

Automating Quality Control for Structured Standardized Radiology Reports Using Text Analysis

Anjani DHRANGADHARIYA^{a,1}, Sandy MILLIUS^a, Cyril THOULY^b, Benoit RIZK^b,
Dominique FOURNIER^b, Henning MÜLLER^{a,c}, and Hugues BRAT^b

^aUniversity of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland

^bInstitut de Radiologie de Sion (IRS), Groupe 3R, Sion, Switzerland

^cUniversity of Geneva, Geneva, Switzerland

Abstract. Radiology reports describe the findings of a radiologist in an imaging examination, produced for another clinician in order to answer to a clinical indication. Sometimes, the report does not fully answer the question asked, despite guidelines for the radiologist. In this article, a system that controls the quality of reports automatically is described. It notably maps the free text onto MeSH terms and checks if the anatomy and disease terms match in the indication and conclusion of a report. The agreement between manual checks of experienced radiologists and the system is high with automatic checks requiring only a fraction of time. Being able to quality control all reports has the potential to improve report quality and thus limit misunderstandings, losing time for requesting more information and possibly avoid medical mistakes.

Keywords. Radiology Reports, Quality Control, Text Analysis

1. Background

Radiologists routinely document, report and communicate diagnostic information to the referring clinicians through structured standardized reports (SSRs). Different from unstructured documents, a typical radiology report is divided into the following sections: type of examination, clinical indication, technique, findings and conclusion [1,2]. A referring clinician conveys any identified clinical question requesting a radiology examination. This clinical question is repeated in the *indication* section of a report and the aim of a radiologist is to clearly communicate the answer to this clinical question through the *conclusion* section [3]. Despite guidelines for structuring reports, several problems can prevent clear communication between a radiologist and a clinician. One of the most frequent problems found in reports is a difference between the content of the indication and the conclusion sections whereby a report either fails to answer the clinical question of the referring clinician or has an unclear answer [4]. For example, we identified a report where a clinician requested a MRI (Magnetic Resonance Imaging) scan upon indications of tendinopathy and bursitis

¹ Corresponding Author: E-mail: anjani.dhrangadhariya@hevs.ch.

(comorbidities). The *conclusion* section confirmed tendinopathy but failed to validate or contradict the clinicians’ question about bursitis.

Such human errors could be avoided through automated quality control (Q.C.) of reports. The work of this article describes a Java-based software prototype that employs data integration and semantic comparison to identify the reports with potential differences in reporting and informs the concerned radiologist by flagging them. We assess prototype performance by measuring overall concordance between the indication and conclusion sections using a data set of radiology reports by comparing the results of automated evaluation with human evaluation.

2. Methods

2.1 Data set used

A set of 200 randomly chosen, anonymized, French-language reports was obtained from Institut de Radiologie de Sion (IRS), a private radiology center located in the French speaking part of Switzerland. An individual report comprises the following sections: indications, description and conclusions (see Figure 1).

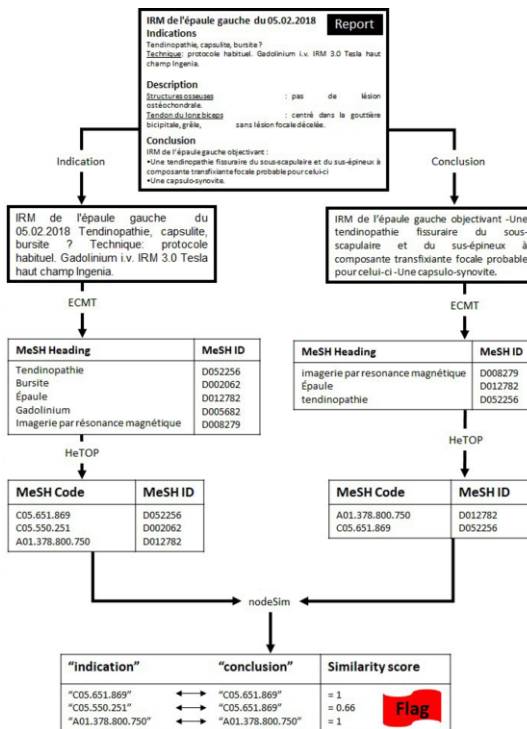


Figure 1. Picture showing the processes followed in the developed prototype.

The data set was not only used to assist prototype development but was also used to evaluate the prototype.

2.2 Medical Subject Headings

Medical Subject Headings (MeSH) are a hierarchically-organized terminology mainly used for organization and cataloguing of the biomedical information and literature. It is organized into a tree structure with its root node branching into 16 broad thematic categories like “Anatomy”, “Diseases”, “Chemical and Drugs”, etc. These categories are further divided into subcategories or MeSH headings. In this way, MeSH enforces uniformity and consistency across the terminology in a way that articles corresponding to a particular topic are indexed under a particular MeSH heading. For example., all the studies about *benign cancer* are indexed under MeSH heading *Neoplasm*². Each MeSH heading in the MeSH tree has a code indicating its location in the tree and a unique MeSH ID [5].

2.3 MeSHSim

MeSHSim is an R-package with “nodeSim” functionality that measures semantic similarity between any two MeSH codes as a function of their length and their position in the MeSH tree. The similarity values range between 0 and 1 [6].

2.4 HeTOP

HeTOP (Health Terminology/Ontology Portal)³ is a portal, housing 70 health terminologies in 32 languages including MeSH and Radiology Lexicon (RadLex). It comprises a semantically interoperable network of 2 million cross-lingual, health-domain related concepts. For example, querying the French term *radiologie* leads to its MeSH counterpart “Radiology” from other multi-lingual terminologies⁴. This rich multilingual content is available for research and can be programmatically accessed via authenticated Representational State Transfer (REST) or Simple Object Access Protocol (SOAP) requests [7].

2.5 ECMT

L’Extracteur de Concepts Multi-Terminologique (ECMT) combines a rule-based and an NLP (Natural Language Processing)-based approach to extract health-related concepts from French-language texts using French-language terminologies in HeTOP. ECMT service can be programmatically accessed via authenticated REST or SOAP requests [7,8].

3. Results

3.1 Prototype developed

The developed prototype fetches a non-empty, radiology report file from internal storage followed by rule-based extraction of free text from the indication and

² <https://meshb.nlm.nih.gov/record/ui?ui=D009369>

³ <https://www.hetop.eu/hetop/fr/?q=&home>

⁴ <https://www.hetop.eu/hetop/#rr=MSH D 011871&oti=all&q=Radiology>

conclusion sections (see Figure 1). An authenticated REST request to ECMT then extracts French-language health-related concepts from the free text. ECMT returns an XML (eXtensible Markup Language) file that is parsed to extract only the MeSH Headings and unique ID's corresponding to Anatomy [A] and Disease [C]. Then, an authenticated REST request to HeTOP returns an XML file with MeSH codes and paths corresponding for the requested MeSH headings. Next, weighted-link semantic similarity between individual anatomy and disease codes is calculated from indication and conclusion using the nodeSim functionality from MeSHSim. If every anatomy and disease code in indication has a similarity score of 0.85 or more with their counterparts in conclusion, the report is considered as

Q.C. passed, otherwise it is flagged. A threshold of 0.85 was chosen, as it pertains to path length three or less between MeSH codes being compared, based on experiments with a few validation cases. This threshold accounts for the conceptual similarity between the MeSH headings being compared. [6]. For example, a similarity score between MeSH anatomy terms “knee” and “toes” is 0.815, but between “toes” and “hallux” it is 0.948.

3.2. Prototype evaluation

To evaluate the performance of the prototype, overall percentage concordance between indication and conclusion for all the radiology reports in the data set was automatically assessed. Out of the 200 reports in the data set, a small subset were flagged by the prototype because the text corresponding to the *conclusion* section was missing in these reports. The *conclusion* text from these reports was deliberately removed for testing this scenario. Of the remaining reports, MeSH terms and codes for the anatomy and diseases were extracted from the indication and conclusion free-texts using ECMT and HeTOP. Only 93 and 143 reports respectively contained MeSH terms corresponding to anatomy and disease. Of these 143 reports, semantic similarity between *indication* and *conclusion* MeSH terms was obtained using nodeSim and total concordance was calculated. To establish a ground truth, two expert radiologists repeated the above procedure manually and calculated the total concordance.

Upon total concordance assessment between the indication and conclusion, the prototype reported 45% concordance for the anatomy information, while manual evaluation reported a 61% concordance. For disease information, the prototype reported 80% concordance versus 88% with the peer review. Based on ground truth, the algorithm accuracy for anatomy was 84% and for disease 92%. It took the experience radiologists about 1.5 hours to conduct manual quality checks, but the prototype did it in a matter of seconds.

4. Discussion and Conclusion

Reports with perfect match for the anatomy and disease MeSH terms between indication and conclusion correctly passed Q.C, multiple reports were flagged when a particular anatomy term (e.g. knee) was mentioned in *indication*, but was missing in the *conclusion*. The prototype success, however, is subject to ability of ECMT to accurately extract the concepts from free texts and also upon availability of a particular concept in the MeSH vocabulary. Currently, the prototype accounts only for the MeSH terminology, which lacks some radiology-specific terms [9]. For example, a report

mentioned *epicondyle* in *indications* section, but not in the *conclusion* section. This report passed Q.C. because no MeSH term corresponding to *epicondyle* was identified from indications. Indeed, this term was absent in MeSH, but present in RadLex⁵ and hence complementing the two can likely improve the results for the missed radiology terms. However, only a part of RadLex is currently available in French. In another instance, a disease term "Déchirure" from free text indication was not identified by the ECMT, but a standardized, English counterpart for "Déchirure" existed in MeSH as "Rupture". This instance was reported to the ECMT team and the term was later added and semantically linked to its MeSH counterpart. Overall prototype performance should also be improved with lexical standardization of disease terms in radiology reports, and further developments in ECMT algorithm.

In conclusion, the automated prototype showed good performance compared to the expert radiologists for assessment of indication and conclusion concordance in the radiology reports. It also reduces time and cost required for the quality control. The prototype additionally identified the lack of systematic reporting of relevant anatomy information in the conclusion sections, a fact that can be highlighted in guidelines for radiologists.

References

- [1] R. B. Gunderman and L. R. McNeive, "Is structured reporting the answer?," *Radiology* **273**(1), pp. 7–9, 2014. PMID:25247561.
- [2] A. P. Brady, "Radiology reporting-from hemingway to hal?," *Insights into Imaging* **9**(2), pp. 237–246, 2018.
- [3] A. Wallis and P. McCoubrie, "The radiology report—are we getting the message across?," *Clin. Radiol.* **66**, pp. 1015–1022, Nov. 2011.
- [4] T. Laswad, D. Fournier, J. Moreau, H. Brat, M. Deac, T. Tancredi, F. Bonvin, T. Zand, and P. Benedict, "Structured reports: 10 years experience in private radiology clinics," in *European Congress of Radiology 2014*, Springer International Publishing, (Vienna, Austria), 2014.
- [5] F. B. Rogers, "Medical subject headings," *Bull Med Libr Assoc* **51**, pp. 114–116, Jan 1963.
- [6] J. Zhou, Y. Shui, S. Peng, X. Li, H. Mamitsuka, and S. Zhu, "MeSHSim: An R/Bioconductor package for measuring semantic similarity over MeSH headings and MEDLINE documents," *J Bioinform Comput Biol* **13**, p. 1542002, Dec 2015.
- [7] A. Benis, M. Crisan-Vida, L. Stoicu-Tivadar, and S. Darmoni, "A Multi-Lingual dictionary for health informatics as an international cooperation pillar," *Stud. Health Technol. Inform.* **262**, pp. 31–34, 2019.
- [8] M. H. Metzger, Q. Gicquel, D. Rabarijaona, F. Segond, A. Bittar, L. Dini, L. F. Soualmia, and S. J. Darmoni, "A secondary use of the electronic health record for epidemiological studies and evaluation of health care quality: the SYNODOS project," *Journes Francophones d'Imagerie mdicale* **1379**, January 2014.
- [9] D. Demner-Fushman, S. E. Shooshan, L. Rodriguez, S. Antani, and G. R. Thoma, "Annotation of chest radiology reports for indexing and retrieval," in *Multimodal Retrieval in the Medical Domain*, pp. 99–111, Springer International Publishing, 2015.

⁵ <http://www.radlex.org/RID/RID38647>