Integrating Wikipedia Articles and Images into an Information Resource for Radiology Patients

Teresa Martin-Carreras¹ · Charles E. Kahn Jr.^{1,2,3}

Published online: 6 November 2018
© Society for Imaging Informatics in Medicine 2018

Abstract

Wikipedia—an open-access online encyclopedia—contains a large number of medically relevant articles and images that may help supplement glossaries of radiology terms. We sought to determine the extent to which concepts from a large online radiology glossary developed as part of the Patient-Oriented Radiology Reporter (PORTER) initiative could be mapped to relevant Wikipedia web pages and images using automated or semi-automated approaches. The glossary included 4090 concepts with their definitions; the concept's preferred name and lexical variants, such as plurals, adjectival forms, synonyms, and abbreviations, yielded a total of 13,030 terms. Of the 4090 concepts, 3063 (74.9%) had a corresponding English-language Wikipedia page identified by automated search with subsequent manual review. We applied the MediaWiki application programming interface (API) to generate web-service calls to identify the images from each concept's corresponding Wikipedia page; three reviewers selected relevant images to associate with the glossary's concepts. Licensing terms for the images were reviewed. For 800 randomly sampled concepts that had associated Wikipedia pages, 362 distinct images were identified from the MediaWiki library and matched to 404 concepts (51%). Three images (1%) had unspecified licensing terms; the rest were in the public domain or available via a Creative Commons license. Wikipedia and the MediaWiki library offer a large collection of medical articles and images that can be incorporated into an online lay-language glossary of radiology terms through a semi-automated approach.

Keywords Radiology · Reporting · Patient- and family-centered care · Vocabularies · Wikipedia

Introduction

Patients and their family members frequently seek health information online, and various Internet-based information resources have been developed to meet their needs [1, 2]. The “patient portal”—a means to access the electronic health record (EHR)—has given further support and impetus for patients to access their personal health information. In radiology, there is a growing awareness of patients' needs for information, and for tools to meet those needs [3–5].

Various health information resources are available online, and one of the most frequently accessed is the English-language Wikipedia website (<https://en.wikipedia.org>) [6]. Wikipedia is a multilingual, multi-authored, web-based encyclopedia of freely available and openly editable content. Its name is a portmanteau of the words “wiki” (a technology for creating collaborative websites, from the Hawaiian word “wiki,” meaning “quick”) and “encyclopedia.” As of July 14, 2018, the English-language Wikipedia site incorporated 5,680,749 content articles, 45,355,735 pages in total, and 870,643 uploaded files [7]. Wikipedia included 30,000 medical articles and another approximately 6000 articles on human anatomy in the English language as of March 2017 [8, 9]. Wikipedia's content licensing allows text and most images to be reused in other settings as long as one includes attribution to the material's source.

Various glossaries of radiology terms have been developed [10–15]. To our knowledge, these glossaries have been deployed as stand-alone information resources or contain definitions suited for medical professionals. In an effort to help lay readers better understand their radiology reports, a glossary of radiology terms has been developed as part of the Patient-Oriented Radiology

---

✉ Charles E. Kahn, Jr.
ckahn@upenn.edu

¹ Department of Radiology, Hospital of the University of Pennsylvania, 3400 Spruce St., Philadelphia, PA 19104, USA

² Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, USA

³ Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, USA

Reporter (PORTER) initiative, which seeks to annotate imaging procedure reports with lay-language definitions of radiology terms [16]. The PORTER radiology glossary was created initially as a set of 310 terms in the domain of MRI of the knee [17]. Glossary terms that appear in radiology reports were linked to their referenced Wikipedia pages; upon hovering over a glossary term, the selected image from Wikipedia was displayed.

To extend the PORTER glossary, a list of frequently appearing terms was generated from a separate set of 1000 randomly selected radiology reports from the authors' radiology department. These reports were broken into single words and word pairs. Terms that occurred with the greatest frequency were considered for incorporation, excluding stopwords (frequently appearing words, such as "the"). In addition, terms were incorporated from local radiology department reporting templates and from the RSNA report template library (<http://www.radreport.org>). Various RadLex terms, imaging signs, and Fleischner Society thoracic imaging terms were added [16]. Lexical variants were incorporated by manual review, such as adjectival forms (e.g., *hepatic* for *liver*). Abbreviations were added only when unambiguous. Plural forms were generated through a semi-automated approach using a set of heuristics. Definitions for all of the glossary's terms were developed to be readable by an average US adult; about 90% of definitions were readable at an 8th-grade level or lower [16].

PORTER's glossary is used to annotate radiology reports with lay-language definitions, relevant illustrations, and links to Wikipedia pages. In a clinical trial of a prototype version of PORTER, 77% of patients found its lay-language definitions to be helpful, and 91% indicated that the illustrations which accompanied many of the terms were helpful as well [18]. In the prototype system, references to Wikipedia pages and their embedded images were selected manually. This approach was highly suitable for a prototype that included only 180 concepts limited to the domain of knee MRI. As PORTER's glossary has grown to accommodate terms for all areas of radiology, we sought to determine the extent to which concepts from the PORTER glossary could be mapped to Wikipedia web pages and images using automated or semi-automated approaches.

Methods and Materials

PORTER Radiology Glossary

At the time of this study, PORTER's glossary incorporated 4090 concepts, all of which included lay-language definitions. Each concept included one or more terms: the name of the concept itself plus its lexical variants, such as plurals, adjectival forms, synonyms, and abbreviations. For example, the concept *meniscus* included three terms: the concept name itself (*meniscus*), adjectival form *meniscal*, and plural form *menisci*. Similarly, the concept *hepatocellular carcinoma*

included the synonym *hepatoma* and the abbreviation *HCC*. The glossary contained a total of 13,030 terms.

Identifying Corresponding Wikipedia Pages

An automated script was developed to search for relevant Wikipedia pages using Wikipedia's application programming interface (API); for each glossary concept, the script searched first using the concept name and then using the lexical variants. Wikipedia includes numerous synonyms and alternate names that "redirect" a web browser to the primary Wikipedia page. For example, entering the PORTER glossary term *occipital horn* found the Wikipedia page for "Lateral ventricles." Wikipedia includes "disambiguation" pages that identify terms with multiple meanings; such pages provide links to multiple pages. For ambiguous terms, the relevant Wikipedia page was selected by manual review. Matching Wikipedia pages were reviewed to assure that the pages would be useful for patients. For example, although the glossary provided a definition for the concept *accuracy*, the matching Wikipedia page "Accuracy" was considered to be of little value to a patient, and hence no Wikipedia reference was listed for that concept. We tallied the number of glossary concepts that were mapped accurately and meaningfully to Wikipedia pages.

To determine if Wikipedia could provide relevant images for lay readers, we developed a semi-automated approach to identify images from Wikipedia references. Three individuals—a radiologist with more than 20 years of experience, a radiology resident, and college-educated individual without medical background—participated as reviewers. Before setting out to address all of the 3063 concepts that had matching Wikipedia references, we wanted to assess a manageable sample that would have an error less than $\pm 4\%$. We randomly selected 800 glossary concepts from among those 3063 concepts that had an associated Wikipedia entry. An automated script retrieved the Wikipedia page's image references and presented them to the reviewer. The reviewer selected the image that he or she judged to be most relevant and useful for a patient. If none of the images was considered suitable for a lay viewer, then the reviewer selected "None" for that glossary concept. Image metadata, such as caption, image dimensions, file size, author credits, license, and copyright status, were recorded for all imported images. We tallied the number of concepts with selected images and the number of images overall. For each image incorporated into the collection, we tallied the models under which they were licensed for use.

Results

Of the glossary's 4090 concepts, 3063 (74.9%) were assigned a corresponding English-language Wikipedia page. Examples are shown in Table 1. Of 800 concepts sampled randomly

Table 1 Example terms from the glossary and their matching Wikipedia entries

Glossary concept	Wikipedia entry
Afebrile	Human body temperature
Lingual nerve	Lingual nerve
Lingual thyroid	Thyroid dysgenesis
Swan-Ganz catheter	Pulmonary artery catheter
Chondromalacia	Chondromalacia patellae
Post-operative	Surgery
Bunionectomy	Bunion
CADASIL	CADASIL syndrome
Retinoblastoma	Retinoblastoma
Ventromedial hypothalamic nucleus	Ventromedial nucleus of the hypothalamus
Occult fracture	Bone fracture
Biliopancreatic bypass	Bariatric surgery
Occipital horn	Lateral ventricles

from among the 3063 concepts that had an associated Wikipedia page, an appropriate image from the Wikimedia Commons library was identified for 404 concepts (51%); the 95% confidence interval ranged from 47 to 54%. Thus, one can estimate that 1547 (51% of 3063) concepts with Wikipedia references will have images, which represents 37.8% of all 4090 glossary concepts. A total of 362 distinct images were identified for the 404 image references, as several concepts were matched to the same image. For instance, the terms *pulmonary vessels* and *pulmonary artery* were matched with the same image (Fig. 1). Details of the images' licenses are listed in Table 2; three of the 362 images (1%) had unspecified licensing terms.

Discussion

Health Information Resources

Online information resources have sought to improve patients' access to health information. MedlinePlus, a web-

based consumer health information resource developed by the US National Library of Medicine (NLM), provides access to reliable full-text health information. The site matches consumer language to medical terminology and includes a preformed PubMed search query for each health topic [19, 20]. RadiologyInfo.org is a web-based information resource developed by the Radiological Society of North America and the American College of Radiology that explains more than 100 radiologic examinations and treatments [21, 22]. PORTER matches significantly more terms in radiology reports than MedlinePlus and its definitions are more readable than those of the RadLex vocabulary [16]. In a clinical trial of the prototype 310-term PORTER glossary, 77% of patients found the lay-language definitions to be helpful and 91% indicated that the illustrations that accompanied many of the terms were helpful as well [18].

Wikipedia as a Health Information Resource

Wikipedia and the Wikimedia Commons library offer a large collection of medical images and illustrations. The Wikipedia

Fig. 1 Public-domain illustration of pulmonary circulation from Wikipedia (https://commons.wikimedia.org/wiki/File:Illustration_of_pulmonary_circulation.jpg) selected for PORTER terms *pulmonary vessels* and *pulmonary artery*

Table 2 Licensing models of images identified from Wikimedia Commons

Abbreviation	License name	Number of images
PD	Public Domain	136
CC0	Creative Commons “No Rights Reserved”	9
CC BY*	Creative Commons “Attribution”	76
CC BY-SA*	Creative Commons “Attribution Share-Alike”	138
(None)	Unspecified	3
Total		362

*The license abbreviation includes a version number (e.g., CC-BY-2.0); here, the licenses are aggregated by type

API allowed our semi-automated approach to efficiently match relevant pages and illustrations to terms in the PORTER glossary. Such information provides freely available public-domain information to supplement the lay-language definitions of radiology terms. There is no restriction to reference Wikipedia documents through a link; the text content can be incorporated into a document as long as attribution is given to the source. We chose to explore references to Wikipedia because it was free of advertising, and because content such as images could be repurposed freely under the existing license.

More than 99% of the images selected for the glossary had licensing terms that allowed reuse. The Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>) allows the licensee to copy and redistribute the material in any medium or format, and to transform and build upon the material for any purpose, even commercially, with attribution to the licensor. The licensee must give appropriate credit, provide a link to the license, and indicate if any changes were made in any reasonable manner that does not suggest the licensor endorses the licensee or his/her use of the material. The licensee cannot place any additional restrictions by applying legal terms or technological measures that legally restrict others from doing anything the license permits. The Creative Commons Attribution-ShareAlike License (<https://creativecommons.org/licenses/by-sa/4.0/>) includes the additional provision that that licensee must make any adapted version of the material available under the same license provisions. Materials in the public domain and under the Creative Commons No Rights Reserved License may be used without restriction.

How complete, accurate, reliable, and readable is Wikipedia’s health information? Internal and external quality metrics have shown that its articles are highly variable in quality, but are improving [8]. Pharmaceutical information in Wikipedia has been found to have inconsistent quality [23, 24]. Another potential drawback is that Wikipedia pages may require college-level reading ability, which is higher than that recommended for lay readers [24–27]. Unlike many health information websites, Wikipedia does not include advertisements. Although the authors have reservations about the sole use of Wikipedia as a health information resource, a great majority of the terms in the PORTER glossary reference

anatomy and medical terminology, and many of these terms have standardized and accepted definitions. We believe that Wikipedia’s more detailed information may provide a useful counterpoint to the PORTER glossary’s simplified definitions.

Conclusion

The current work allowed rapid and efficient selection of related Wikipedia articles and images to enhance the utility of an online glossary of radiology terms. Wikipedia articles were assigned for about 75% of total concepts in the PORTER glossary, and we estimate that suitable images can be assigned for 38% of all concepts. Work is underway to implement the online glossary into the patient portal in a large US health system, which will allow assessment of the utility of the Wikipedia links and illustrations.

References

1. Ball MJ, Lillis J: E-health: transforming the physician/patient relationship. *Int J Med Inform* 61:1–10, 2001
2. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW: Patients’ use of the Internet for medical information. *J Gen Intern Med* 17:180–185, 2002
3. Arnold CW, McNamara M, El-Saden S, Chen S, Taira RK, Bui AA: Imaging informatics for consumer health: towards a radiology patient portal. *J Am Med Informatics Assoc* 20:1028–1036, 2013
4. Bruno MA, Petscavage-Thomas JM, Mohr MJ, Bell SK, Brown SD: The “open letter”: radiologists’ reports in the era of patient web portals. *J Am Coll Radiol* 11:863–867, 2014
5. Itri JN: Patient-centered radiology. *RadioGraphics* 35:1835–1846, 2015
6. Laurent MR, Vickers TJ: Seeking health information online: does Wikipedia matter? *J Am Med Informatics Assoc* 16:471–479, 2009
7. Wikipedia: About. Available at <https://en.wikipedia.org/wiki/Wikipedia:About>. Accessed 16 July 2018
8. Shafee T, Masukume G, Kipersztok L, Das D, Haggstrom M, Heilman J: Evolution of Wikipedia’s medical content: past, present and future. *J Epidemiol Community Health* 71:1122–1129, 2017
9. Heilman JM, West AG: Wikipedia and medicine: quantifying readership, editors, and the significance of natural language. *J Med Internet Res* 17:e62, 2015
10. Hunter TB, Taljanovic MS: Glossary of medical devices and procedures: abbreviations, acronyms, and definitions. *RadioGraphics* 23:195–213, 2003

11. Buscombe J: A new glossary for nuclear medicine. *Nucl Med Commun* 26:85–86, 2005
12. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Muller NL, Remy J: Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 246:697–722, 2008
13. Kundu S, Itkin M, Gervais DA, Krishnamurthy VN, Wallace MJ, Cardella JF, Rubin DL, Langlotz CP: The IR RadLex project: an interventional radiology lexicon—a collaborative project of the Radiological Society of North America and the Society of Interventional Radiology. *J Vasc Interv Radiol* 20:433–435, 2009
14. Langlotz CP: RadLex: a new method for indexing online educational materials. *RadioGraphics* 26:1595–1597, 2006
15. D'Orsi CJ: American College of Radiology: 2013 ACR BI-RADS Atlas: Breast Imaging Reporting and Data System. American College of Radiology: Reston, 2014
16. Martin-Carreras T, Kahn CE Jr: Coverage and readability of information resources to help patients understand radiology reports. *J Am Coll Radiol*, 2017. <https://doi.org/10.1016/j.jacr.2017.11.019>
17. Oh SC, Cook TS, Kahn CE Jr: PORTER: a prototype system for patient-oriented radiology reporting. *J Digit Imaging* 29:450–454, 2016
18. Cook TS, Oh SC, Kahn CE Jr: Patients' use and evaluation of an online system to annotate radiology reports with lay-language definitions. *Acad Radiol* 24:1169–1174, 2017
19. Burgess S, Dennis S, Lanka S, Miller N, Potvin J: MedlinePlus Connect: linking health IT systems to consumer health information. *IT Prof* 14:22–28, 2012
20. Miller N, Lacroix EM, Backus JE: MEDLINEplus: building and maintaining the National Library of Medicine's consumer health Web service. *Bull Med Libr Assoc* 88:11–17, 2000
21. Ellenbogen PH, Tashjian JH: RadiologyInfo: reaching out to touch patients. *Radiology* 246:49–55, 2008
22. Rubin GD, Krishnaraj A, Mahesh M, Rajendran RR, Fishman EK: Enhancing public access to relevant and valued medical information: Fresh directions for RadiologyInfo.org. *J Am Coll Radiol* 14: 697–702 e694, 2017
23. Kupferberg N, Protus BM: Accuracy and completeness of drug information in Wikipedia: an assessment. *J Med Libr Assoc* 99: 310–313, 2011
24. Candelario DM, Vazquez V, Jackson W, Reilly T: Completeness, accuracy, and readability of Wikipedia as a reference for patient medication information. *J Am Pharm Assoc* (2003) 57:197–200.e191, 2017
25. Modiri O, Guha D, Alotaibi NM, Ibrahim GM, Lipsman N, Fallah A: Readability and quality of Wikipedia pages on neurosurgical topics. *Clin Neurol Neurosurg* 166:66–70, 2018
26. Heilman JM et al.: Wikipedia: a key tool for global public health promotion. *J Med Internet Res* 13:e14, 2011
27. Thomas GR, Eng L, de Wolff JF, Grover SC: An evaluation of Wikipedia as a resource for patient education in nephrology. *Semin Dial* 26:159–163, 2013