Using Image References in Radiology Reports to Support Enhanced Report-to-Image Navigation

Thusitha Mabotuwana, PhD, Yuechen Qian, PhD, Merlijn Sevenster, PhD
Philips Research North America, Briarcliff Manor, NY

Abstract

Radiology reports frequently contain references to image slices that are illustrative of described findings, for instance, "Neurofibroma in superior right extraconal space (series 5, image 104)". In the current workflow, if a report consumer wants to view a referenced image, he or she needs to (1) open prior study, (2) open the series of interest (series 5 in this example), and (3) navigate to the corresponding image slice (image 104). This research aims to improve this report-to-image navigation process by providing hyperlinks to images. We develop and evaluate a regular expressions-based algorithm that recognizes image references at a sentence level. Validation on 314 image references from general radiology reports shows precision of 99.35%, recall of 98.08% and F-measure of 98.71%, suggesting this is a viable approach for image reference extraction. We demonstrate how recognized image references can be hyperlinked in a PACS report viewer allowing one-click access to the images.

Introduction

On a routine basis, radiologists have to work with an increasing number of imaging studies to diagnose patients in an optimal manner. Patients, especially ones with cancers, frequently undergo imaging exams and over time accumulate many studies, and thereby reports, in their medical records. Each time a new study is acquired, the radiologist would typically open one or more prior radiology reports to establish the patient's clinical context. A similar practice can be observed by the consumers of the radiology reports as well, such as oncologists and referring physicians.

In the radiology workflow, after a patient has had an imaging study performed, using X-ray, CT, MRI (or some other modality), the images are transferred to the picture archiving and communication system (PACS) using Digital Imaging and Communications in Medicine (DICOM) standard [1]. Radiologists read images stored in PACS and generate a radiology report generally using a reporting software (e.g., Nuance PowerScribe 360 [2]). The report is then transferred to the PACS or the Radiology Information System (RIS) depending on the hospital's IT infrastructure via Health Level 7 (HL7) standard [3]. The radiology reports are narrative in nature and typically contain several institution-specific section headers such as *Clinical Information* to give a brief description of the reason for study, *Comparison* to refer to relevant prior studies, *Findings* to describe what has been observed in the images and *Impression* which contains diagnostic details and follow-up recommendations.

Oftentimes, the radiology reports also contain references to specific images of findings, for instance, "Neurofibroma in the superior right extraconal space (series 5, image 104) measuring approximately 17 mm". In the current workflow, if a radiologist (or a downstream consumer of the report) wants to look at the images referenced in a prior report, for instance, for assessing interval change, the user has to manually open the prior study in the PACS, find the series of interest (series 5 in this example), and then navigate to the corresponding image slice (image 104 in this example). Imaging studies can have multiple series, such as, for instance, axial, coronal and sagittal series and series with dedicated window-level settings. Typically the series description (e.g., AXL W/) is displayed in the PACS reading environment, but the series index number (e.g., 5) is not (which is typically used in the references in reports). This makes locating the series of interest non-intuitive. Furthermore, a series may contain multiple images (typically several hundred in the case of MRIs) making the navigation from report text to images a time consuming process.

If referenced images can be integrated deeply into the radiologist's reading environment, for instance, by hyperlinking image references, this will increase the efficiency and accuracy of the mental process of synthesizing a comprehensive patient history. In this paper we describe the development and evaluation of a method for recognizing and normalizing image references in narrative radiology reports. We also show how the normalized data can be used to support enhanced navigation from the report to the referenced image in a commercial PACS environment.

Methods

Dataset and strategy

We used de-identified production radiology reports from an academic hospital in the Midwest. All electronic protected health information (ePHI) was removed (with dates shifted) from the dataset per HIPAA regulations. Institute Review Board was waived (Protocol Number: 11-0193-E). The dataset contained over 20,000 radiology reports dictated by over 40 different radiologists for over 10,000 patients across ten different modalities.

The development and evaluation plan is divided in three phases. In the pre-development phase, all image reference patterns were charted based on a pre-development data set comprising of 250 reports. A recognizer is constructed based on these patterns and obvious generalizations. In the development phase, using a ground truth corpus based on 500 sentences, the recognizer is evaluated and refined accordingly. In the test phase, using a ground truth based on another 500 sentences, the final recognizer is evaluated. The data sets used in the three phases are strictly disjoint.

References to images in the report

Using a pre-development set of 250 randomly selected reports, we first examined the different ways radiologists referred to specific images in their free-text reports. The sentences containing image references were manually copy-pasted into a master document to understand variations. Once all the sentences containing image references were identified, we tabulated the 'high-level' categories that need to be addressed by the image reference extraction algorithms. Table 1 shows a number of typical variations and their interpretation.

Algorithm development

We identified several components that would be needed to accurately extract image references from narrative radiology reports: a natural language processing module to determine sentence boundaries, a measurement and date identifier, an image reference extractor to extract the series and image information, and a module to determine the temporal context. Following is a brief description of each module in the order they are used in the overall text processing pipeline.

Sentence boundary detector:

In order to support report-to-image navigation, we need to know which study the image reference is on. It is assumed that a reference without an explicitly-specified date or mentioning of comparison/priors typically is on the current study. Therefore, extracting image references at a sentence level would make the most logical sense, as opposed to, for instance, using information at a paragraph/report level. This assumption was supported by our observations of the 250 reports.

To accomplish this, a sentence boundary detection algorithm was developed that recognizes sections, paragraphs and sentences in narrative reports. The algorithm implements a maximum entropy classifier that assigns to each end-of-sentence marker ('.', ':', '!', '?', '\n') in the running text one of four labels:

- Not end of sentence (e.g., period in 'Dr. Doe');

- End of sentence, and sentence is a section header (e.g., Colon in *Findings*: section);

- End of sentence, and sentence is the last sentence in a paragraph;

- End of sentence and none of the above classes.

Using this classification, the entire section-paragraph-sentence structure can be reconstructed. In a post-processing step, recognized section headers (e.g., *Findings*) were normalized with respect to a list of known section headers.

Table 1: Different ways of referencing images within radiology reports

Narrative Context	Image Reference	Comments
…in the superior right extraconal space (series 5, image 104) measuring approximately 17 mm	Series 5, image 104	This is the most common pattern.
Subcutaneous nodule on (image 23; series 9)…	Series 9, image 23	Similar to previous, but "image" appears first. Various punctuation can appear between series and image (e.g., comma, semicolon, colon, hyphen, '/', 'and', 'on').
…low attenuation in right paramedian anterior to the surgical site (axial image 6)	Axial series, image 6	Reference to series is via an imaging plane.
Enhancing nodule in the inferior right breast (63/145) minimally increased compared to prior.	Image 63 of series with 145 images	Has a number-over-number pattern. Interpreted as "index of the image/number of images in series".
Enlarged right hilar lymphadenopathy measuring 11 mm (80232/49), unchanged.	Series 80232, image 49	Similar to above, but first number is greater than second. Interpreted as "series/image".
…luminal cross-sectional dimension of 11 x 10mm (series 3 image 125/152), unchanged.	Series 3, image 125	Reference to 152, the total number of slices, is redundant since series is explicitly specified.
…precentral gyrus which has been displaced medially against the falx (axial image 108 series 605; coronal image 213 series 865)	Series 605, image 108 Series 865, image 213	References to imaging viewing planes are redundant since series is explicitly specified.
…as seen on recent upper GI examination (series 7 images 33, 35)	Series 7, images 33, 35	Multiple images referenced on the same series.
This lesion is best seen on images 51-59, series 15	Series 15, images 51-59 (9 images)	Multiple images referenced on the same series as a range.
Root thickening and enhancement (series 2801 and series 3201 image 8)	Series 2801 and 3201, image 8	Multiple series referenced (this is a rare scenario) for same image.
…hyperdensity embedded in the right posterior wall of glottis (image 56 of series 3 and image 36 of series 8025)	Series 3, image 56 Series 8025, image 36	Multiple image references on different series.
On the 16-October-2011 scan, nodule measures 3cm (series 5, image 26)	Series 5, image 26 on 16-Oct-2011 study	Refers to an image on a prior study.
…white matter signal increase (series 601 image 44 on 03-Jun-2010) is not as clearly seen today.	Series 601, image 44 on 03-Jun-2010 study	Similar to above, but study date is after image reference.
8mm nodule at the left base (image 60; series 5) is unchanged compared to the prior exam (image 61; series 6; 22-Jul-2001 study).	Series 5, image 60 on current; Series 6, image 61 on 22-Jul-2001 study	Multiple image references – one on current, one on prior. Word 'comparison' is sometimes used instead of 'prior'.
5 mm (series 4, image 72), previously 9 mm, (series 4, image 51).	Series 4, image 72 on current; Series 4, image 51 on prior	Similar to previous, but no explicit date reference to prior study.
On the 12/19/10 scan as measured on series 2 image 120 is 3.9 x 2.7 cm and on the 26-Feb-2011 CT series 3 image 109 is 3.8 x 2.6 cm	Series 2, image 120 (on 19-Dec-2010 study); Series 3, image 109 (on 26-Feb-2011 study)	Multiple image references belonging to multiple prior studies.

Measurement and date identifier:

Phrases with numerical characters act as distractors to the image reference recognition process, if they appear close to image references. The most prevalent distractor categories are dates and measurements, e.g., "…nodule is unchanged compared to the prior exam (series 8; image 14; 20-Sep-2003 study)". In this case, the reference extraction algorithm needs to determine that the referenced image is 14, instead of ignoring punctuations and extracting *images 14-20* or *images 14 and image 20*. We used two regular expressions to label phrases that denote measurements and dates (e.g., 5 mm, 5 x 3 mm, and 10-July-2011, 20/12/2010). The output of the distractor recognition step is a labeling of substrings in a given sentence, which can then be ignored by the image reference recognition step.

Image reference extractor:

We took a rule-based approach to extract the references. Each rule was implemented by means of a regular expression. Four primary rules were identified:

1. The first rule extracts references where both the series and image information are provided explicitly along with one or more numeric values (e.g., *series 55, images 45, 46 and 48*; *image number 14 on series 20*; *image 50/series 4*; *series 4, image 43, image 44*; *series 3 image 125/152*; *coronal series 5, image 33*).

2. This rule is similar to the first, but a series can be referred to using a scanning plane, for instance, *coronal image 118*; *axial image 5/44* and *axial images 5-12*.

3. Third rule extracts information from sentences that contain series related information (via the word "series" or reference to a scanning plane), but do not contain "image" along with the corresponding numeric values (e.g., *series 80254/48; coronal/35* and *280; series 8*).

4. The last rule looks for the number-over-number pattern, not matched by any of the prior rules. An example of this is *left lower lobe has increased in size, measuring 2 x 1.6 cm (65/131).* By convention, we interpret the first number as the particular image number (65 in this case) and the second number as the total number of images in the series (131 in this example) if the first number is less than or equal to the second. Similarly, the result is interpreted as series/image if the first number is greater than the second, for instance, *Enlarged right hilar lymphadenopathy measuring 11 mm (80232/49).* For this rule to hold, the image value or the number of slices needs to be out of date ranges (e.g., *11/25* will not be matched by this rule). Our design is based on the assumption that in a clinical environment, it would be more favorable to interpret image references accurately than introducing false positives (i.e., minimizing false-positives at the expense of having false-negatives).

Our method allows for finding multiple image references in a single sentence using different rules, for instance, two references will be extracted from *Mild centrilobular and paraseptal emphysema. 2.3 x 2.3 by 2-cm (16/49, coronal/35) lobulated nodule in the left lower lobe adjacent to the hilum* using rules 4 and 3 respectively.

Temporal context extractor:

To determine to which study an image reference points, we look for combinations of keywords (e.g., 'scan' and 'study') and dates that occur to the left of an image reference (e.g., *On the 12/19/10 scan as measured on series 2 image 120…*); dates that occur to the right of an image reference (e.g., *...white matter signal increase (series 601 image 44 on 03-Jun-2010) is not as clearly seen today*); and keywords such as 'prior' and 'previous' occurring near the image reference (e.g., *...measures 3 mm (series 4, image 52), previously 4 mm (series 4, image 32)* and *...on image 17 series 4 of prior study*). If this search yielded one or more results, we determined that the image reference does not point to the current exam. In addition, if a date was found, we determined that this was the prior study's date.

Ground truth construction

With the assistance of the sentence boundary detection module, we first parsed all reports (excluding the 250 used for algorithm development) to extract the sentences within the reports. Determining the ground truth for all these sentences was not practical, and therefore we filtered out sentences that did not contain any numerical characters.

This reduced the total number of sentences to 85,157. This set was further filtered based on whether the sentence contained the strings 'image', 'series', 'coronal', 'axial', 'sagittal', 'oblique' or a number-over-number pattern ('…/…'). This resulted in 4,992 sentences. From this set, we randomly selected 500 sentences as the development set and another 500 sentences as the test set.

We developed a ground truth annotation tool for capturing the image reference(s) mentioned in a sentence. For a selected sentence, the user can manually enter the series number, image number and the temporal resolution (i.e., study date) which would be considered the 'ground truth'. To speed up the annotation process, we pre-populate the image references as determined by our algorithms and display this to the user. For instance, in Figure 1, two references were extracted from the highlighted text (shown in blue) – one from the current study, and one from the study dated 22-Jul-2001.

Figure 1: Tool to establish ground truth

In this instance, if the extracted information is correct the user can move onto the next sentence upon which the row will be highlighted in grey indicating that this row has already been processed. If extra information is mistakenly extracted or the correct reference is not extracted, the user can enter the correct information directly into the table (towards the bottom of the screen).

Validation

In our evaluation, an instance is a substring of a sentence, a date and a non-empty list of series–image pairs. An instance is positive if its associated substring contains an image reference and its associated list contains precisely the series–image pairs to which the substring refers. A negative instance is any non-positive instance. In the context of a classifier, a positive instance is true positive, if the classifier correctly recognizes the image reference in free text and correctly retrieves the referenced date and series–image pairs. The notions of false positive and false negative are defined similarly. As an example, if there are two image references in a sentence and the algorithm only extracted one of them correctly (or extracted both references, but series number/image number/temporal resolution extraction was incorrect), the sentence is classified as a false-negative.

Let *tp*, *fp* and *fn* denote the number of true positives, false positive and false negatives, respectively, for a given classifier. Then, the classifier's precision and recall are defined as $tp/(tp+fp)$ and $tp/(tp+fn)$ respectively. The F-measure of the classifier is defined as the harmonic mean of its precision and recall: $2\times\text{precision}\times\text{recall}/(\text{precision}+\text{recall})$.

Results

Development set

We evaluated our system on the development set before refinement. Author TM annotated the sentences in the development set which contained 371 image references. In this evaluation, two false negative were found (see Table 2) resulting in F-measure 99.73% (precision 100%, recall 99.46%).

Table 2: 2x2 contingency table for evaluation on development set (n = 371)

		Ground Truth	
		True	False
Extracted References	True	369	0
	False	2	-

The reason for the two false negatives was that in one instance, the word 'image' was repeated (*series 7 image image 295/314*). If the repeat would not have been there, it would have matched the first rule. Initially we coded our algorithms so that for the number-over-number rule to hold, it needs to appear within parenthesis since this was a commonly seen pattern. In the other false-negative, an image reference was interpreted as a date (*...nodule in the right upper lobe (5/19) unchanged*). The latter is the expected behavior in order to minimize false-positives and consistent with our fourth rule. The reference extraction algorithm was refined so as to recognize number-over-number patterns that are not enclosed by parentheses and to handle duplicate words.

Test set

Author MS annotated the sentences in the test set. 314 image references were found for 263 sentences from 228 unique reports. F-measure on the test set was 98.71% (precision 99.35%, recall 98.08%). Table 3 shows the 2x2 contingency table for the validation results. 276 of the references matched the first rule, 11 matched the second rule, 9 matched the third and 18 matched the fourth. We observed a maximum of 8 references in a single report.

Table 3: 2x2 contingency table for evaluation on test set (n = 314)

		Ground Truth	
		True	False
Extracted References	True	306	2
	False	6	-

The two false positives were found in the following two sentences.

1. *Additionally, a focus of increased attenuation is noted in the manubrium (65; sagittal series 80213) likely representing an inflammatory process* – in this case, the algorithm extracted two references: 'Series: sagittal, image: 65' and 'Series: 80213, image: 65'. Since the sentence refers to only one image, we counted this as a false positive.

2. *An example of this lesion class would include right parietal cortex lesions in series 6 image 9 (compared to series 6 image 10 on 17-Oct)* – in this instance, 17-Oct was not correctly marked as a valid date, and

therefore the algorithm extracted Series: 6, image: 10, 17 resulting in an additional reference. As a result, the temporal resolution for the second reference was incorrect as well.

The six false negatives were found in the following sentences:

1. *An index nodule measures 5 mm on image 55 Jan-2005 study.* – this did not match any of our rules.

2. *Anterior to the stent there is heterogenous density of the native mural thrombus, present on non-contrast and delayed images (image 116, narrow windows), consistent with heterogenous evolving thrombus.* – we had not considered 'narrow window' as being a valid series.

3. *Using similar axial dimensions for comparison to the prior study from 3/2009, the sinus of Valsalva in the axial plane is approximately 53 mm (series 301 image 13), similar to 2009 study (series 401 image 12).* – temporal resolution for the second reference was incorrect since '2009 study' was not detected as a valid prior date.

4. *However, there are areas of poor visualization of the dura in this region (series 1601 images 8 versus 9), and there is a moderate quantity of soft tissue (paraspinal) edema that extends inferiorly from this region down to approximately C5.* – 'versus' was not considered as a valid separator and therefore only *Series: 1601, image: 8* was extracted by the algorithm.

5. *Anterior mediastinal mass measuring 8.9 x 4.9 cm and axial dimensions (3/42), previously measured 10.3 x 6.0 cm (2/26).* – '2/26' was not extracted as being a valid reference. This is as expected since the '2/26' can be also interpreted as a date.

6. *Anterior mediastinal soft tissue mass measures 2.2 x 3.1 cm (3/25), previously 2.2 x 3.2-cm.* – '3/25' was not extracted as being a valid reference. This is as expected (similar to previous case).

PACS integration

To demonstrate how the developed algorithms can be used in practice, we integrated our algorithms into Philips iSite Radiology PACS [4] where a user can simply click on the hyperlink in the custom-rendered radiology report to open the corresponding image (in the correct study, and series). Figure 2 shows how the image references have been hyperlinked, and how clicking on a particular reference has opened the corresponding image in a separate window.
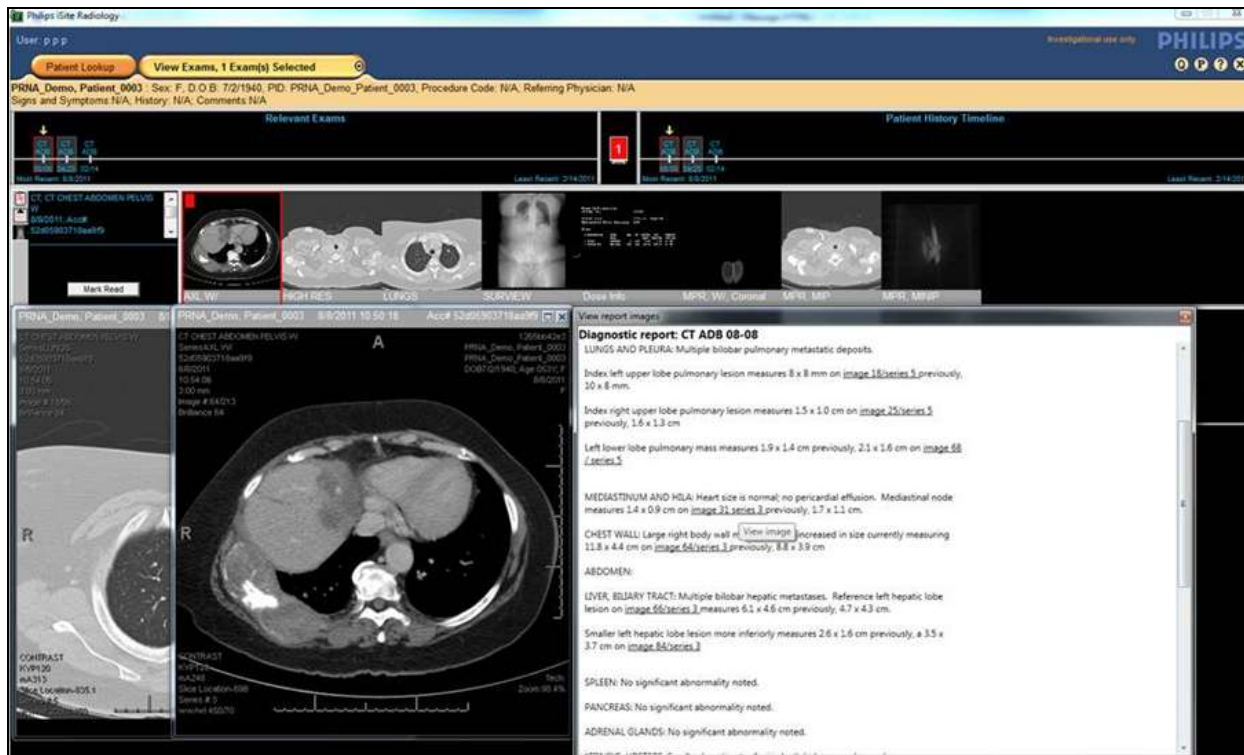
Figure 2: Application of image reference extraction algorithms in iSite PACS for a selected patient. The report window (on the right) shows the hyperlinked references while the two image windows (on the left and middle) show the corresponding images that have been opened by clicking on the linked references.

Discussion

In this work we have demonstrated how a pattern-matching approach can be used to extract image references from radiology reports. As demonstrated, the developed algorithms can be used in the context of an image reference extraction module which can then be integrated into a broader framework as needed, for instance, to allow radiologists and other consumers of radiology reports to easily navigate from a report to image(s). Also, the proposed technique was validated using production study descriptions with very high accuracy indicating that a pattern-matching approach can be used for image reference extraction, for instance, opposed to more involved, data-intensive statistical methods. We also demonstrated an application of the algorithms by integrating the algorithms into iSite Radiology.

Information extraction from narrative radiology reports has been an important area of radiology informatics. Much of the past research has focused on extracting clinical findings from reports. For instance, Prasad et. al. [5] developed a medical finding extraction module to automatically extract findings and associated modifiers to structure radiology reports. Mamlin et. al. [6] used natural language processing (NLP) techniques to extract findings from cancer-related radiology reports. From an implementation point of view, the work by Lakhani et. al. [7] is perhaps the most closely related to our work. They developed a rule-based algorithm to identify critical results from radiology reports. Conceptually, our work contributes to realizing the notion of 'multimedia radiology report' [8] in which radiological findings, recommendations and diagnoses are connected with other information sources, such as medical encyclopedia [9] and images. Hyperlinking image references, has been alluded to in the literature [10], but to the best of our knowledge this is the first attempt where image references have been extracted from the reports and integrated within a PACS.

The current study has several limitations. The most notable is that the reports we used were from one hospital, and therefore the style of reporting image references may not be generalizable to other institutions, although the reports were produced by over 40 radiologists providing a large spectrum of writing styles. Secondly, the algorithm integration into iSite Radiology currently only works for image references with named series, for instance, those matched by our first rule. We intend to make the integration more dynamic so that the integration engine can

dynamically determine the series that contains a given number of slices (e.g., a reference matched by the fourth rule). The integration engine currently does not account for prior studies either; for instance, the reference to the prior study in *5 mm (series 4, image 72), previously 9 mm, (series 4, image 51)* is indicated by the word 'previously'. In this case, the integration engine will need more advanced NLP techniques to determine the exact prior study, for instance, by extracting it from the *Study Comparison* section of the report. Thirdly, our interpretation of the number-over-number pattern *n/m*, is based the assumption that *n* is the image index and *m* is the number of slices if $n{\leq}m$, and that *n* is the series index and *m* is the image index if $n{>}m$. Further research will be needed to disambiguate between image/series or series/image versus image/total images, perhaps by using more dynamic PACS integration with cascaded rule checking. Lastly, this research did not explore the clinical validity of providing image references to PACS users.

Several of the false-negative sentences contained references that were difficult to distinguish from a date (e.g., 3/25). Our observation is that series references often contain either a single digit, or five digits (typically <10 or >80,000 in our dataset). We have also noticed that series '4' typically refers to a lung series. Such observations could be site and/or institution specific, but it may be possible to create specific algorithms to resolve date-related ambiguities using heuristics, more contextual information and rules as well as developing active learning techniques.

As future work, we are working on improving the PACS integration as well as exploring how the extracted image references can be used in the actual radiology workflow to improve workflow efficiency. Further, exploring algorithm integration options into the report creation workflow itself could be useful since it can be validated by the radiologist at the point of report creation which may also encourage/train them to create better linked reports.

References

1. *DICOM Homepage*. [cited 2012 Aug 4]; Available from: http://medical.nema.org/.
2. *PowerScribe 360 RIS & PACS Solutions*. [cited 2013 Feb 18]; Available from: http://www.nuance.com/for-healthcare/capture-anywhere/radiology-solutions/index.htm.
3. *Health Level Seven International* [cited 2013 Feb 18]; Available from: http://www.hl7.org/.
4. *Philips Healthcare*. [cited 2013 Feb 19]; Available from: http://www.healthcare.philips.com/.
5. Prasad, A., Ramakrishna, S., Kumar, D., et al., *Extraction of Radiology Reports using Text mining*. International Journal on Computer Science and Engineering, 2010. **2**(5): p. 1558-1562.
6. Mamlin, B.W., Heinze, D.T., and McDonald, C.J., *Automated extraction and normalization of findings from cancer-related free-text radiology reports*. AMIA Annu Symp Proc, 2003: p. 420-4.
7. Lakhani, P., Kim, W., and Langlotz, C.P., *Automated detection of critical results in radiology reports*. J Digit Imaging, 2012. **25**(1): p. 30-6.
8. Chang, P., *Re-Engineering Radiology in an Electronic and "Flattened" World: Radiologist as Value Innovator*, in *RADExpo*. 2011: Virtual conference.
9. He, J., de Rijke, M., Sevenster, M., et al. *Generating links to background knowledge: a case study using narrative radiology reports*. in *20th ACM international conference on Information and knowledge management* 2011. New York.
10. Schomer, D.F., Schomer, B.G., and Chang, P.J., *Value innovation in the radiology practice*. Radiographics, 2001. **21**(4): p. 1019-24.