

ontologies Radiological Lexicon (RadLex) [2] and System-integration of pro les published by Integration Healthcare
 atized Nomenclature of Medicine Clinical Terms (SNOMED Enterprise (IHE) [6]. Having a repository of pathology-
 CT) [3]. These ontologies provides a standardized, multi-proven cases in a dashboard also has the potential to enhance
 lingual vocabulary of clinical terminology that is used by and encourage the formation of accurate teaching les, as
 physicians and other health care providers for the electronic well as educational publications in the form of case series
 exchange of clinical health information. In our system wear “case of the day” submissions [7]. As the use of positron
 integrated radiology teaching le data sources and, while emission tomography computed tomography (PET-CT) has
 integrating these data sources, we performed experiments increased rapidly, there is a need to retrieve relevant medical
 to evaluate the accuracy of search results. This evaluation images that can assist an image interpretation. Building
 concluded that integration of medical ontologies is necessary database which may provide integrated repository with
 to improve the search quality (the results are discussed images to improve diagnosis accuracy [8]. Larger clinical
 in Section 4). Successful ongoing integration of medical reference datasets that are relevant to a larger number of
 ontologies demonstrates that data integration is a continuous patients may help to retrieve complex query results. (e.g.,
 process – integrating data sources (with teaching les) is retrieve the PET-CT study containing the lymph node
 not suf cient without also integrating the related metadata lesion, which showed no interval change for more than 2
 sources. We further incorporated support for content base years”). Data integration and a centralized data repository
 image retrieval (CBIR) and observed that searching medical for clinical data, patient history, physical exam ndings,
 image data sources enabled us to get better results. In this laboratory data, imaging data is important as a reference
 paper, we discuss the need of data integration for healthcare during the diagnostic process. Authors of [9] discussed how
 domain and how metadata further supports this process. big data analysis could be helpful for radiologist daily work.

In Section 2 we describe RSNA MIRC, MyPacs, RadLex From our survey we can conclude that in radiology there is
 and SNOMED CT ontologies, along with prior data in- a need to integrate clinical reports and images and develop
 tegration work in the radiology domain. In Section 3 we a uni ed reference database. The following is the list of
 focused on data integration methods used in this research sources and repositories we have evaluated in determining
 work. In Section 4 we discussed our current results to show what databases are currently available to radiologists. We
 how integration of data sources and medical ontologies cannot tend to integrate these sources as we have integrated
 help radiologists in the diagnostic process. We expect to RSNA MIRC, MyPacs, RadLex and SNOMED CT.
 speed up reference search for radiologists by providing them RSNA MIRC: It is a large repository with 2,500 teaching
 with an integrated teaching le database solution. Otherwise including the information about the history of patients,
 they may have to refer to different heterogeneous sources diagnosis, differential diagnosis, ndings, discussion as well
 making it dif cult to nd and retain information. Overall, as external references (journal articles). Radiological terms
 this case study shows that data (and metadata) integration are highlighted and linked to RadLex browser (see dis-
 improves the search accuracy and performance. In Section cussion about RadLex below). However, search is done
 5 we summarize the conclusions of this work and describe verbatim with no processing to interpret the goals (e.g.,
 planned future work. synonyms, negation). No image-based search is possible.

2. Related work

Our literature survey is based on articles from Journal of records based on anatomy, pathology, modality, age, gender,
 Radiology, Radiographics, Digital Imaging, IEEE and other etc. Limitations of this search engine include lack of con-
 established medical publication venues. We reviewed papers sideration for synonyms, negation, or image-based search.
 that discussed the need for big data integration of health RadLex [2] Radiology Lexicon term browser. RadLex is
 care systems. There are many papers that argue the need and ontological system that provides a comprehensive lexicon
 for big data utilization and disparate source integration to vocabulary for radiologists. RadLex browser was developed
 better serve the medical eld, which greatly inspired us to by RSNA 11.2 Td [loped ar
 proceed with building IRIS engine. Ron Gutmark [4] argued
 for building a system that reduces errors in radiological
 images using teaching le database. Easy-to-use computer
 teaching les are useful for training physicians, serve as
 a reference tool for experienced physicians and help them
 improve diagnostic accuracy. The work in [5] discussed how
 critical radiologic images are for diagnosis, teaching needs
 and research. They were particularly interested in using
 case-based radiologic teaching les for radiology teaching.
 Their proposed architecture was meant to be integrated
 with existing medical image databases (featured by MIRC
 interoperability), but it is not publicly available. Availabil-
 ity of a large and diverse set of clinical cases need the

gine of the National Library of Medicine enables search and retrieval of abstracts and images (e.g., charts, graphs, clinical images) from the open source literature and biomedical image collections. Searching may be done using text queries as well as query images. Open-i provides access to over 3.7 million images from about 1.2 million PubMed Central articles. Open-i is great source of image collection, however this data source does not include categories such as history or diagnosis information for the patient case.

EURORAD (European Society of Radiology) [12] is a peer-reviewed educational tool based on teaching cases. There are a more than 7,000 teaching cases – similar to other teaching le sources there is no support for negations, synonyms, or image-based search. There are many

Medicine (DICOM) images that provides image modality as a label, the remaining 90,000 images are JPEG or PNG format. The lack of labelled data encouraged us to integrate labelled modality images. We used ImageCLEF [16] dataset that provided 5,000 of modality labelled images, further demonstrating that data integration is an iterative process.

4. Results

In this section we present results from integration of additional data sources. Initially, we used a naive method for data integration (without our proposed logical schema or explicit integration). This method involved comparing the query term in each teaching le body (text) in the database; it was not only time consuming but also error prone. For example, for a “renal artery” query the naive approach would have to match both words in the text exactly; we could also search for individual words (“renal” and “artery”), but that would generate too many false-positives. Results discussed here use 5 sample queries to illustrate how integration of an additional ontology improved IRIS results. We compared our initial IRIS search (IRIS 1.0 with RadLex ontology) with new IRIS 1.1 (with RadLex and SNOMED CT ontologies). Table 1 shows that adding another ontology greatly improved search results. Search for “chiari” produced 153 results in IRIS 1.0; however, adding a second ontology improved results by 39 matches. After query expansion with “hindbrain hernia” and “arnold–chiari malformation” synonyms, the search resulted in 192 relevant teaching les. This search was able to nd so many matches by applying both ontologies. “Hindbrain hernia” is a synonym from RadLex ontology and which is not present in SNOMED