

Bridging the Text-Image Gap: a Decision Support Tool for Real-Time PACS Browsing

Merlijn Sevenster · Rob van Ommering · Yuechen Qian

Published online: 2 August 2011
© Society for Imaging Informatics in Medicine 2011

Abstract In this paper, we introduce an ontology-based technology that bridges the gap between MR images on the one hand and knowledge sources on the other hand. The proposed technology allows the user to express interest in a body region by selecting this region on the MR image he or she is viewing with a mouse device. The proposed technology infers the intended body structure from the manual selection and searches the external knowledge source for pertinent information. This technology can be used to bridge the gap between image data in the clinical workflow and (external) knowledge sources that help to assess the case with increased certainty, accuracy, and efficiency. We evaluate an instance of the proposed technology in the neurodomain by means of a user study in which three neuroradiologists participated. The user study shows that the technology has high recall (>95%) when it comes to inferring the intended brain region from the participant's manual selection. We are confident that this helps to increase the experience of browsing external knowledge sources.

Keywords Human–computer interaction · Image navigation · Image segmentation · Natural language processing · Artificial intelligence

M. Sevenster (✉) · R. van Ommering
Philips Research Europe,
Prof. Holstlaan 4,
5656AA Eindhoven, the Netherlands
e-mail: merlijn.sevenster@philips.com

Y. Qian
Philips Research North America,
345 Scarborough Road,
10510 Briarcliff Manor, NY, USA

Background

Due to technological advances, radiologic data are increasingly stored and transferred in digital form. Picture archiving and communication systems (PACS) allow for storage and retrieval of digital radiology images, whereas global connectiveness and communication standards in medicine (e.g., Health Level 7 and Digital Imaging and Communications in Medicine) enable instant exchange thereof. The digitalization of radiologic data radically changes the radiology workflow and the role of radiology in the care process [1]. It has been recognized that the digitalization of radiology is leading to a *commoditization* of radiology services; a lively debate has ensued if this is a desirable trend and how quality of care can be safeguarded [2–6].

In the meantime, we see a steady increase in the number of digital knowledge sources available to radiologists that can be consulted in the diagnostic and research process. As examples of such knowledge sources, we mention reports of imaging studies stored in a PACS [7], teaching files [8–10], archives of scientific publications (e.g., PubMed¹), expert-validated encyclopedia (e.g., STATdx²), and community-based content (e.g., Wikipedia,³ MyPACS,⁴ and Yottalook⁵).

However, as yet, these knowledge sources function as passive silos, not integrated into the radiology workflow. As a consequence, radiologists are less exposed to potentially relevant information than they could be.

On a technical level, the absence of a tight integration lays bare the lack of an agreed-upon means to represent the state of the radiology workflow that captures, for instance,

¹ www.ncbi.nlm.nih.gov/pubmed

² www.amyrsis.com

³ www.wikipedia.org

⁴ www.mypacs.net

⁵ www.yottalook.com

the clinical question, the level of expertise of the radiologist, his or her (sub)specialty, the anatomy of the case he or she is viewing, and the region of interest in the anatomy. Once we have a workflow-state interface between viewing environment and knowledge sources, we can expect the advent of *workflow-aware*, real-time clinical decision support (CDS) tools [11]. For instance, we can refine search results based on the workflow state, and we can even automatically fire queries that anticipate the radiologist's information needs.

Correlation of image and text data is one of the fundamental technological hurdles that need to be overcome to realize the workflow-aware paradigm of radiology support tools. Such technology would for instance enable the correlation of the region of interest in an MR image with the textual data from a given knowledge source pertaining to this area. Besides advanced CDS solutions, technologies that correlate medical image and text data can also be used to realize other applications, such as automatic generation of key images from free-text reports and graphic visualizations of report contents [12].

For an optimal browsing experience, it is essential that these CDS solutions require minimal understanding of their technological details and the structure of the knowledge sources linked to. For instance, we do not want to require that the user has knowledge about the way in which the image processing software recognizes the body structures, nor do we want to presume that the user knows the lemmas of a targeted medical encyclopedia.

In this paper, we propose a multi-modal algorithm that detects the anatomies of interest from a selection manually drawn on an MRI slice and correlates them to pertinent textual data. The phrase multi-modal refers to the algorithm's capability to interconnect image and text data. We embed this algorithm in an application that allows the user to find information relevant to selected anatomies in a PACS repository of neuro-MRIs and accompanying reports. This application tightly integrates the radiology workspace with PACS repositories, by allowing and encouraging real-time specification of information needs, that is, during the diagnostic process and within the very viewing environment. In practice, this system can be used as a CDS system for benchmarking hard cases, or in other applications that require anatomy-based retrieval of cases, as envisioned in [13].

The proposed algorithm addresses two technical challenges. First, it infers the anatomies of interest from a manually drawn selection on a neuro-MRI. To this end, we develop a scoring function that scores the anatomies based on four pixel-based variables. Second, it correlates the (potentially) limited set of anatomies recognized by the image processing engine with the anatomy terms used in the reports. This challenge is addressed by means of an ontology-based approach that allows us to make hierarchical inferences between anatomical concepts.

The algorithm described in this paper presumes that we have means to structure both the image and the text data by anatomy. In the case of the image data, this means that we have a module at our disposal that assigns an anatomy label to each pixel in a neuro MRI. To structure the reports, we assume we have a natural language processing engine that extracts body locations from reports and correlates them with reported findings and diagnoses. We briefly outline how such modules can be realized in the “**Methods**” section. The “**Discussion**” section of this paper contains three use cases that illustrate the system.

The contributions of this paper are twofold. In terms of feasibility, it shows that multi-modal algorithms can be construed that relate manually selected regions of interest to textual data that pertains to the selected brain regions. The evaluation of the system, which focuses on its usability, is the second contribution.

Methods

In “**Semantic Database of Reports**” and “**Image Segmentation**” sections, we outline two proprietary methods that structure free-text reports and MR images by anatomy, respectively. In “**Interpreting Region of Interest**” section, we describe how to infer the intended anatomies from a region of interest manually selected on an MRI slice. In “**Ontology-Based Score Propagation**” section, we describe our ontology-based approach that matches anatomies from MR images and narrative reports of previous cases. “**Graphical Interface to Semantic Database**” section sketches how these technologies can be embedded in a retrieval system.

Semantic Database of Reports

The proposed system has access to a *semantic database* with records of the form:

Document pointer–observation–body location–presence

The first field points to a position in a document. The observation field contains clinically relevant information found at this position, such as findings and diagnoses. The body location field holds the body location of the observation. We shall assume that all body locations are specified according to the vocabulary of the Systemic Nomenclature of Medicine–Clinical Terms (SNOMED CT).⁶ The fourth field is a Boolean indicating if the observation is present or absent.

The semantic database can be created manually, or generated automatically by means of medical natural language processing (NLP) techniques that detect the

⁶ www.ihstdo.org/snomed-ct/

Table 1 Four semantic database records

Document pointer	Observation	Body location	Presence
1.1.5	Neoplasm	Brainstem (180925009)	1
1.1.10	Glioma	Brainstem (180925009)	1
1.1.12	Astrocytoma	Brainstem (180925009)	1
1.2.2	Midline shift		0

grammatical structure of sentences, chunk them in meaningful parts [14], detect if the chunks are negated [15], map the chunks to concepts in a controlled vocabulary [16], and relate the extracted concepts [17–20]. Dedicated medical NLP techniques exploit the fact that the clinical language has a relatively restricted vocabulary and grammar [21]. On some natural language processing tasks in the medical domain, NLP tools perform on a par with human experts [22–24].

In this paper, we use a semantic database with 16,482 records based on a corpus of 860 anonymized neuroradiology reports in the English language. The semantic database was automatically populated by a reasoning engine that extracts and correlates observations and body locations from radiology reports, which is described and evaluated in [25]. As an example, observe the records in Table 1 that are extracted from the following two sentences: *This is suspicious for a neoplasm such as a brainstem glioma or astrocytoma. No significant midline shift is noted.* The semantic database contains 447 unique body structures, which were manually mapped to 226 unique SNOMED CT concepts.

Image Segmentation

We assume that a labeling of the pixels in the MR images in terms of SNOMED CT body structure concepts is given, that is, each pixel is assigned (at most) one body structure concept. Image segmentation algorithms, atlas-based methods [26, 27] and shape-constrained deformable models [28], do this automatically. Such algorithms have been applied to various organs and imaging modalities [28–32]. Advanced extensions have been proposed to segment images of deformed organs (e.g., due to neoplasms or surgical interventions) [33–35].

In our system, we deploy a shape-constrained deformable model that works on 3D, T1-weighted neuro MRIs. In the shape-constrained deformable model approach, one first considers a generic geometric organ structure, a *mesh* model, which models anatomical knowledge. Then, given a stack of images of the appropriate modality, slice thickness, and protocol, the mesh model is transformed to fit the particulars of the image.

The shape-constrained deformable model was made available to us as a black box by colleagues who were not

involved in this research project. An earlier version of the model was described and evaluated in [36]. The quantitative evaluation indicates that it segments the brain accurately, especially the cerebellum, brainstem, and corpus callosum areas. The model used in our system has 69 body structures, mostly in the subcortex, see Fig. 1. We manually mapped the 69 body structures to 46 unique SNOMED CT concepts. Most of the left–right variations are mapped to the same concept; for instance, “left optic nerve” and “right optic nerve” are mapped to optic nerve.

Interpreting Region of Interest

Our algorithm enables the user to select a brain region by drawing a rectangular region of interest (ROI) on an MRI slice. This user interface functionality resembles the functionality of PowerPoint® by which users select graphical elements (e.g., text boxes and figures). In PowerPoint®, all elements that completely fall inside the ROI are selected. However, we do not want to pollute the viewing environment with the labeling of the brain regions. Therefore, unlike in PowerPoint®, the user does not know the exact position of the elements he wants to select on the image. We shall thus refrain from using an all-or-nothing selection strategy and adopt a graded strategy instead that assigns a value in the interval [0,1] to body regions inside the region of interest.

We represent a ROI as the collection of labeled pixels it encloses. For the slice s on which the ROI was drawn, we let:

- $ROI(s) = \#$ labeled pixels on s inside the ROI

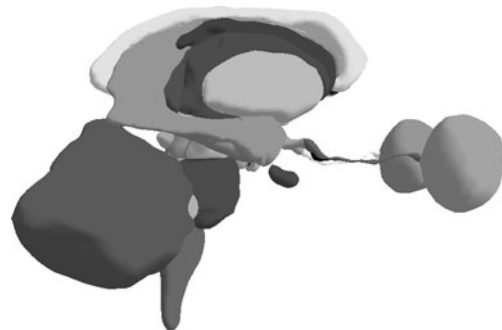


Fig. 1 3D visualization of the generic geometric brain structure

ROI(s) does not necessarily coincide with the total number of pixels inside the ROI, as the slice may contain unlabeled pixels. For a body structure b , we let:

- $\text{Body}(s, b) = \# \text{ pixels on } s \text{ labeled } b$
- $\text{Body} \cap \text{ROI}(s, b) = \# \text{ pixels on } s \text{ inside the ROI labeled } b$

The fourth variable gives the number of pixels of b on the slice on which it is most prominent:

- $\text{MaxBody}(b) = \# \text{ pixels Body}(t, b) \text{ for the slice } t \text{ that maximizes Body}(t, b).$

See Fig. 2 for a diagram, explaining the first three pixel variables.

The four pixel variables are combined in three parameters:

- The portion of b in ROI:
 $\text{PortionBodyInROI}(s, b) = \text{Body} \cap \text{ROI}(s, b) / \text{Body}(s, b)$
- The portion of ROI in b :
 $\text{PortionROIInBody}(s, b) = \text{Body} \cap \text{ROI}(s, b) / \text{ROI}(s)$
- The relative prominence of b on s :
 $\text{RelProminence}(s, b) = \text{Body}(s, b) / \text{MaxBody}(b)$

The first two parameters measure the extent to which the ROI approximates the area of the body structure b at hand. The third parameter is independent of the ROI. The intuition behind this parameter is that if a user wishes to select the medulla, say, he or she will do this on a slice on which the medulla is relatively prominent.

We combine these parameters in the following scoring function:

- $\text{PortionBodyInROI}(s, b)^k \times \text{PortionROIInBody}(s, b)^m \times \text{RelProminence}(s, b)^n$

for some real values k, m , and n .

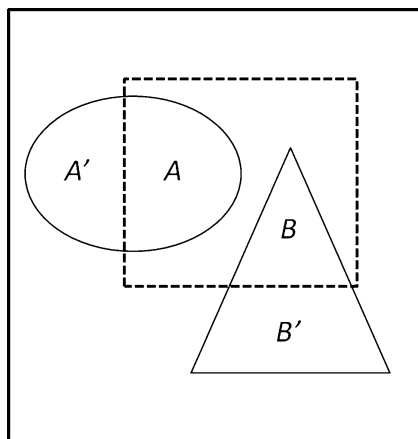


Fig. 2 A slice with two body structures (one oval, one triangular) and a ROI (dotted rectangle). On this slice, $\text{ROI} = A + B$. For the oval body structure, on this slice, $\text{Body} = A + A'$ and $\text{Body} \cap \text{ROI} = A$. If the oval body structure is the largest on this slice, compared to all other slices on which it appears, then its $\text{MaxBody} = A + A'$

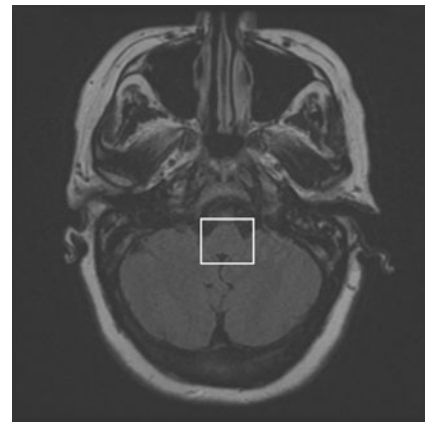


Fig. 3 ROI expressing an interest in the medulla

In informal experiments, we observed that users select body structures by drawing the ROI *around* them, so that the intended body structure is completely contained in the ROI. We concluded that if the user intends to select a particular body structure, he or she will maximize the portion of that body structure in the ROI, that is, the first parameter. Hence, we choose $k=1.3$ greater than one so as to penalize body structures that fall inside the ROI only partially. We observed that the sizes of the ROIs that are drawn around an intended body structure vary significantly. This leads us to set $m=0.4$ quite low, so that the second parameter primarily penalizes body structures that are small with respect to the ROI. We use $n=1$.

For instance, consider the ROI in Figs. 3 and 4 that expresses an interest in the medulla. Besides the medulla, the ROI intersects the cerebellum and the arbor vitae. The values for the pixel variables $\text{Body}(s, b)$, $\text{Body} \cap \text{ROI}(s, b)$, and $\text{MaxBody}(s, b)$ of these three body structures b on the slice s at hand are given in Table 2. The values of the three parameters are computed from these values using the above equations, see again Table 2. In this example, the scoring

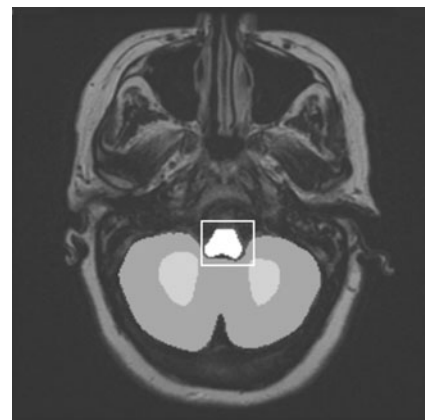


Fig. 4 The slice from Fig. 3 with visualization of the labeling of cerebellum (gray), arbor vitae (light gray), and medulla (white) pixels

Table 2 Pixel variables and parameter values for the ROI in Figs. 3 and 4

	Pixel variables			Parameters			Score
	Body	Body∩ROI	MaxBody	PortionBodyInROI ^{1,3}	PortionROIInBody ^{0,4}	RelProminence	
Medulla	305	305	337	1	0.827	0.905	0.748
Arbor vitae	5,612	179	6,118	0.011	0.668	0.917	0.007
Cerebellum	1,043	7	1,687	0.001	0.183	0.618	0

The number of pixels of ROI is independent of the body structures. On this slice, ROI=491

function correctly points out that the medulla is the intended body structure as it receives the highest score.

In the case of this example, we could have concluded that the medulla was the intended body structure from PortionBodyInROI alone. We experienced that the system is more robust when we use more parameters in the sense that the intended body structures end up with high scores (>0.6) and the unintended body structures with low scores (<0.1).

The scoring function ranks the labels of the body structures recognized by the image segmentation software. Body structures that are not recognized by the image segmentation software receive no score, even though they may be visible on the slice and intended by the user. These structures are scored by means of hierarchical reasoning, detailed in the next section.

Ontology-Based Score Propagation

The mesh model recognizes 46 concepts, which we call *image concepts*. The semantic database, on the other hand, has 227 unique body structure concepts, of which only 18 overlap with the 46 image concepts. The remaining 209 concepts we call *non-image concepts*. This is a quantitative illustration of the semantic gap between image data and the semantic database. Note that this gap is not caused by a lexical mismatching problem as we assume that the body structures in the semantic database as well as the image labels are coded as SNOMED CT concepts.

To bridge the semantic gap, we developed an ontology-based module that estimates the values of the pixel variables for the structures in the semantic database that are hierarchically related to the structures in the mesh model. The scores of these body structures are then computed with the scoring function from the previous section applied to their estimated pixel variables.

We define a spatial model of the body structures in the semantic database by merging SNOMED CT’s Is-a and Part-of relations. According to this spatial model, for instance, pons and tegmental pons are hierarchically related in the sense that pons is a superstructure of tegmental pons and that tegmental pons is a substructure of pons.

The pixel variables of a body structure *b* are estimated iteratively from the pixel variables of the body structures *c* to

which *b* is hierarchically related, assuming the pixel variables of *c* are known, either through direct measurement or estimation:

$$\text{Body}(s, b) = \sum_c W(b, c) \times \text{Body}(s, c)$$

$$\text{Body} \cap \text{ROI}(s, b) = \sum_c W(b, c) \times \text{Body} \cap \text{ROI}(s, c)$$

$$\text{MaxBody}(b) = \sum_c W(b, c) \times \text{MaxBody}(c).$$

In these equations, *W* is a matrix of weight constants *W*(*b*, *c*) that model the spatial proportions between *b* and *c*. We created the weight matrix manually, distinguishing two types of pairs (*b*, *c*): the ones in which *b* is a substructure of *c* and the ones in which *b* is a superstructure of *c*.

For pairs of the first type, the weight *W*(*b*, *c*) estimates the probability that a randomly drawn *c* pixel is a *b* pixel. The weight *W*(*b*, *c*) can be computed from a reference MRI whose pixels are labeled with *b* and *c*:

- $W(b, c) = \# \text{ pixels labeled } b / \# \text{ pixels labeled } c$

Computing the weight matrix in this way is probably more accurate than our estimation approach, but annotation of a reference MRI is very labor and knowledge intensive. For this reason, we preferred our estimation approach.

For pairs of the second type (i.e., *b* is a superstructure of *c*), we set *W*(*b*, *c*) to 1 if *b* is a container concept for its substructures and if its substructures are mesh concepts or are themselves container terms. The pons is an instance of such a container concept. Its constituting substructures are the tegmental and ventral pons, both of which are recognized by the mesh model, see Fig. 5 for a graphical representation. The number of pons pixels on a given slice (the shaded plane) is equal to the number of tegmental pons pixels plus the number of ventral pons pixels on that slice. If not all constituting substructures are in the mesh model or are hierarchically related to a mesh structure, *W*(*b*, *c*) estimates the probability that a randomly drawn *c* pixel is a *b* pixel.

Using the weight matrix, we can estimate the pixel variables of non-image concepts, their parameter values, and finally its score. Thus, the scores of non-image concepts are computed in the same way as we compute the scores of image concepts, except that the pixel variables

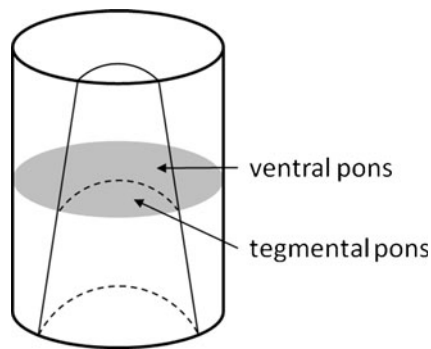


Fig. 5 Graphical representation of the pons (the *cylindrical shape*) and its two constituting substructures, the tegmental and ventral pons

are estimated instead of measured from the slice. This approach should be contrasted to an alternative estimation procedure that assigns to a non-image concept the (weighted) average scores of its hierarchically related structures.

As an example, consider the selection of the pons in Fig. 6. The pixel variables of its constituting parts and the estimates of its own pixel variables are given in Table 3. We see pons receives highest score, that is, the score function correctly infers that the user is more interested in the pons than in its parts (for simplicity, we ignore the other body structures with which the ROI overlaps; the actual scores are very close to the scores given in Table 3). The estimation procedure that takes the weighted average would not be able to assign pons a higher score than its constituting substructures.

Graphical Interface to Semantic Database

The components described above can be combined to define a retrieval system for MRI cases with a graphical interface. Given a manually selected ROI, we compute the scores of the image concepts and estimate the scores of the non-image concepts. We show the list of all body structures to the user, sorted by score, omitting concepts with score 0.

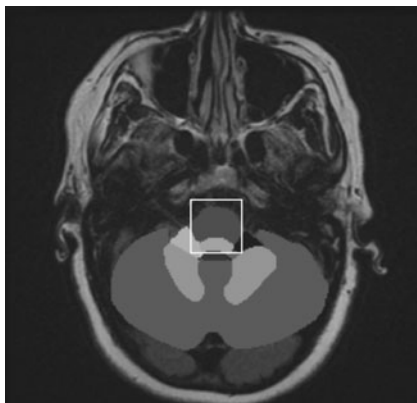


Fig. 6 Selection of the ventral pons (upper structure in the ROI) and tegmental pons (lower structure)

The semantic database is queried for records that pertain to the body structures with score higher than 0. The reports related to the retrieved records are presented to the user. If desired, search can be restricted to records with observations that are present.

In addition to this, we extract the unique body structures and disorders from the retrieved records and present them to the user in two lists, ranked by frequency. The user can select a body structure and/or observation from these respective lists to narrow down the search. Three use cases of this system are illustrated in “[Three Use Cases](#)” section.

Results

We evaluate the system’s ability to transfer a manually selected ROI to a ranked list of body structures. We set up an experiment in which each subject is prompted by a series of brain regions, one at a time, and is asked to carefully draw a rectangle that maximally overlaps the target brain region and minimally overlaps surrounding tissue. During the experiment, one normal MRI of the brain was used in axial view. The subjects were instructed to skip a case if for any reason they found themselves incapable of pointing out the target brain region.

The list of brain regions was compiled as follows: First, we extracted the body location concepts in the semantic database that are hierarchically related in the spatial model to at least one imaging concept. Then, we removed high level concepts such as “brain,” “skin,” and “bone.” Assuming there would be little variation between the selections of left/right mirror image regions (e.g., left and right thalamus), we replaced such pairs with their laterality-independent generalization (e.g., thalamus). The remaining list consisted of 61 brain regions. Sixteen of these regions are mapped to non-image concepts; four of which are a subconcept of an image concept (e.g., uncus of hippocampus is a subconcept of hippocampus); the remaining 12 are superconcepts of an image concept (e.g., pons is a superconcept of ventral pons). Three neuroradiologists from a US-based university hospital participated in the experiment: subject A, fellow neuroradiologist, female, 35 years old; subject B, fellow neuroradiologist, female, 38 years old; subject C, attending neuroradiologist, male, 42 years old.

On average it took each subject roughly 30 minutes to complete the experiment. It was remarked four times that the resolution of the notebook’s monitor on which the experiment was conducted (14.1”, 1,440×900 pixels) is too low to make a reliable selection; 26 times it was remarked by one of the subjects that they would normally take a coronal or sagittal view on the target brain structure. In all cases, the supervisor of the experiment (MS) encouraged the subjects to make their best selection nonetheless. No further difficulties were reported during the experiment.

Table 3 Pixel variables and parameters for the ROI in Fig. 6

	Pixel variables			Parameters			Score
	Body	Body∩ROI	MaxBody	PortionBodyInROI ^{1,3}	PortionROIInBody ^{0,4}	RelProminence	
Tegmental pons	155	155	155	1	0.464	1	0.464
Ventral pons	542	536	927	0.986	0.767	0.585	0.442
Pons	697	691	1,082	0.988	0.847	0.644	0.539

On this slice, ROI=1,056

The subjects’ selections were logged along with the list of suggested brain regions ranked by the scoring function. Thus, 183 data points were collected (3 subjects×61 brain regions). The participants skipped selecting a structure 21 times (11.5% of the total number of 183 data points). The corresponding data points are removed from our analysis. We also removed 24 data points that record a selection of a brain region that does not appear on the slice on which the selection is made, according to the image segmentation module. These data points display a gap between the accuracy of the image segmentation module and the expertise of the subjects and are therefore not suitable for evaluating the scoring function. Thus, 138 data points remain.

One hundred thirty data points (94.2%) record a selection in which the target brain region was in the list of the suggested regions. In 66 data points (47.8%), the target region was ranked first by the scoring function, whereas it was among the top five suggestions in 102 (73.9%) of the data points and among the top ten suggestion in (94.6%) of the data points, see Fig. 7.

Figures 8 and 9 differentiate between the data points with brain regions recognized by the mesh model versus brain regions not in the mesh model, respectively. Ninety-seven

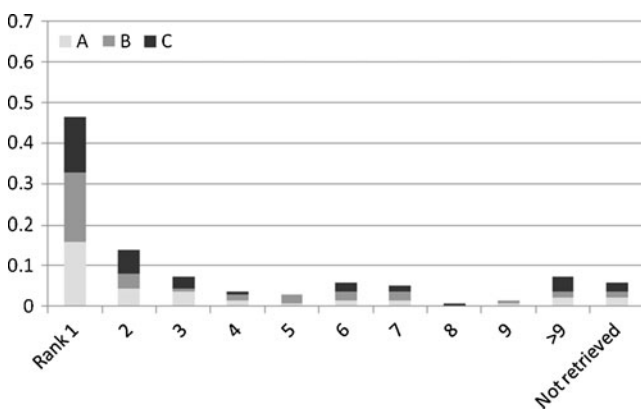


Fig. 7 Histogram of the 138 data points that remain after excluding the skipped data points and the data points in which the subject made a selection on a slice on which the target brain region was not present. For instance, the second bin indicates that in more than 10% of the cases the target region had rank 2 on the ranked list of suggested brain regions. The contribution of the three subjects to each bin is represented with different gray tones

data points address brain regions from the mesh. We see that the brain regions that sit in the mesh are more likely to rank in the top five of suggested brain regions than the regions that do not: 80.4% versus 58.5%. On the other hand, the latter type of brain regions always appear in the list of suggested brain regions, which is not the case for 8.2% of the data points whose brain regions are modeled in the mesh.

Figures 7, 8, and 9 indicate that there is minor variation between the subjects. This was confirmed in another analysis we conducted. Excluding four outliers, we saw that the average distance between the slice on which the subjects make their selection ranges from 1.19 (subjects B and C) to 1.69 (subjects A and C). The average distance between the slices on which the subjects select a body structure and the slices on which that body structure is most prominent, ranges from 1.80 (subject B) to 2.13 (subject A).

Finally, we relate the extent to which data points were retrieved to the size of their body locations. To this end, we divided the data points that were retrieved in four bins: the data points with rank 1 (of which there were 66), data points with rank 2–3 (29), data points with rank 4–7 (24), and data points with rank >7 (14). Two more bins contain the data points that were not retrieved (13) and the 24 data points that were left out, respectively. For each bin, we computed the average MaxBody(*b*) value of the body locations *b* of its data points, see Fig. 10.

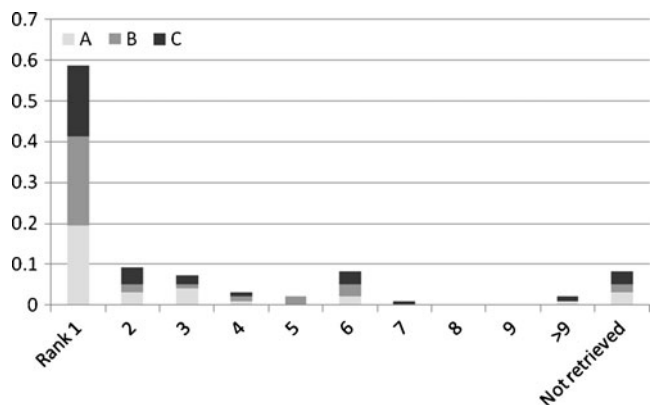


Fig. 8 Histogram of the 97 data points with brain regions mapped to image concepts

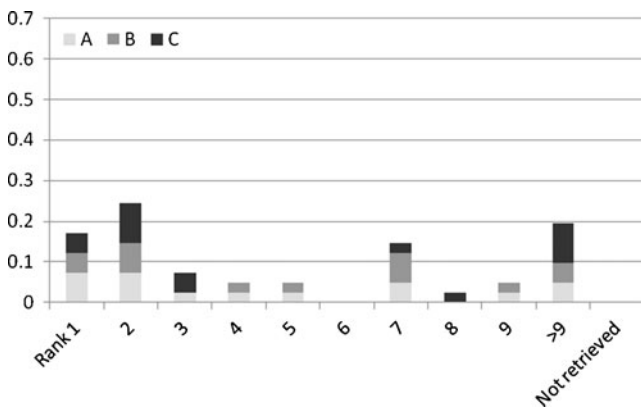


Fig. 9 Histogram of the 41 data points with brain regions mapped to non-image concepts

First, we focus on the 24 data points that were removed because the subject selected it on a slice on which it was not present according to the mesh' labeling. We see that their body locations are relatively small (66.8 pixels on average). The body locations that were not retrieved (208.7 pixels on average) are substantially smaller than the body locations that were (986.9 pixels on average). These findings may reveal a causal relation; for instance, they may indicate that these body locations were not retrieved *because* they are small. Further, we see that the body locations that were retrieved with lower rank (i.e., rank 1–3) are typically smaller than the body locations that were retrieved with higher rank.

Discussion

In this section, we discuss the usability of the system and the extent to which it can be generalized. We first describe

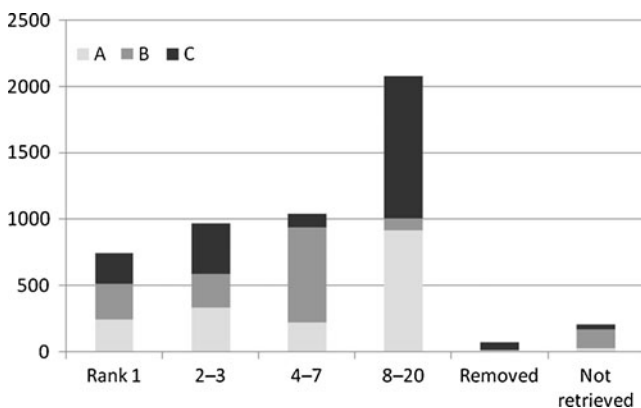


Fig. 10 The average $\text{MaxBody}(b)$ values of the data points' body locations b , broken down in the following bins: data points that were retrieved with rank 1, ranks 2–3, ranks 4–7, rank >7, data points that were not retrieved, and data points that were removed because the subjects' selection was made on a slice on which the body location at hand was not present according to the mesh model

three use cases of the system described in the next section.

Three Use Cases

A simplified rendering of our system illustrates three use cases. In Fig. 11, the user selects the corpus callosum area. The intentions of the user are correctly translated: The corpus callosum ranks the highest in the list of suggested body structures in the top middle pane. Since the user is really interested in the corpus callosum alone, he narrows down his search by selecting the corpus callosum item in the top middle pane. The report previews appear in the bottom pane.

In Fig. 12, we see a configuration of the system in which the user selects the optic nerve region and then clicks "Atrophy." The system correctly concludes that the user is interested in the optic nerve. It ignores the fact that the *right* optic nerve was selected as they are functionally similar. So for the sake of retrieving optic nerve cases, there is no need to discriminate between them. The values between brackets following the observations in the top right pane indicate the number of reports in the semantic database that mention this observation in one or more of the body locations from the top middle pane. Observe that all report previews describe patients with atrophy not just anywhere, but specifically in the optic nerve area.

Finally, in Fig. 13, the user is interested in astrocytomas in the brainstem. To retrieve them, she graphically selects the medulla area. The system correctly concludes that the user may also be interested in the brainstem. The user clicks "Brainstem" and "Astrocytoma" (in any order); only one report is retrieved, as this body location–observation combination is rare in the semantic database. From the report preview, we can see that the system correctly concluded that the astrocytoma appears in the brainstem.

Study Limitations

The pilot study tool recorded the result of the subjects' selections, but not the selections themselves. This is a shortcoming of our study, as it does not allow us to analyze the impact of variations of the proposed system on the algorithm's efficacy. For instance, as was also pointed out by one of the reviewers, this limitation does not allow us to assess the impact of the weighting factors 0.4, 1, and 1.3 of the parameters. Also it does not allow us to zoom in on the contribution of the individual pixel variables. For instance, on the basis of the data that were collected in the pilot study, we cannot investigate if the MaxBody variable alone can be used to fuel the ROI interpretation module.

The goal of the pilot study was to evaluate the ROI interpretation module, not the quality of the mesh model. However, whenever a body location was not retrieved or

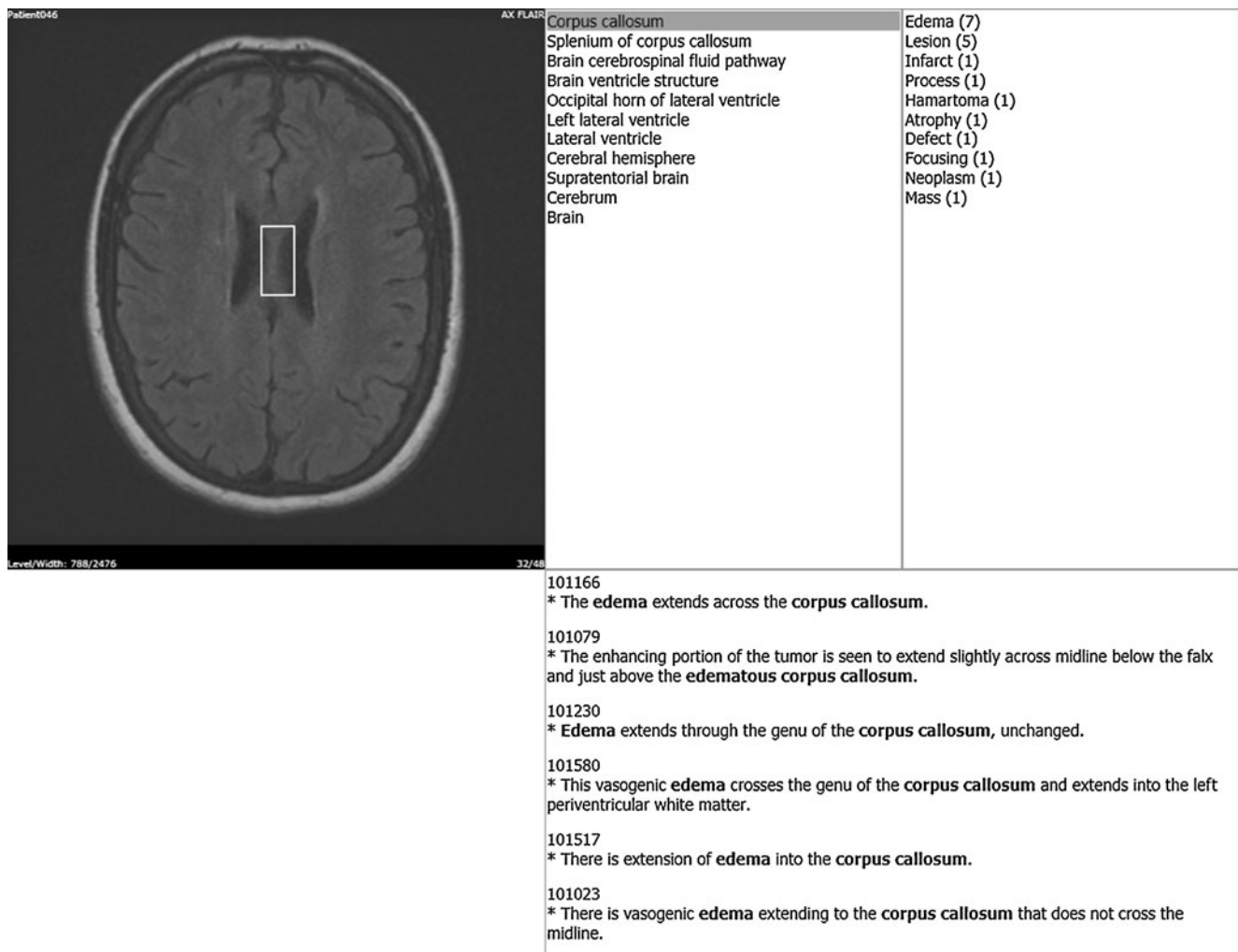


Fig. 11 The user graphically selects the corpus callosum area on the MRI slice. Then, he clicks “Corpus callosum” in the *top middle pane*. The *bottom pane* gives a list report previews

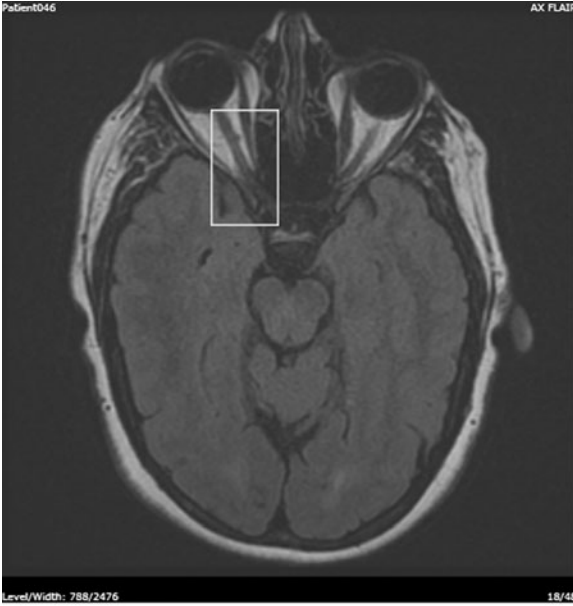
retrieved with high rank (>9, say), our study results do not allow us to say if this is caused by the ROI interpretation algorithm, the mesh model, or by a combination of these factors. Figure 10 lays bare parts of the relation between the algorithm and the annotation of the mesh model, but it is unclear what the causal direction is. A follow-up study may shed light on this relation, for instance, by introducing controlled levels of noise to the labeling of the mesh model. Ideally the three parameters used by the ROI interpretation algorithm would be sufficiently robust that the intended body region can be inferred from the selection, even under considerable levels of noise.

Usability

The aim of the proposed technology is to bring textual knowledge sources closer to the radiology workflow. We stressed that we do not want to presume that the user has a prior understanding of the image segmentation software

(i.e., the regions it recognizes) or the structure of the knowledge source (e.g., its lemmas or indexed body structures). To this end, it is important that the user’s intended body structure is among the given suggestions. Furthermore, whenever the user intends to select a body structure that is not recognized by the system or that is not covered by the semantic database, we still want it to give suggestions that are nearby the intended region. Thus, we believe that recall is the critical factor for browsing external knowledge sources.

Our experiment shows that the proposed region-of-interest interpretation method has a high recall rate: In 94.6% of the cases, the intended brain region was among the top ten suggestions. We further saw that the concepts that are scored by hierarchical reasoning are in the top ten in slightly more than 80% of the cases. The high recall rate guarantees that the user is very likely to find the body structure of interested in the top of the list of the suggested brain regions.



Optic nerve	Swelling (7)
Cranial nerve	Neoplasm (9)
Nerve	Displacement (9)
Right globe	Mass (16)
Eye	Neuritis (80)
	Lesion (7)
	Enlarged (3)
	Meningioma (5)
	Compression (2)
	Atrophy (6)
	Edema (2)
	Glioma (7)
	Leakage of bile (1)
	Neurilemmoma (1)
	Pathology (1)
	Stenosis (2)
	Neuroma (1)
	Fibroma (2)
	Inflammation (3)
	Bowing deformity (1)
	Cholesteatoma (1)
	Focusing (1)
	Fibromatosis (1)
	Pain (11)
	Neuropathy (7)
	Proptosis (5)
	Bell's palsy (1)
	Multiple sclerosis (3)

101601
* Multiple sclerosis with **optic atrophy**.
Orbital MRI Mild T2 hyperintensities are seen within the optic nerves with mild **optic nerve** (right greater than left).

101813
* Findings likely represent relative **atrophy** of the left **optic nerve** since the prior study, which may be attributable to a prior episode of optic neuritis.

101638
* Hemianopsia and **optic atrophy**.

101338
* 75-year-old with the **optic atrophy** and history of breast cancer.

101276
* Severe vision loss, right worse than left, with **optic atrophy** and papilledema.

101627
* Patient with **optic atrophy**, right greater than left.

Fig. 12 The user graphically selects the optic nerve area on the MRI slice. Then, he clicks “Atrophy” in the *right-hand top pane*

The hierarchical reasoning component discloses a significant number of records from the semantic database that would otherwise remain hidden from the user. The image concepts, that is, the concepts recognized by the image segmentation module, pertain to only 416 of the records in our semantic database. The concepts that are hierarchically connected to one or more of the image concepts pertain to 2,029 records. This portion accounts for more than 40% of the records in the semantic database with a body location. We conclude that hierarchical reasoning of body locations improves the recall rate querying the semantic database considerably. This will in turn improve the browsing experience.

Scalability

The proposed system may not be manageable if the granularity of the mesh model and/or the size of the semantic database increase dramatically. In the first case, if the system

is used in combination with a highly granular (brain) model, the number of segments contained in an average region of interest is very high. As a consequence, we might see that the number of regions suggested to the reader may be unrealistically large, especially if the model is not very accurate and/or if the ROI was not drawn carefully enough. This effect can of course be countered by reducing the granularity of the image segmentation software, but in practice access to the software’s source code may be restricted or one is otherwise incapable of modifying it.

An easier solution, leaving the image segmentation software intact, is to reduce the granularity of the range of SNOMED CT concepts to which the images are mapped. We can achieve this by grouping together SNOMED CT concepts, that is, by relabeling all concepts in a group of concepts with the label of its umbrella concept. It is advisable to group together concepts that occupy little space in the geometric model as these are hardest to infer

Medulla Brainstem Brain stem Cerebellopontine angle Infratentorial brain Cerebellum Cerebellar hemisphere Inferior cerebellar peduncle Cerebellar peduncle Cerebellar vermis Brain Cerebellar dentate nucleus	Neoplasm (1) Astrocytoma (1) Glioma (1) Ischemia (1) Mass (1) Focusing (1)
---	--

101638
 * This lesion is suspicious for a neoplasm such as a **brainstem glioma** or **astrocytoma**.

Fig. 13 The user graphically selects the optic nerve area on the MRI slice. Then, she clicks “Brainstem” in the *top middle pane* and “Astrocytoma” in the *top right screen*

from the ROI, see Fig. 10. The rest of the architecture of the proposed information retrieval remains unchanged.

In terms of scaling up the size of the semantic database, the number of *unique* observations and body structures it

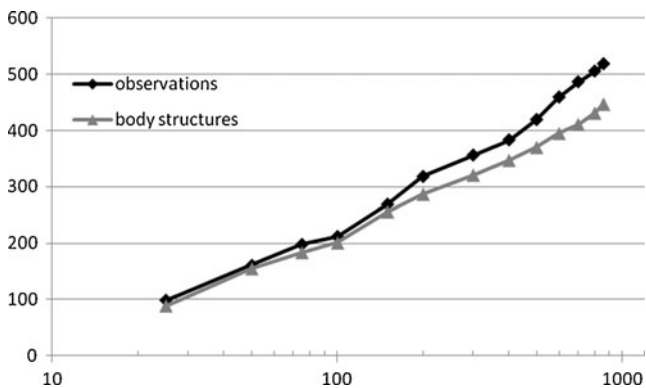


Fig. 14 The number of unique observations and body structures (*y-axis*) appearing in the corpus of radiology reports of varying size (*x-axis*, logarithmic scale)

comprises is critical. It is well known in information retrieval that the number of new unique entities (such as words and phrases) grows sublinearly in the quantity of data that is mined. We found that this law also holds for the number of unique body structures and observations in the semantic database, see Fig. 14. The trendline in Fig. 14 predicts that one million neuroradiology reports contain less than 1,200 unique body structures and less than 1,400 unique observations. We conclude that a corpus of one million neuroradiology reports contains less than three times the number of unique instances than our corpus of less than 1,000 reports. In informal evaluations, the number of unique instances did not negatively affect the usability of the proposed system.

If the number of unique instances reduces the usability of the system nonetheless (e.g., the list of suggestions becomes overpopulated), then one can group together SNOMED CT concepts as described above or categorize them by top-level concepts. In the latter case, the user would first indicate the observation category (e.g., neoplasm, hemorrhage) in a drill-

down menu and then be exposed to the list of suggested observations of the selected category.

Conclusion

We emphasized the importance of workflow state-aware information retrieval and clinical decision support tools, that is, information solutions that use a representation of the radiology workflow to narrow down and anticipate the information needs of the radiologists. Bridging the gap between image and textual data is one of the technological challenges that need to be overcome to realize this workflow state-aware paradigm. To guarantee ease of use, we require that the proposed solution presupposes knowledge neither about the structure of the image processing software that is used to analyze the images nor about the structure of the textual knowledge source to which the solution links.

We propose a system that infers the intended body structure from a manually drawn region of interest. It has a user interface that allows the user to construct a complex anatomical query by selecting the anatomies of interest graphically. The system then correlates the graphically selected anatomies with other anatomies of interest through hierarchical reasoning in SNOMED CT. The ontology-based approach allows us to restrict the range of body structures and observations that are suggested to the user by grouping together concepts. This measure can be taken to manage highly granular mesh models or huge corpora of text data.

In our user experiment, we found that radiologists had no difficulty using the system to select body structures on MRI slices. We found that the system retrieves the intended brain region (and ranks it among the top ten suggestions in almost 95% of the cases) and hierarchically related brain regions. This high recall rate allows the user to browse the external knowledge sources without presuming he or she knows the structure of the image processing software.

We believe that our system paves the way for interesting future research. Ergonomically, we are interested in the way our system integrates in the workflow and how it helps its users to browse external knowledge sources. From an information retrieval point of view, we need to explore what query languages are best able to express the user's information needs. For instance, the system illustrated by the three use cases allows for queries that (only) contain body structures and/or observations. Finally, it is probably most important to investigate the clinical benefits of the proposed system.

Acknowledgments The authors gratefully acknowledge Dieter Geller, Reinhard Kneser, Wim van der Linden, Lina Arbash-Meinl,

Roel Truyen, Paul van der Vet, Lyubomir Zagorchev, Hendrik-Willem van Zutphen, the three annotators, and the anonymous referees.

References

1. Reiner BI, Siegel E, et al: Impact of filmless radiology on frequency of clinician consultations with radiologists. *AJR Am J Roentgenol* 173:1169–1172, 1999
2. Jarvis L, Stanberry B: Teleradiology: threat or opportunity? *Clinical radiology* 60:840–845, 2005
3. Bradley Jr, WG: Off-site teleradiology: the pros. *Radiology* 248:337–341, 2008
4. Maynard CD: Radiologists: physicians or expert image interpreters? *Radiology* 248:333–336, 2008
5. Borgstede JP: Radiology: commodity or specialty. *Radiology* 247:613–616, 2008
6. Boland GW: Teleradiology for auction: the radiologist commoditized and how to prevent it. *J Am Coll Radiol* 6:137–138, 2009
7. Erinjeri JP, Picus D, et al: Development of a Google-based search engine for data mining radiology reports. *J Digit Imaging* 22:348–356, 2009
8. Henderson B, Camorlinga S, et al: A cost-effective web-based teaching file system. *J Digit Imaging* 17:87–91, 2004
9. Wilkinson LE, Gledhill SR: An integrated approach to a teaching file linked to PACS. *J Digit Imaging* 20:402–410, 2007
10. Welte FJ, Kim SC, et al: Incorporation of a formalized emergency radiology curriculum to facilitate population of a MIRC-based digital teaching file. *J Digit Imaging* 23:226–237, 2010
11. Reiner BI: Uncovering and improving upon the inherent deficiencies of radiology reporting through data mining. *J Digit Imaging* 23:109–118, 2010
12. Sinha U, Dai B, et al: Interactive software for generation and visualization of structured findings in radiology reports. *AJR Am J Roentgenol* 175:609–612, 2000
13. Reiner BI: Customization of medical report data. *J Digit Imaging* 23:363–373, 2010
14. Kang N, van Mulligen EM, et al: Comparing and combining chunkers of biomedical text. *Journal of biomedical informatics* 44:354–360, 2011
15. Chapman WW, Bridewell W, et al: A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of biomedical informatics* 34:301–310, 2001
16. Aronson AR, Lang FM: An overview of MetaMap: historical perspective and recent advances. *J Am Med Inform Assoc* 17:229–236, 2010
17. P. J. Haug, S. Koehler, et al: A natural language understanding system combining syntactic and semantic techniques, *Eighteenth Annual Symposium on Computer Applications in Medical Care*, pp. 247–251, 1994
18. P. J. Haug, S. Koehler, et al., Experience with a mixed semantic/syntactic parser, *Nineteenth Annual Symposium on Computer Applications in Medical Care*, pp. 284–288, 1995
19. Sager N, Lyman M, et al: Natural language processing and the representation of clinical data. *J Am Med Inform Assoc* 1:142–160, 1994
20. Friedman C, Alderson PO, et al: A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc* 1:161–174, 1994
21. Friedman C, Johnson SB: Natural language and text processing in biomedicine. In: Shortliffe EH, Cimino JJ Eds. *Biomedical*

- Informatics; Computer Applications in Health Care and Medicine, ch. 8. Springer, New York, 2006, pp 312–343
22. Hripcsak G, Friedman C, et al: Unlocking clinical data from narrative reports: a study of natural language processing. *Ann Intern Med* 122:681–688, 1995
 23. Fiszman M, Chapman WW, et al: Automatic detection of acute bacterial pneumonia from chest X-ray reports. *J Am Med Inform Assoc* 7:593–604, 2000
 24. Hripcsak G, Austin JHM, et al: Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports. *Radiology* 224:157–163, 2002
 25. Sevenster M, van Ommering R, Qian Y: Automatically correlating clinical findings and body locations in radiology reports using MedLEE. *J Digit Imaging* doi:10.1007/s10278-011-9411-0
 26. Fischl B, Salat DH, et al: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neurotechnique* 33:341–355, 2002
 27. Caudra MB, Pollo C, et al: Atlas-based segmentation of pathological MR brain images using a model of lesion growth. *IEEE Trans Med Imaging* 23:1301–1314, 2004
 28. Weese J, Kaus MR, et al: Shape constrained deformable models for 3D medical image segmentation. *Information Processing in Medical Imaging LNCS* 2082:380–387, 2001
 29. Ecabert O, Peters J, et al: Automatic model-based segmentation of the heart in CT images. *IEEE Trans Med Imaging* 27:1189–1201, 2008
 30. Meyer C, Ecabert O, et al: A multi-modality segmentation framework: application to fully automatic heart segmentation. *SPIE* 7259:72594L–72594L-12, 2009
 31. Klinder T, Lorenz C, et al: Automated model-based rib case segmentation and labeling in CT images. *MICCAI* 4792:195–202, 2004
 32. Pekar V, McNutt TR, et al: Automated model-based organ delineation for radiotherapy planning in prostatic region. *SPIE* 6512:65120H-1–65120H-11, 2004
 33. Fletcher-Heath LM, Hall LO, et al: Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. *Artificial Intelligence in Medicine* 21:43–63, 2001
 34. Prastawa M, Bullitt E, et al: Automatic brain tumor segmentation by subject specific modification of atlas priors. *Academic Radiology* 10:1341–1348, 2003
 35. Prastawa M, Bullitt E, et al: A brain tumor segmentation framework based on outlier detection. *Medical Image Analysis* 8:275–283, 2004
 36. Kneser R, Lehmann H, et al: Towards knowledge-enhanced viewing using encyclopedias and model-based segmentation. *SPIE* 7260:72601D–72601D-9, 2009