

# "Knowledge is Power": Constructing Knowledge Graph of Abdominal Organs and Using Them for Automatic Radiology Report Generation

Kaveri Kale<sup>1</sup>, Pushpak Bhattacharyya<sup>1</sup>, Aditya Shetty<sup>2</sup>, Milind Gune<sup>3</sup>,  
Kush Shrivastava<sup>3</sup>, Rustom Lawyer<sup>3</sup> and Spriha Biswas<sup>3</sup>

<sup>1</sup>Department of Computer Science and Engineering, IIT Bombay, India

<sup>2</sup>Breach Candy Hospital, India

<sup>3</sup>Augnito India Private Limited

{kaverikale,pb}@cse.iitb.ac.in,

adityashetty01@gmail.com, dgune@rediffmail.com, {kush.shrivastava, rustom, spriha}@augnito.ai

## Abstract

In conventional radiology practice, the radiologist dictates the diagnosis to the transcriptionist, who then prepares a preliminary formatted report referring to the notes, after which the radiologist reviews the report, corrects the errors, and signs off. This workflow is prone to delay and error. In this paper, we report our work on automatic radiology report generation from radiologists' dictation, which is in collaboration with a startup about to become Unicorn. A major contribution of our work is the set of knowledge graphs (KGs) of ten abdominal organs- *Liver, Kidney, Gallbladder, Uterus, Urinary bladder, Ovary, Pancreas, Prostate, Biliary Tree, and Bowel*. Our method for constructing these KGs relies on extracting entity1-relation-entity2 triplets from a large collection (about 10,000) of free-text radiology reports. The quality and coverage of the KGs are verified by two experienced radiologists (practicing for the last 30 years and 8 years, respectively). The dictation of the radiologist is automatically converted to what is called a *pathological description* which is the clinical description of the findings of the radiologist during ultrasonography (USG). Our knowledge-enhanced deep learning model improves the reported BLEU-3, ROUGE-L, METEOR, and CIDEr scores of the pathological description generation by 2%, 4%, 2% and 2% respectively. To the best of our knowledge, this is the first attempt at representing the abdominal organs in the form of knowledge graphs and utilising these graphs for the automatic generation of USG reports. A Minimum Viable Product (MVP) has been made available to the beta users, *i.e.*, radiologists of reputed hospitals, for testing and evaluation. Our solution guarantees report generation within 30 seconds of running a scan.

## 1 Introduction

Radiology is an integral part of medical care. Radiological imaging-based evidence (X-ray, MRI, CT,

USG, *etc.*) is crucial for determining the nature of the treatment. The usual radiology process is as follows: A patient gets scanned. Then the radiologist prepares the diagnosis notes (referred to as **radiologist's dictation**) and handing them over to a transcriptionist. The transcriptionist opens a scan-specific standardised template (referred to as **normal report template**) and edits it referring to the notes in a more descriptive form (referred to as **pathological description**).

Radiologists are in huge demand since the ratio of radiologists to patients is very low. These ratios in India, the US, and China are 1:100,000, 1:10000, and 1:14772, respectively (Arora, 2014). These low ratios results in a very high patient inflow per radiologist volume, making radiologists incredibly busy and stressed out. The currently adopted transcriptionist-based workflow causes (i) significant delays in report turnaround time, (ii) errors in the reports, and (iii) burnout. To automate the report generation process, domain knowledge is necessary. Domain knowledge can be acquired from already existing radiology free-text reports. We need a structured format for knowledge to be able to use it on a computer. Our research aims to use Natural Language Processing (NLP) (a) to construct abdominal-organ KG and (b) use these KGs for automatically generating radiology reports. Our work is in collaboration with a industry partner. On this project, two experienced radiologists are contributing their domain expertise to our work.

**Our contributions** are:

1. Knowledge graphs of ten abdominal organs- *Liver, Kidney, Gallbladder, Uterus, Urinary bladder, Ovary, Pancreas, Prostate, Biliary Tree, and Bowel*. We will release these constructed KGs and the code for KG construction from free-text reports, for wide use.
2. A radiology dictionary containing 43,304 en-

tries that are adapted from the Radlex-lexicon<sup>1</sup> and enriched with terminology from all forms of scans, *viz.*, USG, CT, MRI, and X-ray.

3. A generic methodology<sup>2</sup> to construct KGs from reports of all kinds of scans, *viz.*, CT, MRI, and X-ray.
4. A fine-tuned KG-BART that is fine-tuned on a parallel corpus of dictations and corresponding pathological descriptions.
5. A radiology report generation pipeline that identifies in a "normal" report the candidate text span for replacement and the patient-specific text that will replace the span.

## 2 Fundamental Definitions

Paulheim (2017) defines Knowledge Graph (KG) as "A knowledge graph (i) mainly describes real-world entities and their interrelations, organized in a graph, (ii) defines possible classes and relations of entities in a schema, (iii) allows for potentially interrelating arbitrary entities with each other and (iv) covers various topical domains." KGs are designed with suitable ontology to store domain knowledge. Ontologies are semantic data models that define the types of things in a specific domain and the properties used to describe those types. Ontologies does not include the details about specific **individuals** in domain. Three main components of ontology are **Classes**, **Relationships**, and **Attributes**. Domain ontology and individual information together form a Knowledge Base (KB). We have defined eight logical relations as follows:

1. **PartOf**: It represents the relation between anatomy and sub-anatomy. For example, right lobe is part of liver.
2. **TypeOf**: It represents the relation between similar type of entities. For example, *cystic lesion* is TypeOf *lesion*.
3. **ModifierOf**: It denotes the descriptors of findings, anatomical locations, property, *etc.* For example, *small* is descriptor of *size*.
4. **ObservationOf**: It denotes the clinical observations observed for particular finding. For example, *acute pancreatitis* denotes the presence of *fluid collection*.
5. **DefaultObservationOf**: It denotes the observation that associated by default with particular

<sup>1</sup><http://radlex.org/>

RadLex is licenced freely for commercial and non-commercial use.

<sup>2</sup>Our code to construct radiology KGs is located at <https://github.com/kaverikale/RadiologyKGConstruction>.

anatomical location or particular finding. For example, *peripancreatic fluid* is observation associated with *acute pancreatitis* by default.

6. **PropertyOf**: It denotes the relation between entities (anatomical entities, finding entities, observation entities, *etc.*) and their properties. For example, *echotexture* is property of the *liver*, *size* is the property of *lesion*, *shape* is the property of *kidney etc.*
7. **DefaultPropertyOf**: It denotes the property that exist by default with particular anatomical location or particular finding. For example, *shrunk size* is the property associated with *chronic pancreatitis*.
8. **FoundIn**: It denotes the relation between findings and corresponding anatomical location. For example, *lesion* found in *segment ii*.

## 3 Related Work

Research is done for automatic radiology report generation based on scanned images. Yuan et al. (Yuan et al., 2019) propose an automated structured-radiology report generation system using extracted features from images. Loveymi et al. (Loveymi et al., 2021) proposed a system that generates descriptions for natural images by image captioning.

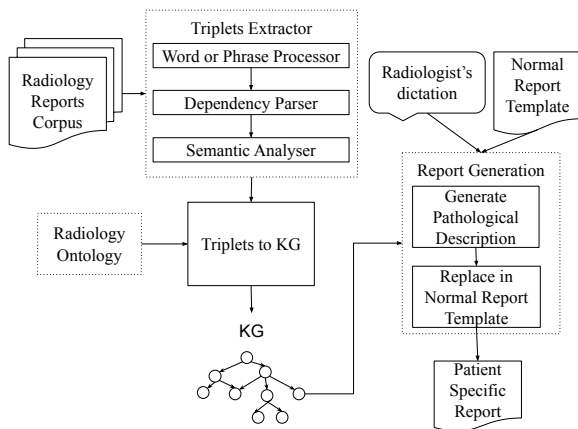
There is a wealth of research done on building medical KG from Electronic Medical Records (EMR). Finlayson et al. (Finlayson et al., 2014) builds a graph from medical text, clinical notes *etc.* Graph nodes represents diseases, drugs, procedures, and devices. Rotmensch et al. (Rotmensch et al., 2017) uses the EMR to construct the graph of diseases and symptoms. Researchers worked on creating medical KG from EMR, but no one has built a KG for the radiology domain except Zhang et al. (Zhang et al., 2020). Graph embedding module is proposed by Zhang et al. (Zhang et al., 2020) that helps to generate radiology reports from image reports. Each node in their KG represents disease. Taira et al. (Taira et al., 2001) developed an NLP pipeline to structure critical medical information. Extracted information includes the existence, location, properties, and diagnostic interpretation of findings from radiology free-text documents. Information is not integrated since they store the structured information for each report separately. Also, this system does not accept reports with different reporting styles. However, this is not always the case. Every radiologist has his or her own dictation

style and reporting style.

IE systems that are based on IE patterns are surveyed by Muslea et al. (Muslea et al., 1999). Ghoulam et al. (Ghoulam et al., 2015) extract signs of lung cancer, their anatomical location, and the relation between the signs and the locations expressed in the radiology reports. Embarek & Ferret (Embarek and Ferret, 2008) used a morpho-syntactic patterns in their rule-based method to find medical entities like symptoms, disease, exams, medicaments, and treatment. Xu et al. (Xu et al., 2009) explains that a pattern is a sub dependency tree that indicates a relation instance. Pons et al. (Pons et al., 2016) give an overview of NLP techniques that can be used in radiology.

## 4 Methodology

As shown in figure 1, we first construct the KGs for each abdominal organ from the ultrasound report corpus. Then we use these constructed KGs to generate ultrasound radiology reports from the radiologist’s dictation. KGs are constructed from anonymized radiology reports provided by our company collaborator. The anonymized report collection consists of approximately 10,000 reports of ultrasound scans.



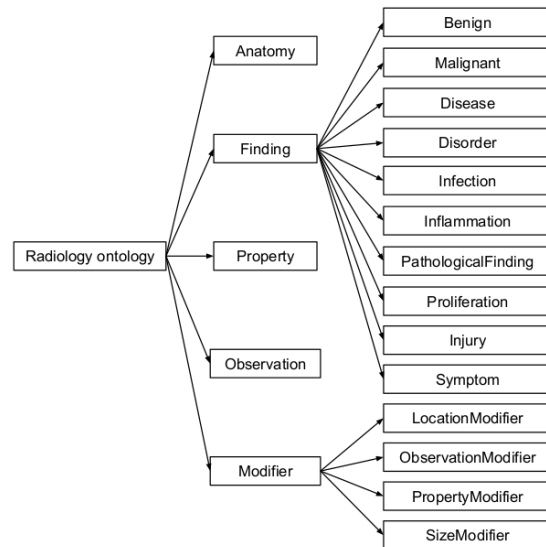
**Figure 1:** The architecture of our system. KG construction and patient-specific report generation are the two main modules in our system.

### 4.1 Ontology Creation

We refer to the RadLex lexicon to create our ontology, which we call "Radiology Ontology" (Langlotz, 2006). Though called a "lexicon," the RadLex is actually an ontology since it has a hierarchical structure. For example, "Solid Organ ← Lobular Organ ← Liver" is a part of the RadLex term

and concept hierarchy. We will use "RadLex Lexicon" to mean RadLex Ontology. The RadLex lexicon includes a total of 46,761 classes and 24,075 individuals. However, there are limitations in the structure. Classes are defined at a very fine-grain level. In our work, we do not need such fine granularity. For example, *liver* is defined as one of the classes in RadLex. Instead of treating it as a separate class, we can define it as an instance of the anatomy class.

We have created our *own* ontology by integrating several RadLex Lexicon class entities. We maintain a coarse level of granularity to ensure a generic ontology. We have defined 8 logical relations as follows: *PartOf*, *TypeOf*, *ModifierOf*, *ObservationOf*, *DefaultObservationOf*, *PropertyOf*, *DefaultPropertyOf*, and *FoundIn*. Definitions and examples of all these logical relations are given in **appendix 2**. Figure 2 shows the class hierarchy of radiology ontology.



**Figure 2:** The class hierarchy of radiology ontology that we have created.

### 4.2 Radiology Dictionary Creator:

Radiology reports contain a large number of medical terms like abbreviations, synonyms, and proper names. The RadLex lexicon is used to create the radiology dictionary. However, we cannot use radiology terms from the RadLex lexicon as it is because the RadLex lexicon contains long phrases, e.g., *fat homogeneous background echotexture* which are not at the right level of granularity for a knowledge graph. Also, there are some radiological terms that frequently appear in radiology reports but are

not present in the RadLex lexicon (e.g., *reflectivity* and *echopattern*). Entities that are missing in the Radlex lexicon have been added from the corpus. Table 1 shows the examples of entities and categories in our dictionary.

Entity	Category	Entity	Category
lesion	observation	small	size-modifier
cirrhosis	pathologic-finding	left lobe	anatomy
hepatitis	inflammation	ankle fracture	injury
size	property	chronic liver disease	disease

**Table 1:** Examples of entities and corresponding categories in our radiology dictionary.

### 4.3 Triplets Extraction

The Triplets-Extraction module extracts entities and relations. For example, for the corpus sentence *Right kidney is normal in size, shape, location and cortical echogenicity*, *Cortical* and *echogenicity* are linked by the relation *ModifierOf*. Also *normal* is a modifier of *size*, *shape*, *location*, and *echogenicity*. However, the state-of-the-art Open Information Extraction tools like OpenIE<sup>3</sup> are not capable of extracting these relations from free-text (Etzioni et al., 2008). Examples of triplets extracted using OpenIE are given in the table 9 in appendix A.3.

Our triplet extraction method combines the dictionary-match, rules and patterns to extract entities and relations. This methodology necessitates the use of (i) Word and Phrase Level Processor, (ii) Dependency Parser, and (iii) Semantic Analyser.

#### 4.3.1 Word and Phrase Level Processor

The input to this stage is a sentence from the corpus, and the output is the sentence-wise syntactic and semantic features of each word and phrase in the sentence. Features include POS tags, lemmas, supersenses, and the root of a noun chunk.

**Lexical Semantic Supersense Tagger:** To extract the relation between two entities connected by a preposition, the machine should understand the meaning of that preposition. As shown in the figure 3, the intuition of the *in* preposition in the first sentence is **characteristic** and in the second sentence is **locus**. Supersenses (Schneider et al., 2015) help get disambiguated senses of these prepositions. We have integrated the pre-written code of a Lexical Semantic Supersense Tagger (LSR)<sup>4</sup> (Liu

<sup>3</sup><https://nlp.stanford.edu/software/openie.html>

<sup>4</sup><https://github.com/nelson-liu/lexical-semantic-recognition>

et al., 2020) to assign supersense tags to prepositions. Figure 3 shows the sentences that contain the prepositions and their corresponding tags.

1. *The/O-DET liver/O-N-n.BODY is/O-V-v.stative normal/O-ADJ in/O-P-p.Characteristic size/O-N-n.ATTRIBUTE and/O-CCONJ echotexture/O-N-n.ATTRIBUTE./O-PUNCT*
2. *Non/B-ADV-!\_enhancing/!~ADJ hypodense/O-ADJ lesion/O-N-n.COGNITION noted/O-V-v.cognition in/O-P-p.Locus right/O-ADJ lobe/O-N-n.LOCATION of/O-P-p.Whole liver/O-N-n.BODY./O-PUNCT*

**Figure 3:** Lexical Semantic supersense Tagger tags all words with the supersense tags. Supersenses of prepositions are highlighted in red. We see *in* tagged with different supersenses.

We have mapped preposition supersenses to their corresponding logical relations. Table 2 shows the examples of supersense classes and their corresponding logical relations.

Supersense	Relation	Supersense	Relation
Locus	FoundIn	Whole	PartOf
Gestalt	PartOf	Manner	PropertyOf
PartPortion	PartOf	Purpose	PropertyOf

**Table 2:** Examples of supersenses and their corresponding logical relations.

Noun phrases are chunked to get the candidate entity phrases. Table 3 shows the combined output of the supersense tagger and the noun phrase chunker.

#### 4.3.2 Dependency Parser

The dependency parser links the entity phrases directly or indirectly. We write rules based on dependency and POS-tags to extract the relations between entities. Spacy<sup>5</sup> APIs are used for dependency parsing. Dependencies are established between phrases instead of words. An example of a dependency tree is given in figure 4.

#### 4.3.3 Entities and Relations Extractor

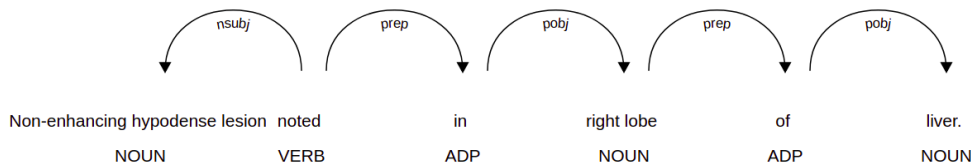
**Dictionary-matching-based Entity Extractor:** The noun chunker gives us noun phrases that are candidate entity phrases. However, not all noun phrases are radiological entities. Hence, to extract proper entities from noun phrases, we search the dictionary for matching entities. If a word or phrase matches multiple dictionary entries through more than one text span, we consider the longest text

<sup>5</sup><https://spacy.io/api/dependencyparser>



Noun Phrases/Words	Token List	Root Token	POS Tags	Lemmas	Supersenses
Non-enhancing hypodense lesion	[Non, -, enhancing, hypodense, lesion]	lesion	[ADJ, ADJ, VERB, ADJ, NOUN]	[non, , enhance, hypodense, lesion]	[B-ADV, L, I ADJ, O-ADJ, COGNITION]
noted	[noted]	noted	[VERB]	[note]	[cognition]
in	[in]	in	[ADP]	[in]	[Locus]
right lobe	[right, lobe]	lobe	[ADJ, NOUN]	[right, lobe]	[OAJ, LOCATION]
of	[of]	of	[ADP]	[of]	[Whole]
liver.	[liver, .]	liver	[NOUN, PUNCT]	[liver]	[BODY, OPUNCT]

**Table 3:** Output of the word and phrase level processing for the input *Non-enhancing hypodense lesion noted in right lobe of liver.*



**Figure 4:** Dependency tree of input sentence *Non-enhancing hypodense lesion noted in right lobe of liver.*

span as the matched entry for entity extraction. For example, in the phrase *right lobe*, although the individual terms *right* and *lobe* exist in our dictionaries, only the longest match, *right lobe*, is used for entity extraction.

**Pattern-based Relation Extractor:** A single noun phrase contains multiple entities. Table 4 shows the patterns to extract these entities. For example, consider the noun phrase *non-enhancing hypodense lesion*. As *non-enhancing* present in the dictionary, it applies the pattern *Modifier Observation* and extracts the triplet (*non-enhancing*, *ModifierOf*, *lesion*). *Hypodense* does not exist in our dictionary; hence, it applies the *ADJ NOUN* pattern and extracts a triplet (*hypodense*, *ModifierOf*, *lesion*).

**Relation Extraction Using Preposition Supersenses:** If two entities are connected with the preposition, then we consider its supersense to find the relation. For example, *lesion in right lobe*, here *in* represents the **locus** supersense and as shown in the table 2, the **locus** is mapped to the *FoundIn* relation. We add a new triplet (*lesion*, *FoundIn*, *right lobe*).

**Relation Extraction Between Different Noun Phrases:** We have discussed how to extract entities and relations between the entities present in the single noun phrase. However, a relation exists between the entities present in the two different noun phrases. In the example shown in the figure 4, there exists a relation between the *lesion* and *right lobe*. We have written rules over the dependency

tree to get the candidate pair of noun phrases. A list of patterns used to extract relations between two entities is listed in the table 5.

#### 4.3.4 Evaluation: Triplets Extraction Module

For each extracted triplet in a sentence, domain experts manually check whether it is correct or not. We calculate precision and recall for each sentence, then calculate the average precision, recall, and F1-Score. Figure 6 shows the evaluation results of our IE and OpenIE systems.

#### 4.4 From Extracted Triplets to the KG

Domain experts create preliminary KGs for each organ containing higher-level entities (i.e., basic hierarchical anatomy, merely 2-3 levels.), keeping the organ name as the root node, e.g., *liver* for the Liver KG. Figure 8 shows the liver preliminary KG. We enhance preliminary KGs by adding extracted triplets to them. The file contains sentence-wise triplets. KG augmentation algorithm steps are below: i) Create a hierarchical graph representation for each sentence's triplets. ii) Find matched pathways in our preliminary KG for each sentence graph path. iii) Add nodes and arcs that are missing in static KG. Figure 5 shows the augmented KG of the liver.

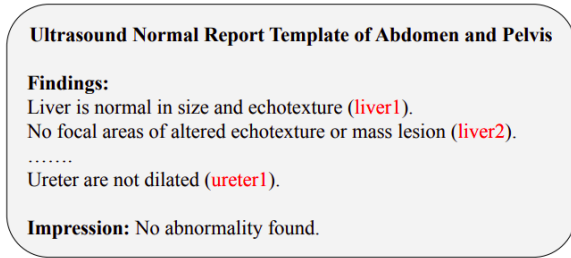
We build ontology using the Protégé<sup>6</sup> (Musen, 2015) the well-known terminology and ontology building and maintenance tool. Using transformation rules<sup>7</sup> we load augmented KG triplets as indi-

<sup>6</sup><https://protege.stanford.edu/>

<sup>7</sup><https://github.com/protegeproject/cellfie-plugin>

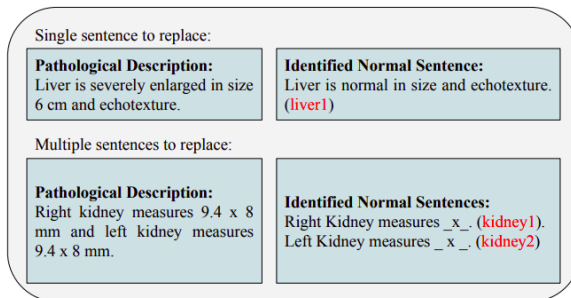


sentences. For example, we label normal sentence *Liver is normal in size and echotexture* as *liver1*. Figure 6 shows the normal report template with a label for each sentence mentioned in brackets. We create a dataset of pathological description sentences, each of which is annotated with corresponding normal sentence labels (i.e., *liver1*, *liver2*, etc.). The span identification problem can now be for-



**Figure 6:** Ultrasound normal report template with a unique label (highlighted in red) for each sentence.

mulated as a multilabel text classification problem. Given the pathological description, we need to decide which amongst the normal report sentences the description targets for replacement. A BERT-based multilabel text classifier takes the pathological description as input and gives multiple labels for the pathological description. Further implementation details are given in [appendix A.5](#). Figure 7 shows the examples of pathological descriptions and their corresponding normal sentences to replace.



**Figure 7:** Examples of pathological descriptions and corresponding normal sentences identified by span identifier. Labels are highlighted in red

**Replacement:** The candidate sentences returned by the BERT classifier are replaced by the pathological description in the normal report. If there are multiple candidates, replace the first sentence only and remove the other candidates. As shown in the second example of the figure 7, for a single sentence pathological description, there are two normal candidate sentences to replace. In that case,

we replace the first normal sentence and remove the second normal sentence.

## 5 Experiments

(Details about the training setup and implementation are in [appendix A.4](#)).

**Baseline and Evaluation:** We compare our fine-tuned KG-BART model with T5-base/large (Raffel et al., 2020) and BART-base/large (Lewis et al., 2019) state-of-the-art pre-trained text generation models. Gold standard pathological descriptions extracted from reports and verified by radiologists are used. Table 7 shows the BLEU (Papineni et al., 2002), ROUGE (Lin, 2004) and METEOR (Banerjee and Lavie, 2005) scores of generated pathological descriptions by KG-BART, T5-base/large and BART-base/large models.

Method	Automatic Evaluation Metrics				
	Bleu-1	Bleu-3	Rouge-L	Meteor	CIDEr
T5-large	0.873	0.780	0.897	0.902	0.892
BART-large	<u>0.887</u>	<u>0.798</u>	<u>0.910</u>	<u>0.916</u>	<u>0.908</u>
KG-BART	<b>0.901</b>	<b>0.830</b>	<b>0.930</b>	<b>0.927</b>	<b>0.928</b>

**Table 7:** The BLEU, ROUGE, METEOR and CIDEr scores of the generated pathological description (best results: bold, second best: underlined).

## 6 Summary, Conclusion and Future Work

We have given a systematic method to construct organ-wise KGs from free-text radiology reports. The KGs are stored in standard RDF format, enabling their application to various medical applications. One such example is the generation of radiology reports, which we have described here. Our KG-enhanced deep learning model improves the reported BLEU-3, ROUGE-L, METEOR, and CIDEr scores of the pathological description generation by 2%, 4%, 2% and 2% respectively. Our approach is generalized for other organs and scanned procedures, as evidenced in the EACL paper<sup>8</sup>.

To the best of our knowledge, this is the first attempt at automatic ultrasound report generation. An MVP (minimum viable product) has been made available to the beta users (practicing radiologists) for testing and evaluation. We are continuously collecting feedback on our system from radiologists and continually refining the tool. We have observed that it can generate a report within 30 seconds of running a scan.

<sup>8</sup><https://aclanthology.org/2023.eacl-main.246/>

## Ethics Statement

Anonymized radiology reports are used to build KGs. The data itself is anonymized; hence, our system does not reveal any patient-specific identity. ML technologies for this kind of work have the potential to gradually become the norm but will always remain as assistive tools for medical practitioners. Hence, while ML technologies of this kind often veer towards the norm, the envisioned assistive nature of this technology, where humans will always have oversight, will address this issue. We have evaluated the outputs of the KG-BART model using automated metrics, but to contextualize the results, a human evaluation metric would have been useful; however, we left this work for the future. An MVP (minimum viable product) has been made available to the beta users (practicing radiologists) for testing and evaluation. Manual evaluation will be done by considering beta users feedback.

## Acknowledgements

## References

- Richa Arora. 2014. [The training and practice of radiology in india: current trends](#). *Quantitative imaging in medicine and surgery*, 4(6):449.
- Satanjeev Banerjee and Alon Lavie. 2005. [METEOR: An automatic metric for MT evaluation with improved correlation with human judgments](#). In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor, Michigan. Association for Computational Linguistics.
- Mehdi Embarek and Olivier Ferret. 2008. [Learning patterns for building resources about semantic relations in the medical domain](#). In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*.
- Oren Etzioni, Michele Banko, Stephen Soderland, and Daniel S Weld. 2008. [Open information extraction from the web](#). *Communications of the ACM*, 51(12):68–74.
- Samuel G Finlayson, Paea LePendou, and Nigam H Shah. 2014. [Building the graph of medicine from millions of clinical narratives](#). *Scientific data*, 1(1):1–9.
- Aicha Ghoulam, Fatiha Barigou, and Ghalem Belalem. 2015. [Information extraction in the medical domain](#). *Journal of Information Technology Research (JITR)*, 8(2):1–15.
- Curtis P Langlotz. 2006. [Radlex: a new method for indexing online educational materials](#).
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. [Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension](#). *arXiv preprint arXiv:1910.13461*.
- Chin-Yew Lin. 2004. [Rouge: A package for automatic evaluation of summaries](#). In *Text summarization branches out*, pages 74–81.
- Nelson F Liu, Daniel Hershcovich, Michael Kranzlein, and Nathan Schneider. 2020. [Lexical semantic recognition](#). *arXiv preprint arXiv:2004.15008*.
- Ye Liu, Yao Wan, Lifang He, Hao Peng, and S Yu Philip. 2021. [Kg-bart: Knowledge graph-augmented bart for generative commonsense reasoning](#). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6418–6425.
- Samira Loveymi, Mir Hossein Dezfoulian, and Muharram Mansoorzadeh. 2021. [Automatic generation of structured radiology reports for volumetric computed tomography images using question-specific deep feature extraction and learning](#). *Journal of Medical Signals and Sensors*, 11(3):194.
- Mark A Musen. 2015. [The protégé project: a look back and a look forward](#). *AI matters*, 1(4):4–12.
- Ion Muslea et al. 1999. [Extraction patterns for information extraction tasks: A survey](#). In *The AAAI-99 workshop on machine learning for information extraction*, volume 2. Orlando Florida.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. [Bleu: a method for automatic evaluation of machine translation](#). In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.
- Heiko Paulheim. 2017. [Knowledge graph refinement: A survey of approaches and evaluation methods](#). *Semantic web*, 8(3):489–508.
- Ewoud Pons, Loes MM Braun, MG Myriam Hunink, and Jan A Kors. 2016. [Natural language processing in radiology: a systematic review](#). *Radiology*, 279(2):329–343.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. [Exploring the limits of transfer learning with a unified text-to-text transformer](#). *J. Mach. Learn. Res.*, 21(140):1–67.
- Maya Rotmensch, Yoni Halpern, Abdulhakim Tlimat, Steven Horng, and David Sontag. 2017. [Learning a health knowledge graph from electronic medical records](#). *Scientific reports*, 7(1):1–11.
- Nathan Schneider, Vivek Srikumar, Jena D Hwang, and Martha Palmer. 2015. [A hierarchy with, of, and for preposition supersenses](#). In *Proceedings of The 9th Linguistic Annotation Workshop*, pages 112–123.



Ricky K Taira, Stephen G Soderland, and Rex M Jakobovits. 2001. [Automatic structuring of radiology free-text reports](#). *Radiographics*, 21(1):237–245.

Hongzhi Xu, Changjian Hu, and Guoyang Shen. 2009. [Discovery of dependency tree patterns for relation extraction](#). In *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation, Volume 2*, pages 851–858.

Jianbo Yuan, Haofu Liao, Rui Luo, and Jiebo Luo. 2019. [Automatic radiology report generation based on multi-view image fusion and medical concept enrichment](#). In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 721–729. Springer.

Yixiao Zhang, Xiaosong Wang, Ziyue Xu, Qihang Yu, Alan Yuille, and Daguang Xu. 2020. [When radiology report generation meets knowledge graph](#). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12910–12917.

## A Appendices

### A.1 Data Preprocessing

To construct KGs, we use the information from the free-text radiology reports. The data preprocessor module takes radiology reports as input. Radiology reports contain Header, Findings, History, and Conclusion/Impression sections. We use simple heuristics like regular expressions to fetch only the Findings and Impression section. Furthermore, we use regular expressions to separate the organ-wise sentences in different datasets. We use SymSpell<sup>9</sup> APIs to correct spelling mistakes, and word-tokenization since the extracted sentences contain spelling mistakes, unwanted punctuation marks, etc. Table 8 shows the samples from the corpus, which we extract from sample reports. In the corpus, there are a lot of extra spaces and unwanted punctuation marks found. We have removed these unwanted characters from the corpus using regular expressions.

For example, *Liver is enlarged in size(16.45cm)& normal in shape and shows raised echo reflectivity. No focal or diffuse lesion is seen. The portal and hepatic veins are normal. In the above example, there is no space between size, (16.45cm) and &. Also, there is no space between . and No and therefore sentence tokenization is challenging. Liver is enlarged in size ( 16.45 cm ) & normal in shape and shows raised echo reflectivity. No focal or diffuse lesion is seen. The portal and hepatic*

<sup>9</sup><https://github.com/wolfgarbe/symspell>

*veins are normal.* The text is then further divided into sentences.

#### A.1.1 Spelling Correction

In corpus, there are a lot of spelling mistakes also. To correct the spellings we have used the SymSpell library.

**Single Word Spelling Correction** We have created unigram and bigram dictionaries for corpus text.

**Unigram Dictionary:** Dictionary of unique correct spelling words, and the frequency count for each word.

**Bigram Dictionary:** Dictionary of the unique correct spelling of a pair of words, and the frequency count for each pair.

Levenshtein algorithm is used to compute edit distance metric between two strings. Edit distance algorithm finds the correct suggestion for words in input text with words in unigram dictionary.

For example, *enlaregd, billiary, radicals* are the incorrect words found in the corpus. In dictionary *enlarged, biliary, radicals* these correct words are present. Edit distance algorithm suggests *enlarged* word for *enlaregd*. Similarly *biliary* for *billiary* and *radicles* for *radicals*.

#### Multi-word Spelling Correction

- We remove mistakenly inserted spaces within a correct word

**Input:** *Liver is normal in size and reveals diffuse hypo attenuation*

**Output:** *Liver is normal in size and reveals diffuse hypoattenuation*

- We add mistakenly omitted spaces between two correct words

**Input:** *Liver appears normal in size and reveals mild generalized increasedparenchymal reflectivity.*

**Output:** *Liver appears normal in size and reveals mild generalized increased parenchymal reflectivity.*

Table 8 shows the organwise samples from text corpus after data preprocessing.

### A.2 Knowledge Graph Augmentation

Preliminary KGs enhanced using triplets extracted by IE module. Steps involved in KG augmentation are explained below:

---

## Samples from organ-wise corpus

---

### Liver

the liver is normal in size and moderate diffuse increase in hepatic echogenicity. no focal lesion is seen. intra-hepatic biliary radicals are not dilated. portal vein is normal. grade ii fatty liver. liver is normal in size and echotexture show no focal areas of altered echotexture or mass lesion. no intra-hepatic biliary radicals dilatation seen. Portal vein appears normal.

### Uterus

uterus is anteverted and normal in size and echotexture. the endometrial echo is midline and regular. endometrial thickness is dim mm. uterus is normal in size of normal echotexture. it measures about 6.7 x 3.1 x 4.5 cms in size. no focal mass seen. endometrial echoes are normal.

### Spleen

spleen is normal in size and echotexture. spleen and Pancreas are normal in size and echotexture. no focal mass lesion noted.

### Kidney

right kidney measures 10.2 cm in long axis. left kidney measures 10.3 cm in long axis. both kidneys are normal in size and echotexture. no evidence of hydronephrosis or calculi seen. both the Kidneys are normal in size and echotexture. corticomedullary differentiation well seen. no hydronephrosis or stones seen. right kidney measures about 11.2 x 4.0 cm. parenchymal thickness is 1.4 cms. left kidney measures about 10.5 x 4.5 cm. parenchymal thickness is 1.3 cms.

### Ovary

right ovary measures 1.2 x 2.3 cm. left ovary measures 2.2 x 1.2 cm. both the ovaries are normal. no adnexal mass noted. right ovary measures 2.1 x 1.4 cm. left ovary measures 2.0 x 1.3 cm.

### Pancreas

pancreas is normal in size and echotexture. spleen and Pancreas are normal in size and echotexture. No focal mass lesion noted.

### Gallbladder

the gallbladder is adequately distended. gallbladder calculus noted measuring 7 mm with no pericholecystic fluid. wall thickness appears normal. the cbd is normal. gallstone with no pericholecystic fluid. gallbladder shows normal distention, no evidence of stones seen or wall.

### Urinary bladder

urinary bladder is normal with no abnormal internal echoes and wall of normal thickness. urinary bladder appears normal. no stone, mass or wall thickening noted.

### Adnexa

no adnexal abnormality seen.

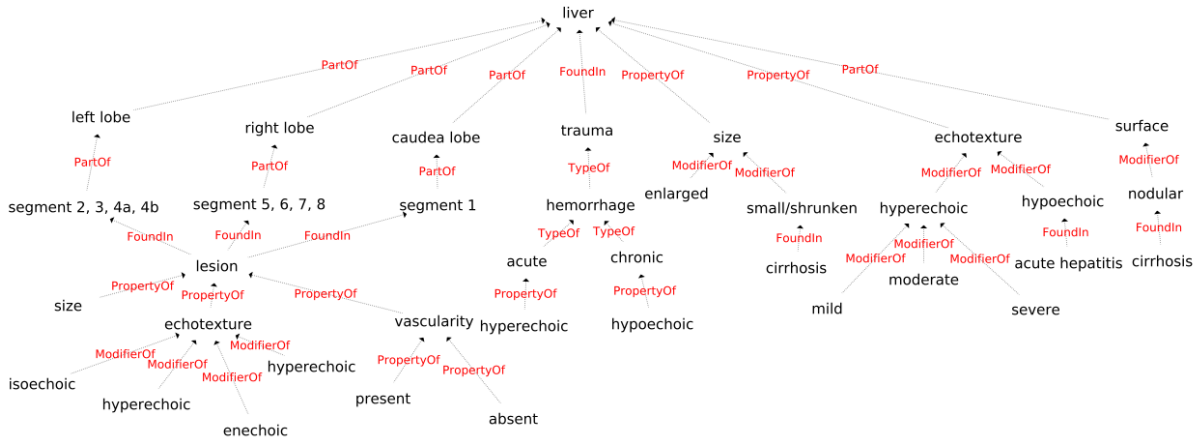
---

**Table 8:** Samples from the corpus following data preprocessing (all text is lower cased, removed unwanted characters, corrected spellings, *etc.*). Samples for all organs are not listed here. For example, samples for prostate, bowel *etc.* are not listed here.

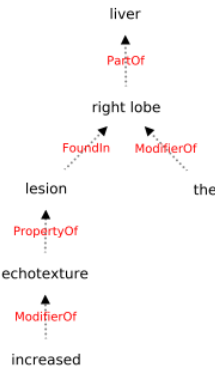
- **Step 1:** Triplets are stored in the file against its input sentence. For example, sentence from corpus is, *A lesion of increased echotexture in the right lobe of liver.* Triplets extracted corresponding to above sentence are, (*increased, ModifierOf, echotexture*), (*echotexture, PropertyOf, lesion*), (*right lobe, PartOf, liver*), and (*lesion, FoundIn, right lobe*).
- **Step 2:** Construct dynamic KG for sentence triplets. [Figure 9](#) shows the dynamic KG constructed for sentence triplets.
- **Step 3:** Find its appropriate matched path in our already built preliminary (static) KG. [Figure 10](#) shows the entities from dynamic KG path matched with static KG path.
- **Step 4:** If a triple is missing in the static KG, then we add a new triple in the static KG. Here in above example triple (*increased, ModifierOf, echotexture*) is missing in static KG. Hence, we will add this triple in static KG. [Figure 11](#) shows the updated static KG.

This is how we update the static KG according to our dynamic KG triplets. We repeat above steps for all sentences in our corpus.

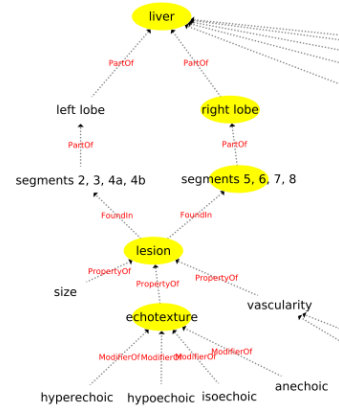
In a static KG, we have multiple instances of the same observations, same properties, and same modifiers. For example, *acute hepatitis* reveals *decreased echogenicity* of the *liver* and *chronic liver disease* reveals *increased echogenicity* of the *liver*. Here for both the findings *echogenicity* is the related observation but their related descriptors/modifiers are not same. *Decreased* is the *echogenicity* modifier associated with *acute hepatitis* and *increased* is the *echogenicity* modifier associated with *chronic liver disease*. Therefore we have created different instances of observation with name *echogenicity* for both the findings. Hence, we use a path from the dynamic KG to find the appropriate entity with identical names from the static KG. In static KGs, we have arranged findings in such a way that its parents represent the anatomical location and its children represent the properties or states of organs related to that finding. [Figure 5](#) shows the augmented KG of the Liver.



**Figure 8:** The augmented KG of the Liver is completed by information extracted from the radiology report corpus. Newly added nodes are highlighted in yellow. This figure shows a partial KG since it is large and cannot represent it in limited page size.



**Figure 9:** Dynamic KG constructed for the triplets of sentence *A lesion of increased echotexture in the right lobe of liver.*



**Figure 10:** The static KG at that instance. Yellow highlighted nodes show the entities from the dynamic KG path matched with the static KG path.

We rely on Protégé<sup>10</sup> (Musen, 2015) the well-known terminology and ontology building and maintenance tool. Extracted triplets are submitted in a CSV file along with classes of entities and relation types. The triplets are then ingested by Protégé to create the KG. We have created a rule-based system to find the class of each entity from our constructed dictionary. For example, given the triplet *lesion-FoundIn-liver*, the class-name for *lesion* is *Finding* and that for *liver* is *Anatomy*. The following are the stages for all triplets (*entity1*, *relation*, *entity2*): i) Protégé creates instances of corresponding classes for *entity1* and *entity2*. ii) Protégé adds the relation between two entities. Figure 5 shows the KG of the liver.

<sup>10</sup><https://protege.stanford.edu/>

### A.3 Datasets and Examples

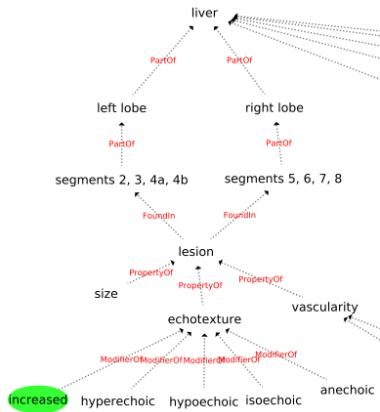
Table 9 shows the triplets extracted by OpenIE tool for given input sentences. As shown in the table 9, for the sentence 1, OpenIE could not find the relation between *calculus* and *middle calyx*, *middle calyx* and *right kidney* and for the sentence 2, OpenIE does not consider the *shape*, *location*, and *cortical echogenicity*.

### A.4 Training Details: KG-BART

Table 10 shows the samples from the parallel (impression and findings) dataset.

We have implemented our own algorithm for KG-grounding task. We use pre-trained KG-BART<sup>11</sup> model which was trained for common-sense reasoning on ConceptNet KG and common-sense dataset. We fine tune this model on radiology

<sup>11</sup><https://github.com/yeliu918/KG-BART>



**Figure 11:** Updated static KG after adding new triple. Newly added node highlighted in green.

Input Sentence	Triplets Extracted Using OpenIE
A 5 mm calculus is noted in an upper calyx and a 4 x 3 mm calculus is noted in a middle calyx of right kidney.	(mm calculus, is, noted in, upper calyx) (mm calculus, is, noted) (5 mm calculus, is, noted in, calyx) (5 mm calculus, is, noted in, upper calyx) (5 mm calculus, is, noted) (mm calculus, is, noted in, calyx)
Right kidney is normal in size 9.6 x 4.0 cm, shape, location and cortical echogenicity.	(kidney, is, normal) (right kidney, is, normal) (right kidney, is, normal in, size) (kidney, is, normal in, size)

**Table 9:** Examples of triplets extracted using OpenIE tool for given input sentences

text dataset that we have constructed. We use byte-pair encoding for tokenization with a maximum length of 32 for the encoder and 64 for the decoder. We set learning rate to 0.00001 and used AdamW with  $1 = 0.9$ ,  $2 = 0.98$  for optimization. We set the batch size to 32. We trained the KG-BART for 15 epochs, and the gradients are accumulated every 6 steps. We apply dropout with a probability 0.1 to avoid over-fitting. We use beam search with beam size 5 and length penalty with factor 0.6 while inferencing. The training time took 7 hrs on a single NVIDIA GeForce GTX 1080 Ti GPU with 11 GB GDDR5X memory.

### A.5 Implementation Details of Span Identifier

We use a BERT-based multilabel text classifier to identify the normal sentences. The last layer uses a sigmoid activation function to generate the probability of a sample belonging to the corresponding class. We used pretrained BERT weights to initialize our model. There are a total of 24 labels, according to the number of nodes in the last layer change. We train all the models on a DGX

A100-SXM-80GB GPU server. For all transformer based models we use hugging face transformer libraries.<sup>12</sup>

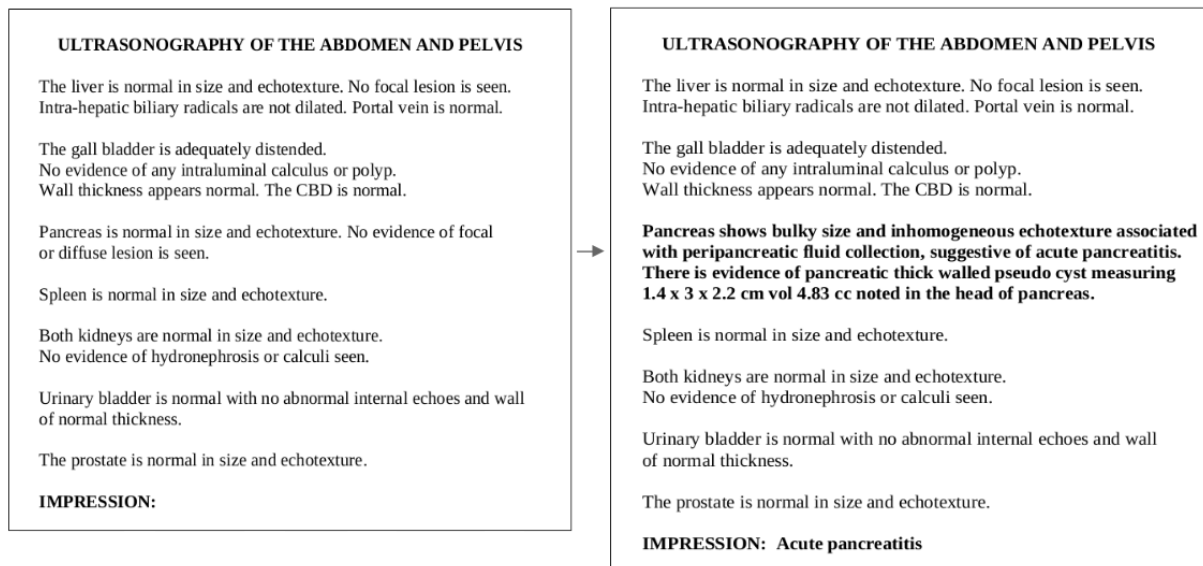
### A.6 Examples of Constructed KGs for Abdominal Organs

### A.7 Example of Normal Report and Patient Specific Report

<sup>12</sup><https://huggingface.co/docs/transformers/index>







**Figure 14:** Left hand side shows normal report template of ultrasonography of the Abdomen and Pelvis, and the right hand side shows a patient-specific report of ultrasonography of the Abdomen and Pelvis.