MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-202

On Building an Ontological Knowledge Base for Managing Patient Safety Events

Chen Liang, Yang Gong

School of Biomedical Informatics, University of Texas Health Science Center at Houston, TX, USA

Abstract

Over the past decade, improving healthcare quality and safety through patient safety event reporting systems has drawn much attention. Unfortunately, such systems are suffering from low data quality, inefficient data entry and ineffective information retrieval. For improving the systems, we develop a semantic web ontology based on the WHO International Classification for Patient Safety (ICPS) and AHRQ Common Formats for patient safety event reporting. The ontology holds potential in enhancing knowledge management and information retrieval, as well as providing flexible data entry and case analysis for both reporters and reviewers of patient safety events. In this paper, we detailed our efforts in data acquisition, transformation, implementation and initial evaluation of the ontology.

Keywords:

Medical error; Patient safety; Ontology; Knowledge management

Introduction

The increasing high rate of medical errors indicates that patient safety is a prominent issue [1, 2]. A recent study reported that preventable medical errors cause annual deaths of 210,000 to 440,000 in the United States [3]. The magnitude of these medical errors, as well as near misses and unsafe conditions, has raised public awareness on patient safety and interest in research. For the purpose of preventing error occurrence, much attention has been drawn into reasoning about systemic factors that contribute to the errors, while the road block appears to be the disclosure of patient safety events [1]. It is documented that the major obstacle to disclosing patient safety events and proposing systematic solutions is due to limited functionalities of a reporting system [4]. The core functionality of a reporting system is thought to be collecting, analyzing and learning from the existing mistakes [5-7]. Therefore, the systems should include functional modules such as data acquisition, knowledge management, information retrieval, and beyond.

Reporting systems are pervasively used in the United States [4, 8], Australia [9], United Kingdom [10] and other countries. Nevertheless, debates on the effectiveness of such systems still remain. So far , there has hardly been any study that reports a decrease of medical errors, mortality or morbidity directly due to the intervention of a reporting system [4]. Obviously, more efforts are needed to prove the value of the reporting systems. A first-line question is why the reporting systems are unsuccessful in demonstrating the value. Among various reasons, data quality is thought to be a major concern. Patient safety reports containing detailed narratives are helpful in replicating the incidents and further translating into improved patient safety. The narratives are different from other data types stored in generic health information systems, such as electronic medical record (EMR) data and clinical notes. Reporting systems collect a variety of data ranging from structured to unstructured formats and therefore may cause a data inconsistency issue, which leads to incompleteness and inaccuracy of the data entry [11]. For example, a reporting system with a structured data entry can oftentimes force the reporters to choose "other or miscellaneous" when they are asked to categorize a patient safety event [12]. In contrast, unstructured data (free text) are unconstrained in offering detailed information but they are not immune from issues. An obvious pitfall for the use of unstructured data in the reporting system is time efficiency. Many reporters are working under time pressure or in a multi-task mode and thus they may not have sufficient time to provide a complete and detailed report to the systems [11]. A recent study using a text prediction method intended to mitigate the abovementioned problems [13]. With this method, the system provides prompting information (suggested words/phrase to use) to the reporters at the time when they are typing free text into the data entry portal. The text prediction technique provides insights into extracting and organizing the semantic information based on the free text with the only limitation being that the prediction list was manually prepared by domain experts. Another pitfall for the use of unstructured data appears in data processing. Similar to clinical notes, typos, abbreviations, and nonstandard acronyms are typically intertwined with free text data. These have become barriers for data pre-processing, such as de-identification, and classification by natural language processing (NLP), and thus cost extra time for reviewers in understanding the data.

Taxonomies have been used to address these problems as they intend to manage patient safety events as a knowledge base. The taxonomies used to document and classify patient safety reports can be traced back to 1987 when the Australian Patient Safety Foundation (APSF) originally reported the Australian Incident Monitoring System [14]. Later on, a series of well known taxonomies were put into use, which include the JCAHO patient safety event taxonomy [15], the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP)'s taxonomy of medication errors [16], the Neonatal Intensive Care system (NIC) [17], the Pediatric Patient Safety taxonomy (PED) [18], the Preliminary Taxonomy of medical errors in Family Practice (PTFP) [19], the Taxonomy of Nursing Errors (TNE) [20], and the Adverse Event Reporting Ontology (AERO) [21]. While these taxonomies served primarily as domain specific knowledge bases, the rapid increase of patient safety data calls for a sharable knowledge base organized by a unified language system.

In sum, more efforts are needed for improving data quality, sharing and learning from patient safety events across the individual systems. A significant challenge to increase data quality remains in the development of a unified domain knowledge base and the effective use of the data. To explore solutions to the problems, we aim to build a semantic web ontology (Medeon) using W3C open standard Web Ontology Language (OWL). The Medeon serves as a unified knowledge base for organizing patient safety events, and as a supporting component that facilitates the user end applications towards enhancing data entry and data quality.

Materials and Methods

Semantic Web Ontology

A taxonomy as a controlled vocabulary in hierarchy has been used in patient safety reporting for years. Ontologies are explicit specifications of conceptualized definitions and relationships where these specifications define a taxonomy of the knowledge [22]. Specifically, an ontology models the real-world knowledge by encoding the entities and the relationships among the entities. We aimed to develop an ontology to replace the role of a taxonomy because the ontology can provide a broader application over taxonomies. We chose OWL and semantic web technologies because they jointly provide a unique advantage for machine understandable semantics and descriptive logic reasoning. OWL has an advantage that allows us to identify unique patient safety terminologies or concepts that may appear under different names or originate from different sources. This advantage largely reduces the ambiguity in medical terminologies, and may advance the knowledge management of patient safety events.

Previous work on building ontologies for patient safety events employed various methodologies (i.e., techniques, tools, procedures and guidelines) [23-25], yet these approaches are lacking computer understandable representations. In the present work, we refer to the *Semantic Web for the Working Ontologist* for theories and general guidelines commonly employed in developing OWL ontologies [26]. Protégé (V4.3.0) was employed to implement the ontology.

The Meta Ontology

Patient safety taxonomies have been used as a reference where hierarchies, entities, and relationships can be used as candidate materials [23]. While these taxonomies served primarily as standards of domain specific taxonomies, the rapid growth in medical information needs a knowledge base for sharing and learning across the reporting systems through a unified language serving as a common denominator. With this view, the Common Formats (v1.2) developed by the Agency for Healthcare and Research Quality (AHRQ) were employed as the taxonomy where we extracted and encoded semantic knowledge into Medeon. Recognized as a unified standard of reporting patient safety events, the Common Formats are designed to specify and collect event information, which range from general concerns to frequent occurrences and serious types of events. We borrowed the hierarchical structure in the Common Formats to build the OWL classes and rephrased the narrative data in the Common Formats to construct OWL incidents and objective properties. At the top level, four OWL classes consist of *Circumstances_of_Event, Patient_Information,Reporting_Reporter_and_Report_Information,* and *Type_of_Event*, with the maximum depth of four. Figure 1 shows the visualization of expanded classes. The OWL Object Property was defined as IsA property, for example, *Inattention IsA Human_factors*.

Data Acquisition

At an initial stage, all the entities and relationships implemented in Medeon were extracted from the Common Formats. Healthcare event reporting form (HERF), patient information form (PIF), and summary of initial report (SIR) in the Common Formats are regarded as a comprehensive and relatively complete collections of entities that can represent patient safety events. Therefore, a direct translation was performed in order to encode those entities from the Common Formats to Medeon. To obtain high quality data in the Common Formats, we followed a set of principles as guidelines [27]. Table 1 provides a brief description on the principles. We borrowed eight principles that were separated into three dimensions to guide the rephrasing of the language used in the ontology. Note that we did not include 'social quality' from the Dimensions in the original literature since this dimension measures the ontology comparing it to the existing ontologies and emphasizes the utility of the ontologies which are not applicable to the project. When the translation of entities was completed, we imported those entities into Medeon via Protégé. Data consistency was checked through Protégé build-in modules to ensure that no logical conflict existed in the ontology.

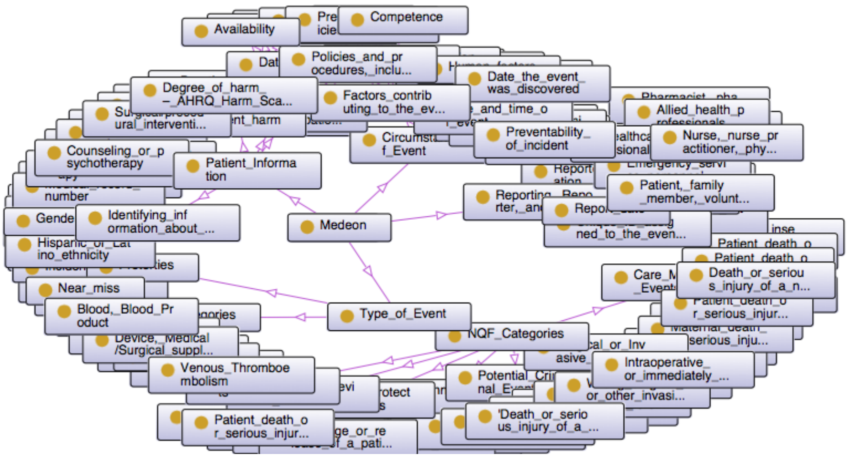

Figure 1 - OWL Class Visualization by OntoGraf (V1.0.1) in Protégé (V4.3.0).

Evaluation

The evaluation is intended to provide a comprehensive report pertaining to the effectiveness and validity of ontology in multi-dimensions. Our evaluation design included the assessment of the ontology itself and the user experience. We designed a questionnaire with a 5-point Likert scale to collect the measurable data from domain experts interested in using the ontology in their daily work.

Below we enumerate a set of questions in the questionnaire. Each question is equipped with answers to a 5-point Likert scale (i.e., 1=very disappointed; 2=disappointed; 3=neutral; 4=good; 5=very good).

1. The phrases used in the vocabulary are well formed and the words are well arranged.
2. The terms used in the vocabulary can explain the meanings of real-world concepts.
3. The terms that appear in the vocabulary are clear.
4. The vocabulary represents the designated domain and provides sufficient knowledge to the user.
5. The claims the vocabulary makes are reasonable.
6. The vocabulary can satisfy your requirements when you use it to categorize the case you are reviewing.
7. Please rate the overall satisfaction based on your experience using the vocabulary.

To make sure that the questionnaire reached the confidence level on effectiveness and validity, we employed a pre-measurement to assess the content-validity and inter-rater reliability to guide the final revision of the questionnaire. The content-validity measures to what extent the designed questions subjectively reflect the tasks they purport to measure. The inter-rater reliability measures the degree of agreement among raters. Three domain experts used the pre-measurement to validate the questionnaire where randomly selected patient safety reports were provided in the task.

The questions listed below were used for measuring content-validity. Each question was instructed to be answered on a 4-point scale (i.e., Not relevant; Somewhat relevant; Quite relevant; Highly relevant)

1. "The phrases used in the vocabulary are well-formed and the words are well-arranged."
   Does the scale purport to measure "The correctness of syntax."?
2. "The terms used in the vocabulary can explain the meanings of real-world concepts."
   Does the scale purport to measure "The meaningfulness of terms."?
3. "The terms that appear in the vocabulary are clear."
   Does the scale purport to measure "The clarity of terms."?
4. "The vocabulary represents the designated domain and provides sufficient knowledge to the user."
   Does the scale purport to measure "The comprehensiveness of the vocabulary in a certain domain."?
5. "The claims the vocabulary makes are reasonable."
   Does the scale purport to measure "The accuracy of information."?
6. "The vocabulary can satisfy your requirements when you use it to categorize the case you are reviewing."
   Does the scale purport to measure "Whether the vocabulary specifies agent's specific requirements."?
7. "Please rate the overall satisfaction based on your experience using the vocabulary."
8. Does the scale purport to measure "The overall satisfaction to the vocabulary."?

Results

Upon the completion of Medeon, we obtained a semantic web ontology in OWL format. The ontology represented, with necessary conceptualization and translation, the entities and relationships of patient safety knowledge that were extracted from the Common Formats. The ontology was constructed in a hierarchy with four top-level classes where each contained sub-classes with a maximum depth of four levels. An example of OWL individuals is shown in Figure 2. As in the preliminary stage, these individuals may be incomplete yet they represent the most frequently used concepts and terminologies appearing in the Common Formats. Note that OWL classes and individuals, as well as the OWL properties, are open to expand. That being said, the knowledge we borrowed from the Common Formats serve as building blocks for further development without limiting patient safety ontology.

Two domain experts participated in the pre-measurement. The results showed a 100% agreement for the inter-rater reliability and 100% for content validity.

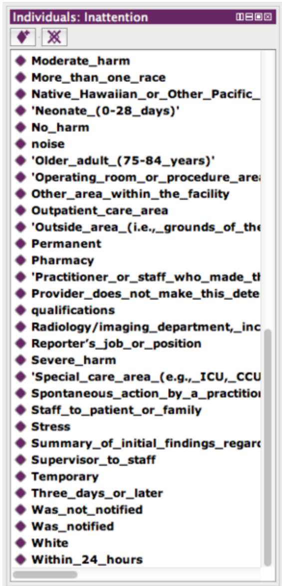

Figure 2 - OWL Individuals in Protégé (V 4.3.0).

Discussion and Future Work

The ontology primarily serves as a knowledge base to model the taxonomies broadly used for patient safety reports. With this role, the proposed ontological approach aims to meet the challenges in the development of reporting systems. Among the many factors fundamental for a successful reporting system, data quality has been a major concern. An outstanding reporting system should be able to collect quality data that link to the procedures and factors threatening patient safety in a timely manner. Nevertheless, a great number of reporting systems are suffering from low quality data due to inefficiency and ineffectiveness of data entry [11, 28, 29]. To improve the data quality, much effort has been made in increasing the number of reports, but the increase in quantity does not improve the system performance since the very crux of the problem remains in knowledge management. In fact, the reporting systems in use generate a great volume of patient safety reports, which on the contrary are becoming a burden for data processing.

The ontology is what we believe is a suitable approach in patient safety reporting. Given a patient safety report, oftentimes it needs to be labeled with multiple categories in a hierarchical knowledge base. Neither plain reporting forms nor patient safety taxonomies can easily solve this problem during the submission of a report to the system or retrieving a report from the system. For example, a case being labeled under 'lighting' may be also labeled under 'contributing factor', 'environment', and 'patient fall', assuming that 'contributing factor' and 'environment' are the super classes of 'lighting', while 'patient fall' is under the other super class yet has certain associations to 'lighting'. It could be time consuming when reporting such a 'lighting' case in plain reporting forms and it could be difficult to extract the association between 'lighting' and 'patient fall', since they belong to different classes or different hierarchical levels. Fortunately, an ontology can contribute to these aspects since it encodes all the relationships among entities and is able to query via path expression.

Ontology provides insights into an efficient and user-friendly data entry framework. Researchers have found close associations between data entry and the system performance in terms of the completeness and accuracy of patient safety reports [30, 31]. The majority of patient safety reports are recorded in free text. Although the free text might be an efficient and natural means for users to deliver an informative case, it could be costly to turn the raw information into a cognitively organized and manageable format for professionals to utilize . Applying semantic web ontology to organize the text data can support effective data entry. Such a use case includes web portals where ontology is used for defining terminologies and concepts (meanings) for an area of knowledge. For example, 'patient fall', 'female', 'slip', and 'emergency room' are terminologies that can be found in the patient safety ontology. The ontology can also define a concept like 'All female slips in emergency room are considered as patient falls.' This provides an interesting angle to look into; the latent yet important information in the patient safety reports since a few distributed semantics (terminologies and relationships) appear to be sufficient for representing the most significant meanings in a report.

To our best knowledge, the use of AHRQ Common Formats as the source taxonomy to build semantic web ontology is an initiative. Patient safety reporting systems never lack taxonomies, but a universal and computer understandable knowledge base. The Common Formats provide a compatible format that includes patient safety events ranging from a series of general concerns, frequent occurrences and serious types of events andmore importantly, the implementation of the Common Formats in an ontological framework offers an open environment to aggregate and share the patient safety knowledge base by cooperating with other data sources and ontologies.

Our design and implementation have challenges and limitations. When building the ontology, it is most challenging in mapping between discrepant data sources due to the distinction among existing taxonomies in terms of the hierarchical structures and synonymous terminologies. We envision the use of NLP techniques and automatic classifier (i.e., k nearest neighbor) could facilitate the process, as we will expand Medeon in the next step. Also, a view on a unified coding system, such as Unified Medical Language System (UMLS), is definitely helpful. On the other hand, debates regarding validity and effectiveness are always in ontological studies Therefore, we must continue the evaluation study of the present ontology.

Table 1 - Design Principles for High Quality Ontologies.

Dimensions	Attributes	Description
Syntactic quality	Lawfulness	Correctness of syntax
	Richness	Breadth of syntax used
Semantic quality	Interpretability	Meaningfulness of terms
	Consistency	Consistency of meaning of terms
	Clarity	Average number of word senses
Pragmatic quality	Comprehensiveness	Number of classes and properties
	Accuracy	Accuracy of information
	Relevance	Relevance of information for a task

Conclusion

The development of a knowledge base for patient safety reporting systems is imperative for both practice and research. With the aim of establishing a comprehensive knowledge base, we employed a semantic web ontology that plays a key role underlying the reporting systems. The present ontology built on the Common Formats serves as the building blocks towards a unified knowledge base, with which the reporting systems are expected to support comprehensive data entry and increase the data quality. We envision that utilizing a semantic web ontology would facilitate information retrieval and reuse of the narrative data for expert review, clinical decision-making and education. Moving forward, the swift growth in aggregate data requires a sustainable knowledge base to keep abreast with the latest reporting events. Our design must be open-minded to glean knowledge from the most recently reported data and tremendous amount of historical data.

Acknowledgments

This project is supported by a grant from AHRQ, 1R01HS022895 and a patient safety grant from the University of Texas system. We thank Dr. Jing Wang for her invaluable comments in designing the evaluation questionnaires for the ontology.

References

[1] Kalra J, Kalra N, Baniak N. Medical error, disclosure and patient safety: A global view of quality care. Clinical biochemistry. 2013;46(13):1161-9.

[2] Kalra J. Medical errors: an introduction to concepts. Clinical biochemistry. 2004;37(12):1043-51.

[3] Joshi MS, Anderson JF, Marwaha S. A systems approach to improving error reporting. Journal of healthcare information management: JHIM. 2001;16(1):40-5.

[4] Woodward HI, Mytton OT, Lemer C, Yardley IE, Ellis BM, Rutter PD, et al. What have we learned about interventions to reduce medical errors? Annual review of public health. 2010;31:479-97.

[5] Leape LL, Abookire S, World Health O. WHO draft guidelines for adverse event reporting and learning systems: from information to action: World Health Organization; 2005.

[6] Olsen S, Neale G, Schwab K, Psaila B, Patel T, Chapman EJ, et al. Hospital staff should use more than one method to detect adverse events and potential adverse events: incident reporting, pharmacist surveillance and local real-time record review may all have a place. Quality and Safety in Health Care. 2007;16(1):40-4.

[7] Weissman JS, Schneider EC, Weingart SN, Epstein AM, David-Kasdan J, Feibelmann S, et al. Comparing patient-reported hospital adverse events with medical record review: do patients know something that hospitals do not? Annals of Internal Medicine. 2008;149(2):100-8.

[8] Kohn LT, Corrigan JM, Donaldson MS. To Err Is Human: Building a Safer Health System: National Academies Press; 2000.

[9] Runciman WB. Lessons from the Australian Patient Safety Foundation: setting up a national patient safety surveillance system—is this the right model? Quality and Safety in Health Care. 2002;11(3):246-51.

[10] Ferlie EB, Shortell SM. Improving the quality of health care in the United Kingdom and the United States: a framework for change. Milbank Quarterly. 2001;79(2):281-315.

[11] Gong Y. Data consistency in a voluntary medical incident reporting system. Journal of medical systems. 2011;35(4):609-15.

[12] Gong Y, Richardson J, Zhijian L, Alafaireet P, Yoo I, editors. Analyzing voluntary medical incident reports. AMIA Annual Symposium proceedings/AMIA Symposium AMIA Symposium; 2007.

[13] Hua L, Wang S, Gong Y. Text Prediction on Structured Data Entry in Healthcare. Methods Inf Med. 2008;47(1):8-13.

[14] Spigelman AD, Swan J. Review of the Australian incident monitoring system. ANZ journal of surgery. 2005;75(8):657-61.

[15] Chang A, Schyve PM, Croteau RJ, O'Leary DS, Loeb JM. The JCAHO patient safety event taxonomy: a standardized terminology and classification schema for near misses and adverse events. International Journal for Quality in Health Care. 2005;17(2):95-105.

[16] Zhang J, Patel VL, Johnson TR, Shortliffe EH. A cognitive taxonomy of medical errors. Journal of biomedical informatics. 2004;37(3):193-204.

[17] Suresh G, Horbar JD, Plsek P, Gray J, Edwards WH, Shiono PH, et al. Voluntary anonymous reporting of medical errors for neonatal intensive care. Pediatrics. 2004;113(6):1609-18.

[18] Woods DM, Johnson J, Holl JL, Mehra M, Thomas EJ, Ogata ES, et al. Anatomy of a patient safety event: a pediatric patient safety taxonomy. Quality and Safety in Health Care. 2005;14(6):422-7.

[19] Dovey SM, Meyers DS, Phillips RL, Green LA, Fryer GE, Galliher JM, et al. A preliminary taxonomy of medical errors in family practice. Quality and Safety in Health Care. 2002;11(3):233-8.

[20] Woods A, Doan-Johnson S. Toward a Taxonomy of Nursing Practice Errors. Nursing Management. 2002;33(10):45-8.

[21] Courtot M, Brinkman RR, Ruttenberg A. The Logic of Surveillance Guidelines: An Analysis of Vaccine Adverse Event Reports from an Ontological Perspective. PloS one. 2014;9(3):e92632.

[22] O'Leary DE. Using AI in knowledge management: Knowledge bases and ontologies. IEEE Intelligent systems. 1998;13(3):34-9.

[23] Mokkarala P, Brixey J, Johnson TR, Patel VL, Zhang J, Turley JP. Development of comprehensive medical error ontology. AHRQ: Advances in Patient Safety: New Directions and Alternative Approaches invited paper, under review. 2008.

[24] Gong Y, Zhu M, Li J, Turley J, Zhang J, editors. Clinical communication ontology for medical errors. AMIA Annual Symposium Proceedings; 2006: American Medical Informatics Association.

[25] Stetson PD, McKnight LK, Bakken S, Curran C, Kubose TT, Cimino JJ, editors. Development of an ontology to model medical errors, information needs, and the clinical communication space. Proceedings of the AMIA Symposium; 2001: American Medical Informatics Association.

[26] Allemang D, Hendler J. Semantic web for the working ontologist: effective modeling in RDFS and OWL: Elsevier; 2011.

[27] Burton-Jones A, Storey VC, Sugumaran V, Ahluwalia P. A semiotic metrics suite for assessing the quality of ontologies. Data & Knowledge Engineering. 2005;55(1):84-102.

[28] Pronovost PJ, Morlock LL, Sexton JB, Miller MR, Holzmueller CG, Thompson DA, et al. Improving the value of patient safety reporting systems. Advances in patient safety: new directions and alternative approaches. 2008;1.

[29] Gong Y, Terminology in a voluntary medical incident reporting system: a human-centered perspective. Proceedings of the 1st ACM International Health Informatics Symposium; 2010: ACM.

[30] Johnson CW. Failure in safety-critical systems: a handbook of accident and incident reporting. University of Glasgow Press: Glasgow, Scotland; 2003.

[31] Barach P, Small SD. Reporting and preventing medical mishaps: lessons from non-medical near miss reporting systems. Bmj. 2000;320(7237):759-63.

Address for correspondence

Yang Gong, MD, PhD

7000 Fannin St, Suite 600

Houston, TX, 77030, USA

Email: Yang.Gong@uth.tmc.edu