

Using Standardized Lexicons for Report Template Validation with LexMap, a Web-based Application

Jason Hostetter · Kenneth Wang · Eliot Siegel ·
Jeremy Durack · James J. Morrison

Published online: 6 January 2015
© Society for Imaging Informatics in Medicine 2015

Abstract An enormous amount of data exists in unstructured diagnostic and interventional radiology reports. Free text or non-standardized terminologies limit the ability to parse, extract, and analyze these report data elements. Medical lexicons and ontologies contain standardized terms for relevant concepts including disease entities, radiographic technique, and findings. The use of standardized terms offers the potential to improve reporting consistency and facilitate computer analysis. The purpose of this project was to implement an interface to aid in the creation of standards-compliant reporting templates for use in interventional radiology. Non-standardized procedure report text was analyzed and referenced to RadLex, SNOMED-CT, and LOINC. Using JavaScript, a web application was developed which determined whether exact terms or synonyms in reports existed within these three reference resources. The NCBO BioPortal Annotator web service was used to map terms, and output from this application was used to create an interactive annotated version of the original report. The application was successfully used to analyze and modify five distinct reports for the Society of Interventional Radiology's standardized reporting project.

Keywords Controlled vocabulary · Data display · Education · Medical · Internet technology · Structured reporting

J. Hostetter (✉) · E. Siegel · J. J. Morrison
Department of Radiology, University of Maryland, 22 S. Greene St.,
Baltimore, MD 21201, USA
e-mail: jhostetter@gmail.com

K. Wang · E. Siegel
Baltimore VA Medical Center, Baltimore, MD, USA

J. Durack
Memorial Sloan Kettering Cancer Center, New York, NY, USA

Introduction and Background

In the field of radiology, unstructured medical data in the form of images and report text are continually being created. After image acquisition and interpretation, the majority of radiologic information becomes archived without an easy way to aggregate common findings among patients or explore relationships between clinical data, imaging findings, and disease. The Radiological Society of North America (RSNA) has promoted the creation and sharing of standardized report formats and vocabulary through the radiology reporting initiative [1]. Separately, standard medical lexicons have been designed to unify reporting terminology and provide context for machine processing. Increased use of ontologies and lexicons such as the Radiology Lexicon (RadLex) [2], Logical Observation Identifiers Names and Codes (LOINC) [3], and Systemized Nomenclature of Medicine Clinical Terms (SNOMED-CT) [4] can improve data mining capabilities as well as facilitate research relative to non-standardized radiology reports [5, 6].

Tools such as the National Center for Biomedical Ontology (NCBO) Annotator [7] and the Radiological Society of North America (RSNA) RadLex Term Browser [8] provide general purpose interfaces for searching standardized vocabularies and basic text annotation. Using these resources, a block of input text can be analyzed by matching words or phrases to corresponding terms in selected vocabularies. The output of these tools can be cumbersome for arbitrary text annotation, particularly, the large blocks of text present in radiology reports or templates. For example, annotating the RSNA RadReport CT Abdomen template [9] with SNOMED-CT using the NCBO Annotator produces 109 direct results, formatted as a table with each result in a separate row. Evaluating a large number of results for clinical or research purposes can be challenging, even from a short source text. The display of results in a table separate from the source text removes

valuable contextual information and requires the user to match result terms with the original input. Alternatively, the RadLex Term Browser produces very simple and easy-to-understand results (i.e., whether a term is contained within RadLex and its hierarchy); however, the interface is limited to single terms entered manually [10].

To overcome these limitations, we sought to develop an interface capable of accepting input text of arbitrary length that produced term maps in a format similar to the original input. The goal was to preserve the original text formatting, enabling a visual overlay of mapped annotations to preserve the relationship to the source text. An additional goal was to create an interactive web-based application that allows for variable levels of detail to be presented according to user requirements. An online tool called LexMap was created to evaluate report terminology during development of structured reporting templates. Using the annotated output from the tool, the vocabulary used in each template was optimized to include the preferred RadLex, SNOMED-CT, and LOINC terminology.

Application Architecture

The application can be conceptually split into two main parts, the ontology/lexicon data analyzer and the user interface. The data analyzer includes the server-side application that interacts with the NCBO Annotator. The user interface includes the text/file input control and results display (Fig. 1).

Data Analyzer

The publicly available NCBO Annotator application programming interface (API) provided at bioontology.org was used to interact with ontological data. The API is a Representational State Transfer (REST) interface, which allows HTTP calls to define parameters for a query. The LexMap server utilizes Node.js [11], a server-side implementation of the JavaScript programming language, and Express [12], a web application framework for Node.js, to perform the HTTP calls and process

the results. A query is sent to the Annotator service containing the text provided by the user, and the NCBO Annotator server responds with an XML formatted document describing matched terms. The returned XML data are converted to a JavaScript object and organized to include only data fields relevant to the result output. For instance, we do not use term hierarchy data such as parent or child terms in our output display, so these data are discarded.

Once the requested JavaScript object is created, it is returned to the browser client in JavaScript Object Notation (JSON) format. Using the JSON format allows seamless transfer of JavaScript Objects between the server and client. The result object consists of an array of matched terms with metadata including the term's beginning and ending character positions within the input text, hierarchical categorization, and a link to the bioontology.org resource listing (Fig. 2).

(Note: At the time of this writing, NCBO bioontology.org is in the process of releasing a new data API supporting multiple output formats, including JSON. Once this upgrade is in place and supported on all ontologies, the need for a conversion step between XML and JSON would no longer be necessary, simplifying the data processing steps above.)

User Interface

The user interface allows two types of text input. A text box allows the user to manually enter or copy-and-paste arbitrary text to be processed. Alternatively, a user can upload Nuance RadPortal (Nuance Communications, Inc., Burlington, MA) dictation templates directly into the application for processing. The relevant text data is extracted from the Nuance RadPortal formatted files (compatible with both Nuance RadWhere and Powerscribe 360) before being sent to the server for annotation.

The main function of the client is to convert the returned term annotations into an interactive display. Several data transformations are performed in order to accomplish this. Since a single term or phrase could match in multiple different contexts within multiple ontologies or lexicons, the result annotations often contain multiple entries for the same term.

Fig. 1 Application data flow diagram

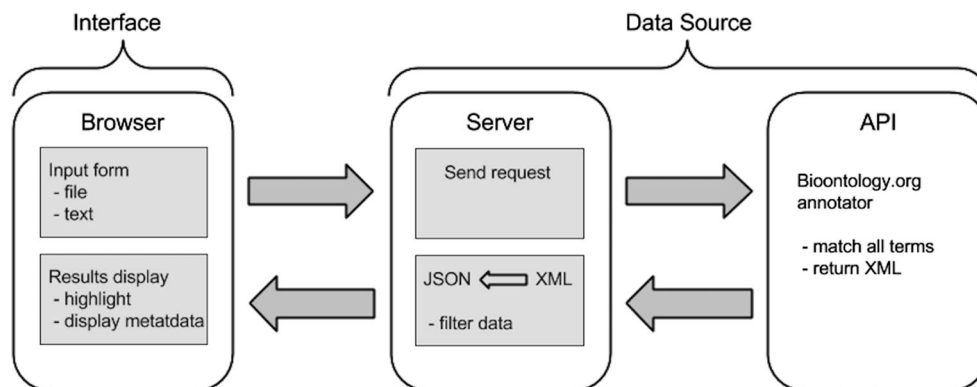


Fig. 2 Example of bioontology.org API output converted to a JSON object

```

1 var terms = [
2   {
3     "term": "malignant",
4     "from": 43,
5     "to": 51,
6     "isA": "RadLex entity",
7     "link": "http://purl.bioontology.org/ontology/RADLEX/RID15655"
8   },
9   {
10    "term": "neoplasm",
11    "from": 53,
12    "to": 57,
13    "isA": "pathophysiologic finding",
14    "link": "http://purl.bioontology.org/ontology/RADLEX/RID3957"
15  },
16  {
17    "term": "embolization",
18    "from": 30,
19    "to": 41,
20    "isA": "RadLex entity",
21    "link": "http://purl.bioontology.org/ontology/RADLEX/RID10404"
22  },
23  ...
24 ]

```

For example, the phrase “uterine artery embolization” matches entries for “uterine artery,” “artery,” and “uterine artery embolization” in RadLex, SNOMED-CT, and LOINC. Within SNOMED-CT, the phrase uterine artery matches both “Structure of uterine artery” and “Entire uterine artery.” These entries are aggregated into a single entry for each term, populating sub-properties of that term with the data from each individual entry. We then rebuild and display the original input text, preserving formatting and highlighting matched terms. When selected, matched terms display additional data including which ontologies matched the term, the type of term (preferred term or synonym), and links to the bioontology.org term definitions (Fig. 3).

The interface utilizes the standard web technologies HTML5, JavaScript, and Cascading Style Sheets (CSS). Multiple open source development libraries were used allowing rapid development of the application. These included AngularJS [13] for JavaScript application development and Bootstrap [14] for CSS development. AngularJS is an open source client-side JavaScript library that provides tools for web application development such as two-way data binding between JavaScript and the Document Object Model (DOM) and an asynchronous HTTP request library. The DOM is the standard convention for HTML representation and interaction within browsers. Two-way data binding provides more flexible, consistent control of interface elements from client-side JavaScript code. A user interface element within the DOM is

bound to a code variable, such that changes in one are instantly reflected in the other without the need to explicitly update values programmatically. Bootstrap is an open source CSS framework that provides visual style definitions for HTML components, page layout, typography, and other web interface elements.

Use Case

The Society of Interventional Radiology (SIR) Standardized Reporting Committee is in the process of creating structured reporting templates to cover the breadth of procedures performed using image guidance. Five sample templates were created for common interventional radiology procedures including venous filter insertion, venous filter removal, central venous access, vertebral augmentation, and uterine fibroid embolization. To facilitate future template development, four sub-templates were also produced that contained common reporting elements such as sedation medications and antibiotic administration. The reporting templates are made up of text selection fields and pick-lists developed and captured using Nuance RadPortal file formats.

Our application was used to verify each of the five full structured reporting templates and four sub-templates against RadLex, SNOMED-CT, and LOINC ontologies. For

Text

File

RadReport

Choose File

Central venou...L export.xml

Verify

FYI: The annotator can take a long time with large files.

Options

Choose which ontologies to verify against:

☒ Radlex
☒ SNOMED
☒ LOINC

Signer name

IR Fellow Name

Resident Name

Mid-level Provider Name

Diagnosis

Indications

Prior imaging/comparison

Access method

Side

Vein accessed

Breast cancer

Lymphoma

Acute renal failure

Chronic renal failure

Healthy donor

Colorectal cancer

Hepatocellular carcinoma

Prostate cancer

None

Study description (study type, date, findings)

Ultrasound guided

Palpation guided

Anatomic landmark guided

Right

Left

Midline

Other (specify)

Internal jugular vein

Subclavian vein

RADLEX

Matched:

PROSTATE CANCER

RADLEX

Matched:

PROSTATE

LOINC

Matched:

PROSTATE

SNOMEDCT

Matched Synonym:

PROSTATE

Preferred Name:

Prostatic structure

Fig. 3 LexMap interface. The *left column* contains text input options. The file upload option allows the annotation of an exported dictation template. The *middle column* displays results in an easily readable format,

and matched terms are highlighted. The *right column* displays additional metadata for matched terms when the term is clicked

example, the venous filter insertion template contains 2145 words, of which 1599 terms were contained within one of the selected ontologies for an overall coverage of 74.5 %.

Through this verification process, a set of terms was identified which were absent from the selected ontologies but considered by the authors to be relevant medical or radiologic words/phrases. A total of 30 radiology-specific terms were subjectively identified which did not currently exist in Radlex, LOINC, or SNOMED-CT. Non-radiologic terms such as “for” or “of” were ignored. These terms were submitted for addition to Radlex, and 17 of the suggested terms have been added to RadLex at this time (Table 1).

Discussion

The Lexmap application provides a simple, easy-to-understand interface for the RadLex, SNOMED-CT, and

LOINC lexicons. The tool can be utilized for many tasks from verifying the existence of a single term within a lexicon to matching terms in a large structured reporting template for maximum compliance with one or more lexicons. Simplified interfaces allow users to interact with and learn about ontologies and lexicons in an intuitive way. RadLex, LOINC, and SNOMED-CT were chosen for this project to suit our specific needs; however, the application architecture could be used with minimal modification for any standardized vocabulary supported by the NCBO Annotator.

As structured reporting becomes more widely utilized, mapping templates to standardized terminologies will become increasingly important. Contextual display of ontology information within reporting systems will be an important consideration for integrated interfaces, and we describe one example of how this might be accomplished. Simplifying the process of mapping will help encourage

Table 1 Selected terms that did not exist in RadLex, LOINC, or SNOMED-CT. These terms were submitted to RadLex for addition to the lexicon. The terms in italics have been added

<i>PICC</i>	<i>Ultrasound-guided</i>	Palpation-guided	Landmark-guided	<i>Cavoatrial</i>	<i>Suprarenal</i>
<i>Infrarenal</i>	<i>Surrogate</i>	<i>Monitored</i>	<i>Fluoroscopic</i>	<i>Mid-level</i>	Fellow
<i>Attestation</i>	<i>Occlusive</i>	<i>Non-occlusive</i>	<i>Fully</i>	<i>Partially</i>	<i>Deployed</i>
<i>Tilted</i>	<i>Aborted</i>	Embolized	Biplane	Sacroplasty	Unipedicular
Parapedicular	Chromophobe	Irreversible	Electroporation	Intercostal	Bipedicular

creators and users of structured report templates to adhere to standard terms as closely as possible.

Using the LexMap interface, nonstandard terms can be more easily identified, and the user can then decide to use an alternative standard term. If a matched term is a synonym of a more preferred term within the lexicon, the term is labeled as a synonym and the preferred term is displayed so the user can decide whether the preferred terminology is more appropriate. Users may also identify nonstandard terms that they feel should be included in the standardized lexicon and submit them for inclusion, as we have done. Ultimately, decisions regarding which terms should or should not be included are made by curators of a given lexicon or ontology.

Our application also demonstrates the utility of developing custom interfaces for interacting with complicated data sets to display information in useful and intuitive ways. The barrier to entry for individuals to develop robust applications using publicly available resources has been lowered with the growth of the open source software community and wide availability of free software development tools. The use of web technologies not only enables rapid development but also facilitates distribution of tools to clinical and research users to improve workflows and drive further innovation.

There are several limitations of the described tool. The interface relies on the Annotator REST service maintained by the NCBO Biportal, and performance is therefore limited in part by network latency and performance of the NCBO Annotator. In our environment, the user could wait up to 8 s for annotation results, which may be unacceptable for some users and workflows. As the developer interacts with the NCBO Biportal Annotator through an API, the underlying code is inaccessible and only client code can be optimized for performance. Therefore, some performance limitations may be insurmountable by the application developer alone.

Application responsiveness is also partly dependent on network speed for sending and receiving requests, which could vary significantly depending on a user's specific network connection. Network speed constraints could potentially be improved by using the NCBO virtual appliance, a freely available virtual machine image that includes the Annotator service software and allows installation within a local network. However, this strategy would potentially incur additional server and maintenance costs. This trade-off of convenience versus control must be considered when developing an application dependent on an external API.

Another aspect of the Annotator interface is its requirement that terms be at least three characters in length. While this is useful to exclude short natural language words such as “is” and “a,” this also excludes terms of interest such as units of measurement (e.g., cm, mm, mL) as well as modality abbreviations (e.g., CT, US). In our use case, these terms were manually validated through the RadLex Term Browser whenever possible. Similar manually entered term browsers are available for LOINC and SNOMED-CT.

Only structured reporting templates for selected interventional radiology procedures were evaluated using LexMap. While the complexity and procedural nature of these reports are not necessarily representative of all radiology reporting, LexMap could easily be applied to any diagnostic imaging report or other text-based document.

Conclusion

We developed an open source web-based application called LexMap to improve radiology report text mapping to large medical lexicons and ontologies. The authors used this application to successfully optimize the terminology used in interventional radiology structured reporting templates, although it could be utilized across the spectrum of radiology. The potential barrier to creating custom interfaces for specific use cases has been lowered due to the widespread availability of free and open source development tools and web-based APIs. Tools such as Lexmap facilitate real-world practical application of medical lexicons and ontologies by lowering barriers to adoption and utilization.

LexMap is freely available at <http://lexmap.xrlabs.com>.

The source code has been made publicly available at <https://github.com/jmhmd/lexmap> under the GNU General Public License (<http://www.gnu.org/licenses/gpl.html>).

References

1. Radiological Society of North America. Radiology Reporting Initiative. Available from: http://www.rsna.org/Reporting_Initiative.aspx
2. Langlotz CP: RadLex: a new method for indexing online educational materials. *Radiographics* 26(6):1595–1597, 2006

3. Inst. R. LOINC 1994 [cited 2014 March 24]. Available from: loinc.org.
4. Organisation IHTSD. SNOMED-CT SNOMED-CT [cited 2014 March 24]. Available from: <http://www.ihtsdo.org/snomed-ct/>
5. Reiner BI, Knight N, Siegel EL: Radiology reporting, past, present, and future: the radiologist's perspective. *J Am Coll Radiol: JACR* 4(5):313–319, 2007
6. Rubin DL, Flanders A, Kim W, Siddiqui KM, Kahn Jr, CE: Ontology-assisted analysis of Web queries to determine the knowledge radiologists seek. *J Digit Imaging* 24(1):160–164, 2011
7. Jonquet C, Shah NH, Musen MA: The open biomedical annotator. *Summit Transl Bioinforma* 2009:56–60, 2009
8. Radiological Society of North America. RadLex 2006 [cited 2014 March 24]. Available from: <http://www.radlex.org>
9. Lakhani P CG, Kahn CE: CT Abdomen 2012 [cited 2014 March 20]. Available from: <http://www.radreport.org/txt/0000234>
10. Woods RW, Eng J: Evaluating the completeness of RadLex in the chest radiography domain. *Acad Radiol* 20(11):1329–1333, 2013
11. Joyent I: NodeJS [cited 2013 December 25]. Available from: <http://www.nodejs.org>
12. TJH. Express [cited 2014 March 20]. Available from: <http://expressjs.com>
13. Google I. AngularJS [cited 2013 December 25]. Available from: <http://www.angularjs.org>
14. Otto M, Thornton J: Bootstrap [cited 2013 December 25]. Available from: <http://www.getbootstrap.com>