SUBMISSION GUIDELINES

AIMS & SCOPE

JDIS devotes itself to the study and application of theories, methods, techniques, services, infrastructural facilities using big data to support knowledge discovery for decision & policy making. The basic emphasis is big data-based, analytics centered, knowledge discovery driven, and decision making supporting. The special effort is on the knowledge discovery to detect and predict structures, trends, behaviors, relations, evolutions and disruptions in research, innovation, business, politics, security, media and communications, and social development, where the big data may include metadata or full content data, text or non-textual data, structured or non-structured data, domain specific or cross-domain data, dynamic or interactive data.

MANUSCRIPTS REQUIREMENTS

All papers are submitted in English with a double-line space. For the assurance that all the materials of the to-be submitted are included, please check the following:

Title page. The title, author list, affiliations, and author contribution statement should all be included on a title page as the first page of the manuscript file.

Structured abstract & Keywords. A structured abstract is required for all research papers in order for readers to acquire the key points of a manuscript.

References. Be sure all the references used should be cited properly in both in text and in bibliography. For the detailed information, please request a copy of Reference Citation Format.

COPYRIGHT

All submitted papers only normally should not have been previously published nor be currently under consideration for publication elsewhere. For all the materials translated or obtained from other published resources, they should be properly acknowledged. All copyright problems should be cleared without any legal entanglements prior to the publication.

NOTES FOR INTENDING SUBMISSIONS

A guide for authors and other relevant information, including submitting papers online, is available at http://www.jdis.org. For any questions, you can e-mail the Office at jdis@mail.las.ac.cn.

The JDIS Editorial Office
National Science Library, Chinese Academy of Sciences
No 33, Beishuan Xilu, Zhongguancun, Haidian District, Beijing 100190, P.R. China
Website: http://www.jdis.org

REFEREING PROCESS

Articles covering the topics or themes mentioned above will be refereed through a double-blind peer review process.

MANUSCRIPTS CATEGORIES

Research Articles represent original research work or a comprehensive and in-depth analysis of a topic. More than 3,000 words are considered as a proper length for such manuscripts, with a structured abstract ca. 200 words.

Application Articles cover the latest development and application in any segment of data analysis and information science field work. The length of the manuscript is preferred to be more than 3,000 words, with a structured abstract ca 200 words.

Review Articles summarize the status of knowledge and outline future directions of research within the scope of the journal.

Perspectives are forward-looking viewpoints that advocate important future directions in the theory or application of data and information science.
Knowledge Representation in Patient Safety Reporting: An Ontological Approach

Liang Chen & Yang Gong†
School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston 77030, USA

Abstract

Purpose: The current development of patient safety reporting systems is criticized for loss of information and low data quality due to the lack of a uniformed domain knowledge base and text processing functionality. To improve patient safety reporting, the present paper suggests an ontological representation of patient safety knowledge.

Design/methodology/approach: We propose a framework for constructing an ontological knowledge base of patient safety. The present paper describes our design, implementation, and evaluation of the ontology at its initial stage.

Findings: We describe the design and initial outcomes of the ontology implementation. The evaluation results demonstrate the clinical validity of the ontology by a self-developed survey measurement.

Research limitations: The proposed ontology was developed and evaluated using a small number of information sources. Presently, US data are used, but they are not essential for the ultimate structure of the ontology.

Practical implications: The goal of improving patient safety can be aided through investigating patient safety reports and providing actionable knowledge to clinical practitioners. As such, constructing a domain specific ontology for patient safety reports serves as a cornerstone in information collection and text mining methods.

Originality/value: The use of ontologies provides abstracted representation of semantic information and enables a wealth of applications in a reporting system. Therefore, constructing such a knowledge base is recognized as a high priority in health care.

Keywords Patient safety; Medical error; Knowledge representation; Health information technology; Ontology

† Corresponding author: Yang Gong (E-mail: Yang.Gong@uth.tmc.edu).
1 Introduction

Medical errors, near misses, and unsafe conditions cause patient harms and reduced healthcare quality. A recent study reported that the estimated annual cost of medical errors in the United States has risen to $17.1 billion (van Den Bos et al., 2011). The growing cost of medical errors is observed in other countries as well and has become a global patient safety concern (Baker et al., 2004; Vanderheyden et al., 2004; Williams & Osborn, 2006). The Institute of Medicine (IOM) and the Agency for Healthcare Research and Quality (AHRQ) recommended the use of patient safety reporting systems (PSRS) to reduce future mistakes from the incurred incidents (Brennan et al., 1991; Erickson et al., 2003; Kohn, Corrigan, & Donaldson, 2000). Moving from paper-based reporting systems to electronic systems, the development of PSRS has been documented since the late 1970s (Elliott, Martin, & Neville, 2014). A well-functioning PSRS benefits the communication efficiency (Cochrane et al., 2009; Elliott et al., 2014), the quality improvement of reports across various healthcare settings and types of errors (Braithwaite, Westbrook, & Travaglia, 2008; Cochrane et al., 2009; Kuo et al., 2012; Levitzion-Korach et al., 2009), and user experience (Braithwaite et al., 2010; Braithwaite, Westbrook, & Travaglia, 2008; Cochrane et al., 2009; Frankel, Gandhi, & Bates, 2003; Keistinen & Kinnunen, 2007; Levitzion-Korach et al., 2009; Mekhjian et al., 2004; Tepfers, Louie, & Drouillard, 2006; Tuttle et al., 2004). Despite diligent efforts and impressive progress in PSRS, ongoing challenges remain: (1) Low quality of data. Many efforts were made toward increasing the quantity of reports, yet data quality remains a major concern. A detailed discussion centers on the dilemma of using structured or unstructured data formats in the reporting (Gong, 2011; Hua, Wang, & Gong, 2014). (2) Challenge of processing text data. Most of the patient safety reports that convey information for analyzing are written in natural language (Lamont et al., 2009; Newman, 2003; Steiner, 2005). However, it has been a technical challenge for analyzing text data in a timely manner (Hsieh & Shannon, 2005; Pope, Ziebland, & Mays, 2000). (3) Lack of a common language system. In the biomedical domain, a controlled vocabulary of terms and concepts can enhance the interoperability of semantic data (Bodenreider, 2004). (4) Difficulty of classification. Classifying patient safety reports is recognized as a cornerstone of reporting and data analysis (Leape & Abookire, 2005). However, developing a mature strategy of classifying patient safety reports remains remarkably challenging (Erickson et al., 2003).

A research agenda to address these problems should include building a patient safety ontology, where a uniform knowledge base for representing patient safety knowledge is in the center of the discussion (Chang et al., 2005). Healthcare
institutes worldwide have been developing such a knowledge base, such as a
taxonomy for classifying and monitoring medical incidents released by Australian
Patient Safety Foundation (APSF) (Brixey, Johnson, & Zhang, 2002; Chang et al.,
2005; Dovey et al., 2002; Greens, 2006; Spigelman & Swan, 2005; Suresh et al.,
2004; Woods & Doan-Johnson, 2002; Woods et al., 2005; Zhang et al., 2004).
Nevertheless, ontology is recognized as an advanced solution for providing machine-
readable representations for semantic information (Allemang & Hendler, 2011;
Ananiadou & McNaught, 2006; Maynard, Li, & Peters, 2008; McGuinness et al.,
2004). Ontologies have several advantages. Firstly, serving as a tool of terminology
management, ontologies provide a clear representation and communication of
complex semantic relationships. Secondly, they support information exchange
among biomedical information systems, especially when the biomedical information
is growing rapidly (Alexander, 2006; Kumar, Yip, Smith, & Grenon, n.d.). Thirdly,
ontologies facilitate knowledge discovery and reuse (Andronis et al., 2011;
Bodenreider, 2008; Gottgtroy, Kasabov, & MacDonell, 2004; Mukherjea, 2005;
Smith et al., 2007). Biomedical knowledge is complex in content and huge in
amount but arduous to process. Ontologies form a number of standards of annotating
concepts and relations and thus make semantic reasoning available.

In this paper, we described our initial efforts to design and implement a patient
safety ontology for US hospitals in the context of PSRS. We used semantic
information from PSRS in US hospitals to generate the ontology. We further
discussed the application of the ontology in PSRS. The World Health Organization
(WHO) has reported initial efforts to achieve better integration and interoperability
of patient safety information in their patient safety program (Larizgoitia, Bouesseau,
& Kelley, 2013; Runciman et al., 2009; Sherman et al., 2009). Some other studies
followed up with a focus of ontological approaches (Rodrigues et al., 2007;
Souvignet et al., 2011; Souvignet & Rodrigues, 2014). Our efforts of constructing
a patient safety ontology fit in the context of patient safety reporting in the US.

2 Design

The development of the ontology follows OBO Foundry principles to incorporate
interoperable and accurate representations from the clinical reality (Smith et al.,
2007). The ontology construction began with designing a concept ontology to
determine the overall structure. An evaluation was conducted in order to validate
this structure. Accordingly, we incorporated annotated terms from real-world patient
safety reports into the concept ontology. Table 1 demonstrates the general workflow.
Table 1. Workflow chart of ontology construction.

<table>
<thead>
<tr>
<th>Project</th>
<th>Task</th>
<th>Materials</th>
<th>Method/tool</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept ontology</td>
<td>Knowledge acquisition</td>
<td>ICPS and the Common Formats</td>
<td>Expert analysis</td>
<td>Semantic knowledge organized in hierarchies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Semantic knowledge</td>
<td>Expert review</td>
<td>A concept ontology with a hierarchical structure of patient safety knowledge</td>
</tr>
<tr>
<td></td>
<td></td>
<td>organized in hierarchies</td>
<td>Ontology engineering</td>
<td>Quality indicators of the classification by domain experts</td>
</tr>
<tr>
<td>Evaluation</td>
<td>Human evaluation</td>
<td>Hierarchical classes from the concept ontology</td>
<td>Survey instrument Statistics</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Real-world reports from Web M&amp;M</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Computational evaluation</td>
<td>Concept ontology in OWL</td>
<td>Statistical analysis Consistency checking</td>
<td>Quantitative indicators of the ontology</td>
</tr>
<tr>
<td>Detailed ontology</td>
<td>Annotation</td>
<td>Concept ontology Dataset from a university hospital</td>
<td>Expert annotation</td>
<td>A detailed ontology with enriched terms, relations, and other ontological specifications</td>
</tr>
</tbody>
</table>

### 2.1 Developing a Concept Ontology

The concept ontology describes the most general concepts and categories across specific domains in patient safety reports. It also serves as a guideline for semantic annotation and integration in the later processes of constructing a detailed ontology, which includes instances of the concepts and other ontological specifications.

#### 2.1.1 Knowledge Acquisition

Semantic patient safety knowledge from the real world is the basic element for constructing a patient safety ontology. We extracted patient safety knowledge from the International Classification for Patient Safety (ICPS) and the Common Definitions and Reporting Formats (a.k.a., the Common Formats) by a method to incorporate the respective advantageous features of the two. ICPS is a conceptual framework developed by the WHO in 2009, representing concepts and preferred terminologies used in patient safety reports (Sherman et al., 2009). The Common Formats, developed by AHRQ, are a set of guidelines and paper-based formats for specifying and collecting safety event information in the US, which range from general concerns to frequently occurred and/or serious adverse event.

#### 2.1.2 Ontology Implementation

A formal language is used to standardize and normalize the expression of objects and their relations, in addition to computerized processing which can be done with XML (Rossi, Consorti, & Galeazzi, 1998). We used Web Ontology Language (OWL) as it represents rich and complex semantic information (Baader, 2003;
McGuinness et al., 2004). The data were implemented in Protégé 4.3.0. We employed an iterative process to construct the ontology, which is described in the following three steps:

1. **Data transformation.** A data transformation was employed to integrate the concepts and terms in ICPS and the Common Formats, where inevitable ambiguities and synonyms exist. Three domain experts (CL, XW, and KA), who have background knowledge in both patient safety and ontology engineering, performed the transformation by reviewing the concepts and terms in ICPS and the Common Formats. A final decision was made only if an agreement was reached among the three experts.

2. **Adjustment of hierarchical structure.** In many cases, a unique concept may be categorized in different classes or even shown under different names. Since the ICPS has been recognized as an adequate classification for representing patient safety knowledge hierarchy (Sherman et al., 2009; Souvignet et al., 2011), we adopted ICPS’s hierarchical structure and made minor adjustments with exceptions when a creation of new classes was necessary. Such adjustments include merging duplicate subcategories, concepts, and terms. Parent-child relations were defined by taxonomic subsumption, ‘isA’ (e.g., A is a subclass of B). Alias relations were defined by ‘EquivalentTo’ (e.g., A is equivalent to B) (Allemang & Hendler, 2011). We also defined other relations such as ‘hasParticipant’, ‘hasOutcome’, ‘involvesActivity’, etc.

3. **Merging the Common Formats with ICPS.** We built the ontology in Protégé to merge structures, concepts, and terms from the Common Formats and ICPS with adequate properties created.

### 2.2 Evaluation

The evaluation examines whether the concept ontology represents an adequate knowledge for patient safety reports. The ontology first needed to pass the machine-based evaluation, by which the Protégé build-in module (HermiT 1.3.8) performed consistency checking (Shearer, Motik, & Horrocks, 2008). Secondly, we employed human evaluation by using survey instruments and statistical analysis. The human evaluation procedure has two phases. In the first phase, we developed a survey instrument for assessing biomedical ontologies in the scope of patient safety events. The questions in the survey instrument were adapted to cover eight dimensions for evaluating an ontology (Brank, Grobelnik, & Mladenić, 2005; Burton-Jones et al., 2005). To ensure that the survey instrument reaches a sufficient confidence level of reliability and validity for use, we employed a pre-assessment to measure its content-validity and inter-rater reliability. The content validity measures to what extent the
designed questions subjectively reflect the tasks they purpose to measure (Lynn, 1986; Polit & Beck, 2006). The inter-rater reliability measures the degree of agreement among raters (Fleiss, Levin, & Paik, 1981). The survey instrument is valid for use only if no major revision is needed. In the second phase, two domain experts (JW and YG) who are experienced in reviewing patient safety reports used the survey instrument to assess the concept ontology. When taking the survey, they were asked to annotate, using the concept ontology, the de-identified patient safety reports from Morbidity and Mortality Rounds on the Web (WebM&M). WebM&M is an online platform that publishes reported patient safety events and expert commentaries (Wachter et al., 2005). Table 2 demonstrates the sample questions in the pre-assessment and the survey instrument.

Table 2. A sample set of questions demonstrates the design of the survey instrument and the pre-assessment for validating the survey instrument.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Questions in the survey instrument</th>
<th>Questions in the pre-assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctness</td>
<td>For the case you reviewed, the terms used in the taxonomy are well-formed and the words are well-arranged.</td>
<td>Does the scale purport to measure “The correctness of syntax”?</td>
</tr>
<tr>
<td>Meaningfulness</td>
<td>For the case you reviewed, the terms used in the taxonomy can represent the concepts in the real-world setting.</td>
<td>Does the scale purport to measure “The meaningfulness of terms”?</td>
</tr>
<tr>
<td>Clarity</td>
<td>For the case you reviewed, the terms that appear in the taxonomy are clear (no ambiguity).</td>
<td>Does the scale purport to measure “The clarity of terms”?</td>
</tr>
<tr>
<td>Comprehensiveness</td>
<td>For the case you reviewed, the taxonomy provides sufficient knowledge in the domain.</td>
<td>Does the scale purport to measure “The comprehensiveness of the taxonomy in a certain domain”?</td>
</tr>
<tr>
<td>Accuracy</td>
<td>The information the taxonomy provides is accurate.</td>
<td>Does the scale purport to measure “The accuracy of information”?</td>
</tr>
<tr>
<td>Specificity</td>
<td>The taxonomy satisfies your needs when you use it to categorize the case you are reviewing.</td>
<td>Does the scale purport to measure “Whether the taxonomy specifies agent’s specific requirements”?</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>Please rate the overall satisfaction based on your experience of using the taxonomy.</td>
<td>Does the scale purport to measure “The overall satisfaction to the taxonomy”?</td>
</tr>
<tr>
<td>Educational value</td>
<td>Please rate the education value of the case you reviewed.</td>
<td>Does the scale purport to measure “The educational value of the case”?</td>
</tr>
</tbody>
</table>

2.3 Developing a Detailed Ontology

A successful concept ontology provides an intrinsic infrastructure of patient safety knowledge, thus paves the way for developing a complete ontology. We performed a set of tasks to populate selected ontology classes with instances from real-world patient safety reports. These reports ($n = 2,919$) were obtained from a University Hospital in the US. The resulting ontology includes instances for classes
that associate with two types of patient safety incidents: ‘patient fall’ \( (n = 346) \) and ‘equipment and device’ \( (n = 170) \). We focused on these two types of incidents in the starting stage for two reasons. Firstly, patient fall usually leads to significant morbidity and mortality in US hospitals. Secondly, during our review of the reports, information describing patient fall and equipment and devices is well documented in narratives and thus can be easily modeled by an ontological representation. For example, in a segment of the reports, ‘Pt was noted to be sitting on the side of the bed as he had done many times before without any difficulty or c/o. Pt was found on the floor next to his bed on his back and yelling for help.’, a subject-predicate-object triple can be determined as ‘Pt-sit-bed’. The populating process involves the extraction of terms from the reports into corresponding classes in the ontology. Two domain experts (SP and QM) completed the ontology population by following these procedures: (1) Each expert is assigned to a set of randomly selected reports and the selected classes of ‘patient fall’ and ‘equipment and devices’; (2) Each expert reviews the reports and annotates terms from the text to corresponding classes; (3) Each expert cross validates each other one’s work; (4) The populating is considered complete after a few iterations when no more revision remains needed.

### 3 Results

#### 3.1 The Ontology

**3.1.1 Ontology Structure**

With minor adjustments, we retained to the largest extent the top-level classes in the ICPS, which are incident type, patient characteristics, incident characteristics, detection, mitigating factors, patient outcomes, organizational outcomes, ameliorating actions, actions taken to reduce risk, and contributing factors/hazards. ‘Process’, which used to be under ‘Incident type’ – ‘Clinical administration’, was defined as a top-level class since it does not fit in any place under any top-level classes. A number of classes were broken down into several newly defined subclasses to better fit in the ontology. For example, ‘Detection’ was replaced by several new classes (i.e. ‘People’, ‘Assessment’, etc.) to accurately describe how the incident was detected. Some other changes worth mentioning are the relocation of ‘Fall’, ‘Pressure ulcer’, and ‘Venous thromboembolism’ since they were not explicitly documented in the ICPS but are significant in clinical cases. Adjustments were also made to the classes extracted from the Common Formats. For example, ‘Surgery’ and ‘Anesthesia’ are defined as top-level classes in the Common Formats. However, they were defined as subclasses of the ‘Process’ in our ontology. Adjustments as such help retain both the original information and the clarity of the ontological
structure. Figure 1 provides a close view of these adjustments by showing the ontology structure in Protégé screenshots.

Figure 1. Protégé screenshots of partial ontology hierarchies. (a) Overall ontology structure. (b) Ontology structure of the classes associated with ‘fall’ incidents. (c) Ontology structure of the classes associated with ‘equipment and device’ incidents.

The current version of ontology has 71 classes, in which 24 classes have equivalent classes from selected existing ontologies from BioPortal. All these ontologies are in the fields of medical incidents or patient safety. In these ontologies, the ICPS ontology is derived from WHO’s conceptual model of ICPS. The Adverse Event Ontology (AEO) encodes terminologies and representations in the scope of adverse events and medical interventions (He et al., 2011). The use of existing ontological terms can reduce repetitive work on future ontology expansion within similar domains. Table 3 shows a summary of the ontological terms.

<table>
<thead>
<tr>
<th>Ontology names</th>
<th>Classes</th>
<th>Object properties</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Safety Ontology</td>
<td>47</td>
<td>3</td>
<td>50</td>
</tr>
<tr>
<td>International Classification for Patient Safety (ICPS)</td>
<td>22</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>Adverse Event Ontology (AEO)</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>5</td>
<td>76</td>
</tr>
</tbody>
</table>

Table 3. Statistics of ontology specific terms and imported terms.
3.1.2 Examples of Ontology Terms

A simple example of medical incidents can be determined by linking a number of terms through object properties. Figure 2 demonstrates two examples in which a ‘patient fall’ incident can be determined. In the examples, a ‘fall’ incident can be inferred by defining semantic rules, in which classes (i.e. Person, Patient, Activity, PatientActivity, PatientOutcome, IncidentType, and Fall), object properties (i.e. involvesActivity, hasParticipant, and hasOutcome), and other predefined properties in the ontology were employed in the reasoning process. By defining more terms, object properties, and rules, we can infer a greater number of semantic evidence.

Figure 2. An example of inferred terms. (a) Two rules that infer a ‘patient fall’ incident. (b) Three inferences suggested by HermiT 1.3.8 in Protégé. ‘Patientfall’ is inferred as an instance of ‘Fall’. (c) A diagram of designed logical path applied in the example.
3.2 Evaluation Results

The ontology passed consistency checking through HermiT 1.3.8 (Shearer et al., 2008). This procedure validated the ontology from a machine-based perspective. In addition, we included human-centered evaluation to ensure the ontology is valid in clinical practice. We used a real-world patient safety report (http://www.webmm.ahrq.gov/case.aspx?caseID=337) in the Morbidity and Mortality Rounds on the Web (WebM&M) for the pre-assessment. Two domain experts (JW and YG) participated in the pre-assessment. See Table 4 for the results. We used a Content Validity Index (CVI) method to calculate the optimized content validity (Polit & Beck, 2006). The CVI for each item and overall are shown in Table 5.

Table 4. Calculation of inter-rater reliability for the evaluation instrument.

<table>
<thead>
<tr>
<th></th>
<th>Item 1</th>
<th>Item 2</th>
<th>Item 3</th>
<th>Item 4</th>
<th>Item 5</th>
<th>Item 6</th>
<th>Item 7</th>
<th>Item 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater 1 (WJ)</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Rater 2 (YG)</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Number in agreement</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total agreement in percentile</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. The eight items are shown in the Table 2, ‘corresponding questions in the pre-assessment’ column. The numbers represent 5-point scale, i.e., 1 = strongly disagree; 2 = disagree; 3 = neither agree nor disagree; 4 = agree; 5 = strongly agree. We count it an agreement when two raters select the same scale or neighbor scales for a given item.

Table 5. Two raters rating on a 4-point scale for content validity.

<table>
<thead>
<tr>
<th></th>
<th>Item 1</th>
<th>Item 2</th>
<th>Item 3</th>
<th>Item 4</th>
<th>Item 5</th>
<th>Item 6</th>
<th>Item 7</th>
<th>Item 8</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater 1 (WJ)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>1.00</td>
</tr>
<tr>
<td>Rater 2 (YG)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>1.00</td>
</tr>
<tr>
<td>Number in agreement</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>Mean I-CVI = 1.00</td>
</tr>
<tr>
<td>Item CVI</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>Mean rater proportion = 1.00</td>
</tr>
</tbody>
</table>

Note. The Content Validity Index (CVI) is calculated as the number of all raters selecting a scale of either 3 or 4, where 1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, 4 = highly relevant. An X stands for a CVI counted. I-CVI stands for the CVI for individual item.

4 Discussion

Ontologies are important tools to structure biomedical domains (Bodenreider, 2008). In the last decade, we have seen a grand challenge for translational research in biomedical domains with increase in both volume and complexity of data. Interpreting these data naturally requires domain knowledge that is usually given by clinical experts. When it comes to a timely response to rapidly growing medical incident data, a machine-readable fashion for such domain knowledge is integral. The broad use of biomedical ontologies has resulted in a community of resources
that can be shared within the domain. Ontology communities as such enable easy data integration and incorporation of individual ontologies with specific text mining applications.

We highlighted a role of ontological knowledge representation in PSRS. This role needs to be interpreted within the context of the existing PSRS’ limitations. Data quality has become a focal issue for performing downstream analyses. Patient safety information exists in various types of medical records, including structured and unstructured data (free text) where a great number of information is reported in free text. While this type of reporting can largely retain invaluable information from natural language, it poses a crucial problem of processing free text. When it comes to the structured data, data quality is usually influenced by a pre-defined categorization (Gong, 2011). In many PSRSes that use a hybrid of both unstructured and structured data entry, conflicts were identified between structured data and free text (Holzmueller et al., 2005; Pronovost et al., 2008). Our study demonstrates a feasible approach to incorporate both structured and unstructured data while creating a machine-readable fashion for data representation. Along with this approach, future efforts should include mapping strategies to merge relational data that are used for representing structured data with ontologies (Cullot, Ghawi, & Yétongnon, 2007; Xu, Zhang, & Dong, 2006).

Text data pose technical challenges to computerized data processing and information retrieval. In the patient safety reporting, aggregate analyses are as important as reviewing a handful of cases since it can effectively alarm and trend recurring incidents (Leape et al., 2005). However, performing a manual review on massive reports is costly and unpractical. It may also bring unacceptable deviations to the outcomes (Itoh & Andersen, 2004). Mature NLP solutions and text mining methods are necessary but require a well-developed knowledge base for support. Our study holds promises to address this problem with two advantages: (1) The patient safety ontology serves as a domain knowledge base that can support text mining tasks such as relation extraction and NER. Moreover, a well-designed ontology by itself also provides semantic reasoning functions that can infer new knowledge and support RCA in part (Allemang & Hendler, 2011). (2) The patient safety ontology accelerates information exchange through an unified language system where an uniformed language system provides not only a controlled taxonomy but also the capacity of data integration and knowledge discovery (Alexander, 2006; Bodenreider, 2008).

Our study also enables a number of demanding functionalities in the PSRS. Firstly, classification of patient safety events is critical yet underdeveloped in the
US and many other countries (Elliott et al., 2014; Leape et al., 2005). The ever-increasing volume and complexity of patient safety events call for a uniformed classification system. We envision that an ontology-based multi-label classifier will improve the performance of patient safety classification. The patient safety report is a typical multi-label classification problem in which a given document can be assigned to multiple classes in a hierarchical structure. Therefore, the classification task is denoted as hierarchical multi-label classification (Tsoumakas & Katakis, 2006). The patient safety ontology will define possible classes for the documents and thus enable multi-label learning. Secondly, discovering relatedness between incidents can help identify the contributing factors and understand if repetitive errors worth a broader attention. Semantic similarity metrics have been successfully applied to the Gene Ontology (Lorit et al., 2003; Wolting, McGlade, & Tritchler, 2006). For patient safety ontology, we determine the similarity between two incidents by measuring the distance between concepts/terms annotated from the incident reports by the ontology. In the proposed PSRS, each report is mapped on to the ontology, therefore a set of ontological features such as classes and object properties are assigned to the report. The calculation of the similarity between any of two reports becomes the calculation between the two sets of features associated with the reports. The distances between these features (i.e. classes) are calculated according to their connections defined by the ontology. For example, the distance between two classes is determined by their position in the hierarchies, as well as their hypernym and/or children, in the hierarchies (Garla & Brandt, 2012). Consequently, front end users (i.e. risk managers in the hospitals) are returned with a list of similarity scores when they query the similarity between two or more incidents.

The present work should be discussed in the context of its challenges and limitations. We demonstrated initial steps for improving patient safety reporting through the use of informatics. Constructing patient safety ontology by aligning with different information sources from different perspectives or standards is challenging and has been recognized as a long-term endeavor. With regard to the generalizability of our work, it is worth noting that we are using a small sample of patient safety reports for ontology development and evaluation in this starting stage. The small sample size limits a comprehensive validation from many perspectives in the medical domain. The use of different data source will help discover more knowledge towards a comprehensive ontology. It could show the benefits of our approach if we further expand the ontology alone with the methods we proposed. In conclusion, our future direction will focus on the follow-up ontology development and ontology-based applications using real-world medical data.
Acknowledgements

This project is supported by a grant from AHRQ, 1R01HS022895 and a patient safety grant from the University of Texas system, #156374. We thank Xinshuo Wu, Qi Miao, Swananda Pandit, Drs. Khalid Almoosa and Jing Wang for their input in constructing the ontology and evaluation.

Author Contributions

C. Liang (chen.liang@uth.tmc.edu) and Y. Gong (yang.gong@uth.tmc.edu, corresponding author) conceived the research design. C. Liang designed and implemented the ontology. C. Liang performed the evaluation and data analysis. C. Liang wrote the manuscript under Y. Gong’s supervision. Y. Gong revised the manuscript and approved the final version.

References


This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (http://creativecommons.org/licenses/by-nc-nd/4.0/).
C1. Information Science Roles in the Emerging Field of Data Science
Gary Marchionini

7
Data Science Altmetrics
Mike Thelwall

13
Comparative Study of Trace Metrics between Bibliometrics and Patentometrics
Fred Y. Ye, Mu-Hsuan Huang & Da-Zen Chen

32
Identifying Scientific Project-generated Data Citation from Full-text Articles: An Investigation of TCGA Data Citation
Tian Li, Si Zhang, Hongyi Kang, Zhen He & Ming Nian

45
Mining Related Articles for Automatic Journal Cataloging
Yuqing Hao & Zhiyong Lu

60
Critical Factors for Personal Cloud Storage Adoption in China
Jianya Wang

75
Knowledge Representation in Patient Safety Reporting: An Ontological Approach
Chen Liang & Yang Gong

92
Program of International Conference on Data-driven Discovery: When Data Science Meets Information Science (June 19-22, 2016, Beijing, China)