving answers of time-related questions by querying and reasoning over the annotated data. In the rest of paper, we introduce our preliminary studies of such a system and discuss lessons learned for improving the system.

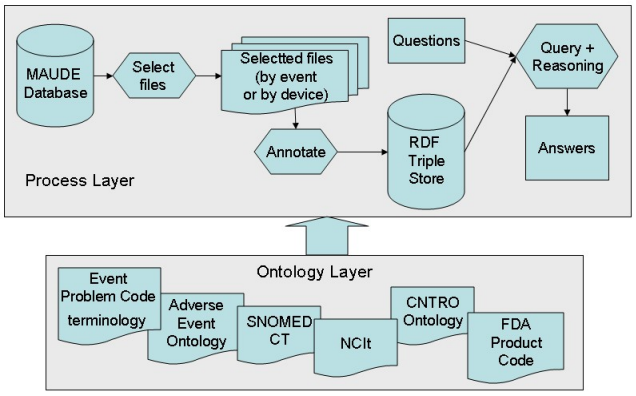

Fig. 1. System Overview

2 Approaches and Challenges

File Selection for devices The MAUDE database provides a way to search files using the FDA product code [3]. Currently the product codes are maintained in a relational database: FDA Product Classification Database. It contains device names and their associated product codes, which identify the generic category of a device for FDA. The database, however, does not contain much hierarchical information of the devices, nor does it contains annotation properties such as synonyms, acronyms, and etc. This makes it difficult to find the right product code(s) for a specific category of devices. For example, to find the code for drug-eluting stent, one has to search by the exact phrase "drug-eluting stent", synonyms such as "drug coated stents" or "medicated stents", or acronym such "DES" would not work. In addition, the product codes do not contain much hierarchical classification. It is hence also difficult to find a category of devices when on a different levels of granularity are used in the classification by FDA. For example, the FDA classification does not provide a category for ventricular assist device (VAD), but rather contains codes for different kinds of VADs. This makes it difficult for computer programs to automatically identify all the codes for VAD without any domain knowledge. We believe a domain ontology for medical device classification is very necessary for the purpose of our research.

The ontology can be created on the basis of the FDA Product Classification Database, with extended hierarchical information, and annotation properties.

File Selection for events The second challenge we encountered is to identify similar kinds of adverse events. The MAUDE database does not provide any searchable feature for adverse events. Currently one has to manually go through the narratives to identify a specific adverse event. A domain ontology would help automate this process. The Adverse Event Ontology (AEO) is a community-driven ontology developed to standardize and integrate data on biomedical adverse events and support computer-assisted reasoning [1]. The current version of AEO, however, focuses on vaccine and drug adverse events as the initial use cases, and only provides a shallow representation for medical device adverse events. It will be an interesting and important task to extend the AEO for representing medical device adverse events. In addition, since our focus is the time aspects in adverse events, we would also like to assert the temporal information in adverse events semantically when possible (e.g. to assert that a late stent thrombosis is a thrombosis that has a duration of 30 days or longer).

Semantic Annotation Since a lot of temporal information is embedded in clinical narratives, extracting information of interest and semantically annotating this information with respect to domain ontologies are a must. A manual annotation process could be very time-consuming and expensive [7]. The information to be annotated for our system includes: adverse events, other related medical events, possible device problems, as well as important temporal expressions and relations. We are trying to adapt NCBO annotator [5] and Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES) [6] for semi-automatic annotation in our system. The NCBO annotator is a user-friendly, scalable web service that is designed for annotating narratives with relevant ontology concepts. The cTAKES pipeline processes clinical notes, identifies important clinical named entities (NEs) as well as semantic features such as negations for the NEs, and assigns attributes for the code in standard ontologies such as SNOMED CT [9] and RxNorm [8] to these NEs. Most recently, a time relation recognizer is under development for cTAKES. Both systems, however, rely on well-developed domain ontologies as the background knowledge base for annotation. Currently, there is a lack of a domain ontology that can serve the purpose as a knowledge base for medical device adverse events. The FDA has developed a set of Adverse Event Problem Codes for standardizing the classification of device and patient problems associated with medical device use [10]. This coding system can provide a good foundation for developing an ontology of the domain. A few extensions, however, need to be done to serve the annotation purpose: (1) expanding hierarchical levels to cover granular representation of data, and (2) adding more lexical properties for each code, i.e, synonyms and acronyms.

Reasoning Framework We are working on building an ontology-based temporal relation reasoning framework for helping clinical researchers answer time-related questions from EHR. We have built an ontology called CNTRO (Clinical Narrative Temporal Relation Ontology) [12], which was designed to model the temporal information in clinical narratives. It models clinical events,

different kinds of temporal expressions (such as time instants, time intervals, repeated time periods, and durations), different levels of time granularity, temporal relations, and time uncertainties. Based on CNTRO, we have also developed a prototype framework for querying and inferring temporal information [11]. We evaluated the feasibility of using CNTRO with existing Semantic-Web technologies and discussed possible limitations and extensions that we found necessary or desirable to achieve the purposes of querying time-oriented data from real-world clinical narratives. Most recently, we have applied our system to the late stent thrombosis adverse events use case (with manually annotated data). Our preliminary study received very promising results (∼89% accuracy) on answering important time-related questions [2].

3 Concluding Remarks

In this paper, we discuss our vision of an ontology-based system for representing both events and temporal relations, annotating the useful information from narratives, and querying and reasoning over the data. Our preliminary study indicates that our system is feasible for retrieving from and reasoning over annotated data from medical device adverse event reports. Future work includes extending or creating domain ontologies for medical devices, adverse events associated with medical devices, as well as problems associated with medical devices.

References

1. Adverse Event Ontology. `http://sourceforge.net/projects/aeo/`.
2. K.K. Clark, C. Tao, D.K. Sharma, and C.G. Chute. Application of a temporal reasoning framework tool in analysis of medical device adverse events. In *AMIA Annual Symposium*, 2011. submitted.
3. FDA product code. `http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpcd/classification.cfm`.
4. MAUDE - Manufacturer and User Facility Device Experience Database -FDA. `http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm`.
5. NCBO Annotator. `http://www.bioontology.org/annotator-service`.
6. cTAKES. `http://www.ohnlp.org`.
7. Angus Roberts, Robert Gaizauskas, Mark Hepple, George Demetriou, Yikun Guo, Ian Roberts, and Andrea Setzer. Building a semantically annotated corpus of clinical texts. *J. of Biomedical Informatics*, 42:950–966, October 2009.
8. RxNorm. `http://www.nlm.nih.gov/research/umls/rxnorm/`.
9. Systematized nomenclature of medicine–clinical terms (SNOMED CT). `http://www.snomed.org`.
10. L. Reed TL T and D. Kaufman-Rivi. FDA adverse event problem codes: standardizing the classification of device and patient problems associated with medical device use. *Biomedical Instrumentation and Technology*, 44(3):248–256, 2010.
11. C. Tao, H.R. Sobrig, D.K. Sharma, W.-Q. Wei, G.K. Savova, and C.G. Chute. Time-oriented question answering from clinical narratives using Semantic-Web techniques. In *ISWC*, pages 241–256, 2010.
12. C. Tao, W.-Q. Wei, G.K. Savova, and C.G. Chute. A semantic web ontology for temporal relation inferencing in clinical narratives. In *AMIA Annual Symposium*, pages 787–791, 2010.