Automated Semantic Indexing of Figure Captions to Improve Radiology Image Retrieval

CHARLES E. KAHN, JR., MD, MS, DANIEL L. RUBIN, MD, MS

Abstract
Objective: We explored automated concept-based indexing of unstructured figure captions to improve retrieval of images from radiology journals.

Design: The MetaMap Transfer program (MMTx) was used to map the text of 84,846 figure captions from 9,004 peer-reviewed, English-language articles to concepts in three controlled vocabularies from the UMLS Metathesaurus, version 2006AA. Sampling procedures were used to estimate the standard information-retrieval metrics of precision and recall, and to evaluate the degree to which concept-based retrieval improved image retrieval.

Measurements: Precision was estimated based on a sample of 250 concepts. Recall was estimated based on a sample of 40 concepts. The authors measured the impact of concept-based retrieval to improve upon keyword-based retrieval in a random sample of 10,000 search queries issued by users of a radiology image search engine.

Results: Estimated precision was 0.897 (95% confidence interval, 0.857–0.937). Estimated recall was 0.930 (95% confidence interval, 0.838–1.000). In 5,535 of 10,000 search queries (55%), concept-based retrieval found results not identified by simple keyword matching; in 2,086 searches (21%), more than 75% of the results were found by concept-based search alone.

Conclusion: Concept-based indexing of radiology journal figure captions achieved very high precision and recall, and significantly improved image retrieval.


Introduction
Images published in peer-reviewed journals provide valuable information for education and clinical decision support. Retrieval of images based on their visual properties and textual captions is an area of active research. The articles in which the figures appear are indexed by Medical Subject Heading® (MeSH®) terms (U.S. National Library of Medicine, Washington, DC), which enables users to find articles using medical concepts. Although MeSH-based search can help find journal articles, it is not well suited to the task of finding particular images in those articles. Such images generally have an associated figure caption. The caption’s text provides more granular information, which can allow more robust search and retrieval of images. Searching the text within figure captions is plagued by the same challenges that are encountered when searching other clinical free-text information, such as radiology reports. We evaluated the impact of semantic indexing—the mapping of unstructured text to controlled terms—to improve retrieval of radiological images from journal articles.

Background
Semantic, or concept-based, indexing allows users to search for information using medical concepts. For example, concept-based searches recognize abbreviations, synonyms, and lexical variants. Most importantly, concept-based retrieval systems recognize subtypes of specific terms; for example, such systems understand that Parosteal Osteosarcoma is a type of Osteogenic Sarcoma, which is in turn a type of Bone Tumor. These systems require a robust model of medical knowledge to understand medical concepts and their interrelationships. Purely text-based retrieval systems are challenged by abbreviations and lexical variants, which stimulated our strategy to employ concept-based indexing.

To facilitate concept-based retrieval of images in articles, one could index the images using concepts extracted from the associated captions. However, it would be extremely laborious to perform this task manually. We explored an automated technique to map the unstructured (“free”) text of figure captions to concepts in a set of controlled vocabularies. Methods such as those described in this report can enable the radiology community to access more effectively the vast amounts of radiological image data being published online.

Several approaches have been explored for concept-based indexing of unstructured biomedical text. Systems such as
MicroMeSH, CHARTLINE, CLARIT, SAPHIRE, Meta-
phrase, and work by Nadkarni, et al have been applied in
a variety of applications to map unstructured text to the
MeSH vocabulary and/or the UMLS Metathesaurus. The
MetaMap program offers a linguistically rigorous con-
cept-discovery approach, and a version of the software can
be obtained without cost. To improve retrieval of radiology
images from the biomedical literature, we explored the use
of MetaMap to index the text of radiology figure captions.

**Methods**

This work had two specific aims: (1) to evaluate the ability of
a concept-mapping algorithm to correctly map free-text
radiology figure captions to controlled vocabulary concepts,
and (2) to measure the impact of concept-based searching on
the performance of an image search engine. First, we used a
concept-mapping algorithm to discover controlled-vocabu-
lar terms in a collection of radiology figure captions and to
index the captions accordingly. We applied standard informa-
tion-retrieval performance metrics to measure the effectiveness
of our semantic indexing process. Finally, we examined the
effects of concept-based retrieval on real-life queries to a
popular image search engine that uses this indexing ap-
proach. This investigation involved only analysis of infor-
mation in the published literature, and did not involve any
human subjects or protected health information; therefore,
this study was exempt from Institutional Review Board
review.

**Source Vocabularies**

The Unified Medical Language System (UMLS®) Metathe-
saurus, licensed from the U.S. National Library of Medicine,
served as the knowledge model for the image retrieval system.
The Metathesaurus is a very large database of biomedical and
health-related concepts, their various names, and the relation-
ships among them. It is built from the electronic versions
of many different source vocabularies, such as classification
schemes, thesauri, and lists of controlled terms used in patient
care, health services billing, biomedical research, public
health statistics, and biomedical literature indexing. To
index the text corpus used in this study, we employed three
source vocabularies from the UMLS, version 2006AA: Sys-
tematized Nomenclature of Medicine Clinical Terminology®
(SNOMED-CT®; International Health Terminology
Standards Development Organisation, Copenhagen, Den-
mark), the Foundational Model of Anatomy, and
the MeSH vocabulary, as these are the dominant sources for
terms relevant to our corpus. The aggregate vocabulary
consisted of 1,735,102 terms representing 662,736 distinct
concepts.

**Concept Mapping Algorithm**

We implemented the National Library of Medicine’s
MetaMap Transfer (MMTx) program to discover Metathe-
saurus concepts in unstructured (free-text) figure captions.
MMTx employs a series of language-processing modules to
map text to concepts in the UMLS Metathesaurus. MMTx
first parses text into components, including sentences,
paragraphs, phrases, lexical elements, and tokens.
Variants are generated from the resulting phrases. Candi-
date concepts from the UMLS Metathesaurus are retrieved
and evaluated against the phrases. The best of the candi-
dates are subsequently organized into a final mapping in
such a way as to best cover the text. We employed MMTx’s
“strict” model of the UMLS Metathesaurus, version 2006AA.
The strict filtering option limits the search to terms that are
supported by both the MetaMap and PubMed Related
Citations indexing methods. This approach tends to give a
small list of very good candidate controlled terms, but may
filter out some good recommendations as well.

**Experimental Dataset**

The ARRS GoldMiner® system (http://goldminer.arrs.org; Amer-
ican Roentgen Ray Society, Leesburg, VA) is a widely used
image search engine that is freely available via the Internet.
Goldminer uses both concept- and keyword-based search
techniques to retrieve images from a large number of
open-access, peer-reviewed journals. To build the experi-
mental dataset, we extracted 84,846 figure captions from the
Goldminer database. The figure captions, derived from
GoldMiner’s initial set of images, were acquired from 9,004
articles published online from 1999 to 2006 in five peer-
reviewed, English-language radiology journals: American
Journal of Roentgenology (AJR), American Journal of Neuroradi-
ology, British Journal of Radiology, RadioGraphics, and Radiol-
ogy. All the articles from which the figures and captions are
derived were available for open access. We created auto-
mated pattern-matching modules to remove hypertext mark
up language (HTML) tags from the figure captions so that
we could build a corpus containing only the text from the
captions.

**Information Retrieval Metrics**

To assess the performance of our concept-mapping ap-
proach, we sought to evaluate the standard information-
retrieval metrics of precision and recall (Fig 1). The
Reference Standard is a Boolean value that indicates, based
upon manual review, whether the specified concept is

<table>
<thead>
<tr>
<th>Indexed</th>
<th>Reference Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>TP, FP</td>
</tr>
<tr>
<td>−</td>
<td>FN, TN</td>
</tr>
</tbody>
</table>

TP = true positive
FP = false positive
FN = false negative
TN = true negative

Precision = TP / (TP + FP)
Recall = TP / (TP + FN)

**Figure 1.** Contingency table, with variables used to com-
pute precision and recall. “Reference Standard” indicates
that a concept is present (+) or absent (−) in a figure caption,
as determined by manual review. “Indexed” indicates
whether the concept has been identified in the figure caption
by the algorithm.
present in the figure caption. The Indexed variable indicates whether MMTx identified the concept in the figure caption. To compute precision and recall exactly, for each possible pairing of concepts and captions one must compare whether the concept is truly present in the caption versus whether the algorithm has assigned it as present. However, for sets of \( c \) concepts and \( f \) figure captions, the cross-product—the set of all concept-caption pairs—has \( c \times f \) elements. Here, the number of concept-caption pairs is 662,736 \( \times 84,846 \), which exceeds 56 billion. Thus, we applied sampling strategies to estimate the precision and recall of the indexing technique.

We used both “microaveraging” and “macroaveraging” to estimate these metrics. Microaveraging considers all concept-caption pairs as a single group. Macroaveraging computes the effectiveness measure separately for the set of captions associated with each concept, and then computes the mean of the results values. Macroaveraging is generally favored because it gives equal weight to each user query.\(^{23}\)

**Reference Standard**

To establish a reference standard, one of the authors (CEK) served as reviewer. The reviewer was presented sequentially with paired figure captions and concepts. For each concept-caption pair, the reviewer viewed the complete free-text figure caption, the UMLS concept unique identifier (CUI), and list of terms for that concept. The reviewer indicated whether the concept was present in the figure caption’s text. To eliminate potential bias, the sequence of caption-concept pairs was randomized; the reviewer was blinded as whether one was determining if the concept might be present or absent within the figure caption.

**Precision**

Precision measures the fraction of retrieved documents that are relevant to a specific query, and is analogous to positive predictive value. To estimate the precision, we randomly selected 250 concepts among those that appeared in the collection. For each concept, we selected a random sample of up to five figure captions in which MMTx identified the concept as present. Those captions were reviewed manually to determine if the caption was indexed by specified concept correctly (true positive [TP]) or incorrectly (false positive [FP]). We computed the precision as the number of captions correctly indexed (TP) divided by the total number of captions indexed (TP + FP). We calculated the 95% confidence interval (CI\(_{95}\)) for precision based on the size of the sample.

**Recall**

Recall measures the fraction of all the relevant documents in a collection that are retrieved by a specific query, and is akin to the concept of sensitivity. Here, recall is the number of figure captions that were indexed by a concept divided by the number of captions in which the concept was actually present. We estimated recall by sampling concepts and captions. We randomly selected 40 concepts, each of which MMTx had indexed in more than 10 figure captions. For each concept, the true positive (TP) value was estimated as the total number of “positive” captions (those indexed by that concept) multiplied by the overall precision value. Then, for each concept, we sampled 25 figure captions from among those that were not indexed by that concept and reviewed those concept-caption pairs. Those captions should be negative; the “Sample TN” is the number of true negative (TN) figures among the 25 sampled for each concept. Based on the Sample TN value, we extrapolated to the entire set of negative captions. Recall was computed as the number of correctly indexed captions (TP) divided by the number of captions that truly contained the concept (TP + TN).

We illustrate our estimation of recall with an example. Consider the concept Liver diseases (C0023895), which was identified in 90 figure captions (Table 2). Given an overall precision value of 90%, there are an estimated 81 “true positive” (TP) captions for this concept. Now we examine the sample of 25 captions not indexed by this concept. If one of the 25 sampled captions in fact contains the concept, then that caption is falsely negative; thus the false-negative fraction would be 1/25. To estimate the number of false negatives in the entire dataset, we multiply the false-negative fraction by the total number of negative captions (84,756 captions = 84,846 − 90) to yield 3,390. Thus, for Liver diseases, the estimated recall would be TP/(TP + FN) = 81/(81 + 3,390) = 0.023.

\[ F_1 = \frac{2 \cdot P \cdot R}{P + R} \]

**Impact on Search Engine Performance**

To evaluate how semantic indexing enhances search, we obtained a set of 10,000 randomly selected entries from the

<table>
<thead>
<tr>
<th>CUI</th>
<th>Concept Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1305775</td>
<td>Entire portal vein</td>
</tr>
<tr>
<td>C0017110</td>
<td>Gases</td>
</tr>
<tr>
<td>C0596601</td>
<td>Gastrointestinal gas</td>
</tr>
<tr>
<td>C0032718</td>
<td>Portal vein structure</td>
</tr>
<tr>
<td>C0205054</td>
<td>Hepatic</td>
</tr>
<tr>
<td>C1278960</td>
<td>Entire vein</td>
</tr>
<tr>
<td>C0042449</td>
<td>Veins</td>
</tr>
<tr>
<td>C0009924</td>
<td>Contrast Media</td>
</tr>
<tr>
<td>C0040405</td>
<td>X-ray computed tomography</td>
</tr>
<tr>
<td>C0441633</td>
<td>Scanning</td>
</tr>
<tr>
<td>C0520510</td>
<td>Materials</td>
</tr>
<tr>
<td>C1201820</td>
<td>Obtained</td>
</tr>
<tr>
<td>C0000811</td>
<td>Termination of pregnancy</td>
</tr>
<tr>
<td>C1278929</td>
<td>Entire liver</td>
</tr>
<tr>
<td>C0023884</td>
<td>Liver</td>
</tr>
<tr>
<td>C0151747</td>
<td>Renal tubular disorder</td>
</tr>
<tr>
<td>C0332208</td>
<td>Tubular formation</td>
</tr>
<tr>
<td>C0205216</td>
<td>Decreased</td>
</tr>
<tr>
<td>C0205100</td>
<td>Peripheral</td>
</tr>
<tr>
<td>C0336721</td>
<td>Arrow</td>
</tr>
<tr>
<td>C0243095</td>
<td>Finding</td>
</tr>
<tr>
<td>C0582254</td>
<td>Intrahepatic portal vein</td>
</tr>
<tr>
<td>C1512948</td>
<td>Intrahepatic</td>
</tr>
</tbody>
</table>
ARRS GoldMiner search engine’s log file. Each log-file entry included the total number of images \(N\) retrieved, the number of images found by concept-based search alone \(C\), and the number found by keyword-based search alone \(K\). Because the total search result is the union of the concept- and keyword-based searches,
\[
\frac{N}{H_{11349}} = \frac{C}{H_{11001}} + \frac{K}{H_{11006}}.
\]
We computed the fraction of results that were contributed by concept-based search alone—that is, \((N-K) / N\)—to assess the extent to which concept-based searching increased the number of total results. Keyword-based search used the MySQL database management system’s case-insensitive, whole-word “FULLTEXT” indexing method.

**Results**

The MMTx program identified 31,108 unique concepts in the radiology figure captions. A figure caption with its indexing terms is shown as an example in Table 1. The number of concepts found per figure caption ranged from 0 to 227 (median, 36; mean ± SD, 38.6 ± 20.1). The distribution of the number of concepts per caption is shown in Fig 2.

At least one concept was discovered in 83,573 (99.95%) of the 83,615 nonempty figure captions. The five most common concepts appeared in 41–62% of all captions, whereas 4,035 concepts appeared only once. The 50 most common concepts (0.2% of all concepts identified) accounted for 25% of references to concepts in the entire collection.

**Precision**

By selecting up to five figure captions indexed for each of 250 randomly selected concepts, 890 figure captions were

### Table 2 - Estimate of Recall from Sample of 40 Concepts

<table>
<thead>
<tr>
<th>CUI</th>
<th>Concept Name</th>
<th>Captions Indexed</th>
<th>Sample TN</th>
<th>Est. Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0011304</td>
<td>Demyelination</td>
<td>51</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0011331</td>
<td>Dental Procedures</td>
<td>18</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0016911</td>
<td>Gadolinium</td>
<td>1,770</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0018099</td>
<td>Gout</td>
<td>43</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0020883</td>
<td>Ileostomy</td>
<td>30</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0021925</td>
<td>Intubation</td>
<td>54</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0023895</td>
<td>Liver diseases</td>
<td>90</td>
<td>24</td>
<td>0.023</td>
</tr>
<tr>
<td>C0030424</td>
<td>Paragoniomiasis</td>
<td>16</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0032463</td>
<td>Polycythemia Vera</td>
<td>174</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0037939</td>
<td>Spinal Neoplasms</td>
<td>291</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0038895</td>
<td>Surgical Aspects</td>
<td>2,121</td>
<td>22</td>
<td>0.159</td>
</tr>
<tr>
<td>C0040132</td>
<td>Thyroid Gland</td>
<td>334</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0042382</td>
<td>Vascularization</td>
<td>53</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0085406</td>
<td>Anisotropy</td>
<td>186</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0149554</td>
<td>Frontal Horn</td>
<td>79</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0179376</td>
<td>Bottle, device</td>
<td>15</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0185792</td>
<td>Incision of sternum</td>
<td>16</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0205556</td>
<td>Qualitative</td>
<td>58</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0225897</td>
<td>Left ventricular structure</td>
<td>648</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0226862</td>
<td>Structure of straight sinus</td>
<td>53</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0280100</td>
<td>Solid tumor</td>
<td>67</td>
<td>20</td>
<td>0.003</td>
</tr>
<tr>
<td>C0332218</td>
<td>Difficult</td>
<td>396</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0332272</td>
<td>Better</td>
<td>1,567</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0428772</td>
<td>Left ventricular ejection fraction</td>
<td>23</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0433343</td>
<td>Unstable status</td>
<td>80</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0494379</td>
<td>Connection</td>
<td>186</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0450195</td>
<td>Cervicothoracic</td>
<td>27</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0489800</td>
<td>Left Calf</td>
<td>59</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0521104</td>
<td>Permission</td>
<td>947</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0522537</td>
<td>Xenograft type of graft</td>
<td>11</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0560737</td>
<td>Bone structure of hamate</td>
<td>28</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C0600080</td>
<td>Stretching exercises</td>
<td>65</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C1269584</td>
<td>Entire posterior semicircular canal</td>
<td>11</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C1278929</td>
<td>Entire liver</td>
<td>4,283</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C1280264</td>
<td>Entire pterygoid muscle</td>
<td>25</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C1280605</td>
<td>Entire infratemporal fossa</td>
<td>19</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C1280839</td>
<td>Entire incus</td>
<td>53</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C1305627</td>
<td>Entire superior ramus of pubis</td>
<td>11</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C1446409</td>
<td>Positive</td>
<td>1,371</td>
<td>25</td>
<td>1.000</td>
</tr>
<tr>
<td>C1457873</td>
<td>Os trigonum disorder</td>
<td>19</td>
<td>25</td>
<td>1.000</td>
</tr>
</tbody>
</table>

**MACRO-AVERAGE**

0.930

For each concept, the table lists the UMLS concept unique identifier (CUI), the concept name, and the number of “positive” captions (indexed by that concept). For each concept, 25 “negative” figure captions (those not indexed by the concept) were sampled. The number of true negatives in that sample (sample TN) is indicated, and the estimated recall value is computed.
The mean precision was 0.897 (CI95, 0.857–0.937).

The term “HCC”. Conceptual indexing allows search engines to intelligently unify such lexical variants, as well as to expand queries to retrieve information based on the meaning of the concept and its relationships to other concepts.

In radiology figure captions, descriptions tend to focus on anatomy, diseases, radiological findings, and imaging techniques—a subset of general language which is much more varied. The focused scope of radiology language may account for the high performance of our approach.

Our concept-based indexing approach has been incorporated into GoldMiner to improve retrieval for user searches. The results for the 10,000 search queries suggests that concept-based indexing substantially increases the number of images retrieved; in fact, our methods have been adopted in the current release of GoldMiner. Given the high precision and recall of the concept-based index, the images retrieved should be highly relevant to the query terms. Identification of age, sex, and imaging-modality metadata in radiology figure captions also can be accomplished with high recall and precision.

Concept-based indexing of text has been undertaken in earlier work. For example, in using the SAPHIRE system to index concepts in radiology reports, the researchers found recall of 63% but a precision of only 30%. The precision and recall found in this study were very high. Some of the differences in our results from that of prior work may relate to the methods for concept recognition and differences in the domain of the text being indexed.

Retrieval of images based on their visual content and textual annotations is an area of active research. The ImageCLEFmed 2008 medical image retrieval task, part of the Cross-Language Evaluation Forum (CLEF) information retrieval challenge, employed a subset of images that have been indexed by ARRS GoldMiner. Yu and colleagues have explored the analysis of figure captions and associated text from journal articles to answer biological questions.

Because of the rich interconnections among its component vocabularies, the UMLS Metathesaurus is an important source of medical knowledge. Indexing of unstructured text to standardized vocabularies—similar to that done in this study—has improved information retrieval in several other biomedical domains. The KnowledgeMap system has been used to identify Metathesaurus concepts in the impression text of electrocardiogram reports. Dermatlas, a Web-based collection of dermatology cases, was indexed to MeSH terms using the National Library of Medicine’s Medical Text Indexer (MTI). Ontology-based indexing has been shown to aid retrieval and extraction of information from the biomedical literature. Shah, et al developed and applied techniques to map free-text annotations of tissue microarray data to structured vocabularies. We chose the approach described here because MMTx offered high-quality indexing, was readily available, and was integrated with the UMLS Metathesaurus. Lexical expansions and exploitation of knowledge in UMLS make this approach particularly advantageous in the radiology domain to improve recall of matching concepts. Preliminary analysis of GoldMiner’s performance showed that this indexing approach functioned well in our domain.

A limitation of our work is that we did not measure recall for all concepts, but estimated it by sampling. To measure
recall most accurately, one would have to determine how many relevant documents are retrieved for each search concept. Given the number of concepts and the size of the database, such measurement would have been prohibitive. We believe our estimated recall based on sampling figure captions and concepts is a reasonable approach to this limitation.

Although widely used, MMTx, which is incorporated into our system, has several limitations: it is relatively slow, it is limited to UMLS vocabularies, and it is unable to process negation. Because figure captions from journal articles are processed as a “background” task, MMTx’s processing speed was not detrimental to our project. Investigators have developed a new MetaMap module that identified 91% of the concepts found by MMTx in 14% of the time taken by MMTx. Alternative algorithms, such as MGREP or MTag, may provide sufficient speed to allow real-time mapping of clinical text to controlled vocabularies. Such systems would allow flexibility to use vocabularies, such as RadLex, which are not yet part of the UMLS. RadLex offers terms for radiology-specific observations that are not found in other terminologies. One goal is to integrate semantic indexing of clinical radiology reports in real time. Real-time indexing could allow integration of clinical systems with ontology-based knowledge resources.

Another limitation of MMTx is that it depends on UMLS for its source terminologies, and UMLS lacks terminologies specific to radiology. RadLex, a unified vocabulary for radiology that is being transformed into an ontology of radiology knowledge may help improve our concept-based image retrieval method. Until RadLex is incorporated into the UMLS Metathesaurus, other tools must be used to map text to terms in that lexicon. Another limitation is that MMTx lacks negation detection, so that both positive and negative statements are indexed equivalently. Although satisfactory for figure captions (which generally mention negative concepts only if relevant, e.g., “no evidence of appendicitis”), such an approach likely would retrieve too many false-positive results when dealing with clinical text such as radiology reports.

Concept-based indexing of clinical documents is an area of active investigation. Although text mining and semantic indexing have been applied successfully to molecular biology and the biomedical literature, relatively few studies have explored their application to clinical content. In radiology, automated techniques have been used to code findings in cancer-related radiology reports, to identify findings of congestive heart failure, and to identify clinically important findings. Semantic indexing has improved noun phrase identification and overall precision of information retrieval in radiology reports. Real-time semantic indexing of the content of radiology reports creates opportunities to integrate the reporting process with clinical decision support and point-of-care learning, and may improve the quality of radiology practice and learning.

**Conclusions**

Our goal was to assess the performance of the MMTx system for concept-based indexing of radiology figure captions. In our study, MMTx demonstrated precision of 0.897 and estimated recall of 0.930. This indexing approach has been incorporated into the ARRS GoldMiner Web-based image search engine. Concept-based indexing allowed retrieval of results not identified by keyword-based retrieval in more than half of all actual search queries, based on a large sample. Concept-based indexing can achieve high precision and recall, and can improve retrieval of radiology images and their textual captions.

**References**


44. Friedlin J, McDonald CJ. A natural language processing system to extract and code concepts relating to congestive heart failure from chest radiology reports. AMIA Annu Symp Proc 2006:269–73.


