Guest Editorial

Special Issue on Auditing of Terminologies

Introduction

This special issue is the first in any journal to deal exclusively with the auditing of medical terminologies and ontologies. In the early stages of many emerging technical fields, the emphasis is primarily on creating systems that work. The emphasis eventually shifts to guaranteeing a high level of quality. In fact, two sure signs that a technical field is maturing are extensive activities regarding standards—both national and international—and quality assurance initiatives.

In the field of medical terminologies, we have seen a great deal of standards activity in recent years, such as the work of the ISO/TC215 Working Group 3 (Health Informatics – Semantic Content) and the American National Standards Institute guidelines for designing controlled terminologies (ANSI/NISO Z39.19-2005). There is also a heightened awareness of the need for terminology quality assurance, which has manifested itself in different ways. For example, at the 2007 AMIA Annual Symposium (Chicago), a session was devoted exclusively to terminology assessments (S10). One can also discern a constant stream of research papers in the medical informatics literature dealing with auditing. The breadth of this work, as measured by the number of different terminologies being considered, is growing. The reference list and Table 6 of the paper by Zhu et al. [1] in this special issue give evidence to these observations.

Furthermore, many health care information systems rely on terminologies to provide semantically uniform access to knowledge expressed in different ways and following various paradigms. As a result of this reliance, an error in a terminology may propagate to errors in systems ranging from decision-support systems to pharmacy information systems that are monitoring, for example, drug allergies and drug-drug interactions. Hence, an error in a terminology may further propagate, resulting in an error in the treatment of patients.

Realizing these facts, the guest editors came to the conclusion that the field of medical terminologies is maturing and that there is enough interest within the research community to warrant producing this unprecedented special issue on “Auditing of Terminologies.” The intention is to capture the current state of the art in the field of terminology auditing. In addition to announcing the Call for Papers for this special issue on the JBI Web-site and in other forums, the guest editors directly solicited prominent researchers in this area to submit papers. Fortunately, most of them agreed. The current issue features many authors with outstanding track records in the areas of medical terminologies and their quality assurance.

To ensure high-quality published papers, many of the authors were asked to serve in the role of reviewer as well. Some researchers who were too busy to prepare papers themselves graciously agreed to be reviewers. As a result, all papers (including those that were invited) received extensive reviews, providing in-depth feedback from the perspective of leading researchers. Moreover, the guest editors did not just see themselves as mediators between the authors, reviewers, and JBI’s Editor-in-Chief Dr. Edward H. Shortliffe, but as guides, helping the authors properly reflect the value and potential of their research in their papers. This role of the guest editors was duly noted and pointed out by one of the submitting authors. Even with all the involvement of the guest editors, it was still necessary to reject some papers submitted to the special issue. It is gratefully acknowledged that Dr. Shortliffe brought his vast experience to bear on the improvement of the accepted papers.

The special issue covers a variety of auditing techniques as they are applied to a wide range of terminologies, ontologies and terminological resources. These include the Unified Medical Language System (UMLS), the Systematized Nomenclature for Medicine (SNOMED), the Gene Ontology (GO), the International Classification of Diseases (ICD-10), the National Cancer Institute thesaurus (NCIt), the Foundational Model of Anatomy (FMA), the standardized nomenclature for clinical drugs called RxNorm, and the Medical Entities Dictionary (MED).

Occasionally, researchers in the field of terminology auditing methodologies encounter the misperception that all auditing techniques will be applicable to the entire spectrum of existing terminologies, or at least to many of them. As it happens, terminologies can differ quite drastically from one another in regard to their properties, organizational paradigms, models, grain sizes, coverage, etc. These differences can be due to the idiosyncratic nature of the domain of a terminology, its intended use, or the design choices made by its creators.

When researchers devise techniques for auditing, they typically exploit the properties of the terminology under consideration. This can result in a methodology that is highly specialized for that one terminology and therefore inapplicable to others. In some cases, an auditing technique may be applicable to a family of several structurally similar terminologies.

The view of the guest editors is that a “one size fits all” auditing solution often cannot be achieved, even if it may be desirable as a goal. Sometimes, there is simply no avoiding the fact that a methodology is tailored specifically to the particulars of some terminology. Hence, the guest editors were compelled to accept some papers on auditing techniques applicable to just one terminology. Since all terminologies need to be audited, all approaches were welcome, as long as they were coherent, effective and lucidly explained. Nevertheless, the authors were encouraged to stress general lessons learned that could be of value in the development and application of future auditing techniques.

Drawing on the personal auditing research of the guest editors, we frequently found ourselves asking questions such as: “How...
does the editorial team of a specific terminology actually do auditing? What techniques do they employ?” Such questions typically arose when encountering a correction (between successive versions of the same terminology) for which it was difficult to discern a reason or interpretation. We often found it necessary to surmise the rationale for a correction, due to a lack of documentation of the quality assurance processes or techniques used. Why is it so difficult to obtain this kind of information?

One reason is that researchers are rewarded for publishing novel ideas and techniques. Routine quality assurance measures typically do not fall into this category and are difficult to publish. As a result, there is seemingly a “black hole” in the literature, concerning best practices for auditing existing terminologies.

One goal of this special issue is to remedy the aforementioned situation and to initiate a change in attitude towards the publishing of papers on best practices in auditing. For this reason, a paper category “experience papers” was included in the call for papers of the special issue, and the guest editors reached out to experienced researchers involved in such auditing and maintenance activities for several terminologies. All experience papers received serious scrutiny and went through the same review and revision process as all other papers. The only difference was that the reviewers were notified that these are experience papers and should be evaluated as such, with less of an emphasis on novel techniques. Three experience papers, for the NCIt, the FMA, and the MED were accepted. It is hoped that with these three papers the goal of unearthing hidden practical auditing knowledge in the field was partially achieved. We believe quality-assurance teams for other terminologies will find these reports useful and trust that they will accordingly be encouraged to disseminate their own practices. In the future the guest editors hope to see more such papers published.

The Papers

The first paper in the special issue, “A review of auditing methods applied to the content of controlled biomedical terminologies” by Zhu et al. [1], is a methodology paper providing a comprehensive view of the field. This is a paper that seeks to impose an order on the auditing knowledge accumulated in the field. Its analytic method clearly benefits from the vast experience of its senior author, Dr. J.J. Cimino, a veteran of biomedical terminologies and their auditing, who himself was the principal designer of the MED at New York Presbyterian Hospital. To the best of our knowledge, Cimino was the researcher who coined the term “auditing” with regards to a terminology in his seminal 1998 paper [2]. The current paper classifies many auditing papers according to several dimensions. One dimension is the kind of knowledge utilized in the auditing process: intrinsic vs. extrinsic. Another dimension makes a three-way distinction among the kinds of techniques used. The first of these is manual review, which is inherently labor intensive. Considering the limited resources available for terminology auditing and the large sizes of many terminologies, this kind of approach cannot be applied extensively for purposes of quality assurance, in spite of its obvious appeal. The other two kinds of approaches are “automated systematic” and “automated heuristic.” Both involve applying computational processes of respective kinds that help in identifying potentially problematic features that might indicate errors with a high likelihood. In both cases, a manual review may follow.

The third classification is by the terminology to which the auditing process is applied. The fourth classification is according to the terminology attributes being audited: terms and concepts, semantic classification and semantic relationships. Another classification is by five quality factors: concept orientation, consistency, non-redundancy, soundness, and comprehensive coverage. The paper offers an extensive reference list consisting of 130 publications. Various tables organize the cited papers according to the above classifications.

The remaining papers in this special issue are organized according to the terminology that is audited. The terminologies themselves are ordered according to the number of pertinent papers in the reference list of the methodology paper [1], starting with the most popular one. Thus, the UMLS leads with 51 publications, followed by the SNOMED with 25. Next in the list are the MED with 19 cited papers and the ICD-9/ICD-10 with 11. The remaining terminologies covered in this special issue are the GO with five citations, the NCIt with three, and the FMA with two. RxNorm did not appear in [1]. Below are brief highlights of the other papers. Following the classification of [1], for each paper the knowledge type, the auditing method, and the terminology are listed according to the judgment of the guest editors, whenever these properties are applicable.

In the paper “Auditing associative relations across two knowledge sources,” Vizenor et al. [3] present a method for auditing associative relations that is based on comparing UMLS Metathesaurus relationships with UMLS Semantic Network relationships. They show that of the 177 associative relationship kinds in the Metathesaurus, 48% exhibit a high degree of consistency with the corresponding relationships in the Semantic Network. Looking at concept-relationship-concept triples, the agreement is even higher. Out of 1.8 million such triples, 63% are consistent with the Semantic Network. While these are excellent results, they show directions to be taken in future work. The authors concluded that the semantics of associative relationships should be defined explicitly by their developers. Adding new relationships to the Semantic Network would allow an even closer correspondence between these two UMLS knowledge sources.

Knowledge used: Intrinsic; Auditing method: Automated systematic; Terminology: UMLS

The paper “Analyzing polysemous concepts from a clinical perspective: Application to auditing concept categorization in the UMLS” (Mougin et al.) [4] analyzes polysemous concepts in the UMLS through their categorization using multiple “semantic groups” [5,6]. By a polysemous concept the authors mean a concept with complementary ambiguity regarding manifestations of the same basic meaning of the word as it occurs in different contexts. That is, the different meanings are related to one another. They have analyzed the inheritance of semantic groups in multiple semantic group concepts and manually reviewed the categorization of multiple semantic group concepts for auditing purposes. A large majority of multiple semantic group concepts are found to be polysemous by convention. Only a small portion (about 5%) of these concepts have their multiple categorizations as a result of the integration process utilized by the UMLS. These need to be corrected.

Knowledge used: Intrinsic; Auditing method: Automated heuristic; Terminology: UMLS

“Structural group-based auditing of missing hierarchical relationships in UMLS” by Chen et al. [7] is a continuation of the research started in [8]. The focus in these two papers is on collectively auditing a set of UMLS concepts having similar semantics. One theme in auditing research is that where one error is found, there is a high likelihood of others. The reason for this is that an error may reflect some misconception in modeling, which is likely to be manifested by additional errors. In the case of the current paper, this theme is exercised in a hypothesis that concepts found
to have an erroneous semantic type assignment (in [8]) have a high likelihood of missing a child-of relationship.

A recursive procedure that allows a domain expert, with the support of an algorithm, to locate missing hierarchical relationships is presented. The procedure uses a divide-and-conquer approach on two levels. It starts with a group of concepts having exactly the same semantic type assignments, following the partition of concepts of the UMLS according to the Refined Semantic Network [9,10]. It then partitions the concepts, based on child-of hierarchical relationships, into smaller, singly rooted, hierarchically connected subgroups. The hypothesis is tested and confirmed for two semantic types of the Semantic Network.

Knowledge used: Intrinsic; Auditing method: Automated systematic; Terminology: UMLS

The paper “The Neighborhood Auditing Tool: A hybrid interface for auditing the UMLS” by Morrey et al. [11] describes the principles and implementation of a software tool, called the NAT, for auditing the UMLS. The NAT is based on the observation that auditing at one moment often concentrates on one “focus concept” and its neighborhood. Thus the NAT allows the user to specify several kinds of neighborhoods of different sizes around a focus concept. The neighborhoods are designed to support the auditor with all the knowledge that may be needed in the auditing process, while avoiding overloading the mental capacity of the auditor with too much information. For this, the auditor controls the choice of neighborhood and the navigation steps. The user can navigate from the focus concept to another concept by clicking on any concept in the neighborhood. Importantly, the NAT bridges the gap between diagram-based representations and text-based displays of the UMLS. It introduces the idea of a hybrid display, which maintains the relative positions of concepts as they appear in a diagram, but eliminates the clutter of connecting parent-child links. Thus, all parents of the focus concept appear above it and all children below it, but without connecting arrows, gaining “the best of both worlds” for the display. The NAT is implemented in Java, is Web-accessible, and runs in the JNLP environment which is browser-independent.

Knowledge used: Not applicable; Auditing method: Not applicable; Terminology: UMLS

Wade and Rosenbloom [12] in their paper “The impact of SNOMED CT revisions on a mapped interface terminology: Terminology development and implementation issues” address the issue of dealing with terminologies that evolve over time. The authors manually mapped 1570 concepts from a proprietary interface terminology to two versions of SNOMED CT, in order to gain insight into the differences between these two versions. They show that for these two versions the majority of concept mappings were influenced by changes in the SNOMED CT model, caused by both a restructuring of the top-level hierarchies and by the attributes that are allowed in the respective versions. Whereas these results may be specific to the two versions of SNOMED CT that were used, the re-mapping method provides a way to assess the impact of changes in SNOMED CT.

Knowledge used: Intrinsic; Auditing method: Manual Review; Terminology: SNOMED CT

As its title conveys, “Practical experience with the maintenance and auditing of a large medical ontology” (Baorto et al.) [13] is an experience paper about auditing in the context of maintaining a large, production-quality terminology, namely, the Medical Entities Dictionary (MED) of New York Presbyterian Hospital. One of the design goals of the MED was to bring together disparate controlled terminologies of source systems into a single comprehensive terminology. Due to this, the MED undergoes frequent updates and expansions in order to stay in sync with its sources. Auditing has thus become an integral component of the MED’s maintenance process. Various auditing regimens carried out regularly on the MED are presented. The issue of their timing (e.g., instant audits vs. retrospective audits) is also discussed. Auditing not only has the benefit of helping keep the MED accurate and error-free but also has led to the discovery of errors in the sources.

Knowledge used: Not applicable; Auditing method: Not applicable; Terminology: MED

In their paper “Formalizing the ICD Coding Rules Using Formal Concept Analysis” (Jiang et al.) [14] the authors apply the method of Formal Concept Analysis (FCA) to audit the hierarchy in ICD-10. The rationale of their approach is to create an FCA-based model of (parts of) the ICD-10. This model enables the specification and application of rules for detecting specific patterns. The authors audit ICD-10 with respect to the existence of non-disjoint classes and to the occurrence of anonymous nodes in the FCA-based model, which indicate classes that may need remodeling. The classes that are highlighted by means of FCA need manual review to determine whether and how they actually can be remodeled. Case studies indicate that applying FCA is an effective way to provide starting points for auditing the ICD-10.

Knowledge used: Intrinsic; Auditing method: Automated heuristic; Terminology: ICD-10

In “Applying Evolutionary Terminology Auditing to the Gene Ontology,” Ceusters [15] presents a technique called “Evolutionary Terminology Auditing” (“ETA”) and reports on its application to the vocabularies of the Gene Ontology (GO). ETA has its basis in the idea that changes appearing in a new version of a terminology—such as concept additions or deletions—are manifestations of flaws in the preceding version. A metric is introduced to quantify the improvement of the terminology as it moves from version to version. In this context, improvement refers to the accuracy of the terminology in reflecting the reality of the domain of interest (here, genes). ETA does require that the terminology auditors supply reasons and justifications for the changes they are making, though an exhaustive account is not needed. The application of ETA to the GO has shown its feasibility in assessing the quality of a terminology. Furthermore, it was demonstrated that ETA has the ability to forecast the quality as it evolves over time.

Knowledge used: Intrinsic; Auditing method: Automated heuristic; Terminology: GO

The paper “NCI Thesaurus Quality Assurance Life Cycle” (by de Coronado et al.) [16] is an auditing experience paper. The NCI thesaurus (NCIt) is a large controlled biomedical terminology based on a description logic model, which is heavily used for cancer research. The paper describes both automated and manual techniques used during the life cycle stages of the editing and production processes. The handling of conflicts between needs of applications and terminological best practices is illustrated. The contribution of external auditing for quality assurance purposes is discussed.

Knowledge used: Not applicable; Auditing method: Not applicable; Terminology: NCI Thesaurus

In the paper “Content-Specific Auditing of a Large Scale Anatomy Ontology,” Kalet et al. [17] describe an auditing methodology
for the Foundational Model of Anatomy (FMA) that is based on building an application program accessing the FMA. The application program focuses on the representation of the lymphatic system in the FMA. It is well known that every path through the lymphatic system must terminate at one of two trunks. Finding paths through a network system is a computationally solvable task. Thus, if an application program fails to find correct paths, this indicates a problem in the modeling or data entry of the FMA. The authors describe problems that were found and indicate how the ontology should be adapted to deal with them. Keeping track of lymphatic paths is potentially useful for radiation therapy planning, as tumors may spread downstream in the lymphatic system. The methodology developed in this paper can be extended to other network systems in the body, such as blood vessels.

Knowledge used: Intrinsic; Auditing method: Automated systematic; Terminology: FMA

In their auditing experience paper “Relationship Auditing of the FMA Ontology,” Gu et al. [18] present a method to point out concepts to which erroneous relationships may have been assigned. The method focuses on transitive relationships (part_of, tributary_of, branch_of) in the FMA. The assignments of these relationships are examined by means of algorithms aimed at detecting five categories of potentially incorrectly assigned relationships: circular, mutually exclusive, redundant, inconsistent, and missed entries. In order to determine whether the assignments are actually incorrect, the concepts that are highlighted by this algorithm need to be reviewed manually. Analysis of samples of highlighted concepts indicates that for the majority of these concepts the relationships have actually been incorrectly assigned.

Knowledge used: Intrinsic; Auditing method: Automated systematic; Terminology: FMA

“A Graph-based Approach to Auditing RxNorm” of Bodenreider and Peters [19] proposes a methodology for auditing the relationships of RxNorm, a nomenclature for clinical drug entities, for consistency and completeness. RxNorm divides the space of drug entities into generic and branded and introduces eight high-level categories (called term types) to organize these. The drug entities themselves are defined as instances of the term types. Relationships among the term types model linkages between clinical drugs and their ingredients. A set of constraints is formulated in order to identify 115 “meaningful paths” through RxNorm’s type level, which can be viewed as a graph structure. The auditing effort focuses on exploring alternate paths that are expected to be consistent and identifying those that are not. Graph traversals among the actual drug entities are performed for this purpose. The application of this approach found inconsistencies such as extraneous concepts as well as omitted and extraneous relationships. In addition to finding errors that had eluded RxNorm’s standing quality assurance regimen and proposing changes to it, the paper recommends certain additions to RxNorm’s published documentation.

Knowledge used: Intrinsic; Auditing method: Automated heuristic; Terminology: RxNorm

In “The caBIG Terminology Review Process” (Cimino et al.) [20], members of caBIG’s “Vocabulary and Data Elements Workspace” working group deal with auditing from the perspective of determining whether a terminology is of high enough quality to be worthy of caBIG-approved status. Toward this goal, a large set of criteria is proposed for assessing four general aspects of a terminology or ontology: structure, content, documentation, and editorial process. These criteria were applied successively to the Gene Ontology (GO), the NCIt, the Common Terminology for Adverse Events (CTCAE), and the Logical Observation Identifiers Names and Codes (LOINC). In the course of these applications, the criteria themselves were assessed. This provided an important feedback cycle allowing for the refinement of the criteria. Overall, the authors conclude that the criteria, presented in the form of a matrix, not only can help in judging the usefulness of a terminology for caBIG participants, but they can help guide terminology developers and curators who are ultimately interested in caBIG certification.

Knowledge used: Intrinsic; Auditing method: Not applicable; Terminology: All potential caBIG terminologies; these demonstrated: GO, NCIt, CTCAE, LOINC.

Conclusion

As mentioned in the introduction, it was the declared goal of this first special issue on auditing of terminologies to capture the current state of the art. While the reader will be the final arbiter whether we succeeded, the guest editors would like to reflect on the final product of the invested efforts.

A methodology review paper, which tries to establish some order in this new field by using a multi-dimensional classification, is presented [1]. As with any classification effort, one may argue with the result, e.g., by requesting additional or alternative dimensions. Undoubtedly, other methodology papers with different classification methods will appear in the future. However, the current classification is an important step toward codifying the area of terminology auditing research.

Papers covering a variety of techniques are presented, written by a large number of experts in the field. The techniques in the papers are so different from one another that it was difficult to find two similar papers. Hence, the special issue clearly covers a variety of methods, terminologies, view points and authors. The inclusion of three experience papers, one each for the NCIt, the FMA and the MED, represents an effort to unearth the vast unpublished knowledge about auditing methods used in the practical maintenance of terminologies.

In contrast to the internal auditing performed by maintenance teams, the external auditing by an independent team dedicated to this task is mentioned by several of the papers. The benefits of such independent audits were previously suggested by some of the guest editors [21], following common practice in Software Engineering.

The need for such an independent evaluation is repeated by Bodenreider and Peters [19] as a lesson learned from their external auditing effort on RxNorm. The virtue of external auditing is also discussed by De Coronado et al. [16] with regards to NCIt auditing. Finally, the FMA audit [18] was partially performed as an external audit, where most of the auditing team, headed by Gu, was external [18], while the final verification of the auditing results was done by Mejino, a member of the FMA design team. We hope that these lessons learned will be heeded by researchers and practitioners in the field.

The guest editors would like to conclude with a personal note about the social, as opposed to the technical, dimension of terminology auditing. Being heavily involved in auditing research ourselves, we repeatedly faced the need to communicate our findings to the teams maintaining the audited terminologies. This kind of communication may sometimes feel awkward. On the technical level, researchers often disagree about modeling a domain; after all, there are always various legitimate ways to model the same knowledge. Furthermore, the editors have a broad view of the terminology as a whole, while auditors tend to concentrate on details.

On an emotional level, a terminology represents the brain-child of its editing team, and who likes to hear criticisms of their
creation? Eventually, the realization that both teams have only the best intentions to improve the terminology helped in smoothing the communication between terminology editors and terminology auditors.

We would like to thank the many researchers with whom this special issue as well as our previous auditing work have brought us in contact. They have served as authors, terminology editors, reviewers, informal advisers and very often as inspirations. We are grateful for the lessons learned from them. Since thank you lists always end up omitting a few very deserving people we chose not to provide one, but we are grateful to all. The guest editors would also like to reiterate their gratitude to Dr. Shortliffe, Janine Burch at JBI/Elsevier in San Diego, and the National Library of Medicine for their support.

References


J. Geller
New Jersey Institute of Technology,
323 Dr. King Blvd., Newark, NJ, USA
E-mail address: james.geller@gmail.com

Y. Perl
New Jersey Institute of Technology,
323 Dr. King Blvd., Newark, NJ, USA

M. Halper
Kean University, Union, NJ, USA

R. Cornet
Academic Medical Center,
University of Amsterdam, Amsterdam,
The Netherlands