# Knowledge-based extraction of measurement-entity relations from German radiology reports

**Heiner Oberkampf, Claudia Bretschneider, Sonja Zillner, Bernhard Bauer, Matthias Hammon**

**Angaben zur Veröffentlichung / Publication details:**

Oberkampf, Heiner, Claudia Bretschneider, Sonja Zillner, Bernhard Bauer, and Matthias Hammon. 2014. "Knowledge-based extraction of measurement-entity relations from German radiology reports." In *Proceedings of the 2014 IEEE International Conference on Healthcare Informatics, 15-17 September 2014, Verona, Italy*, edited by Carlo Combi and Shusaku Tsumoto, 149–54. Los Alamitos, CA: IEEE. https://doi.org/10.1109/ichi.2014.27.

# Knowledge-based Extraction of Measurement-Entity Relations from German Radiology Reports


Oberkampf Heiner*† Bretschneider Claudia* Zillner Sonja*
*Siemens AG
Corporate Technology
Munich, Germany

Bauer Bernhard†
†University of Augsburg
Software Methodologies for Distributed Systems
Augsburg, Germany

Hammon Matthias‡
‡University Hospital Erlangen
Department of Radiology
Erlangen, Germany



***Abstract*—A large percentage of relevant radiologic patient information is currently only available in unstructured formats such as free text reports. In particular measurements are important since they are comparable and thus provide insight into the change of the health status over time, for example in response to some treatment. In radiology most of the measurements in reports describe the size of anatomical entities. Even though it is possible to extract measurements and anatomical entities from text using standard information extraction techniques, it is difficult to extract the relation between the measurement and the corresponding anatomical entity. Here we present a knowledge-based approach to extract this relation for size measurements using a model about typical size descriptions of anatomical entities in combination with hierarchical knowledge of existing medical ontologies. We evaluate our approach on two data sets of German radiology reports reaching an F1-measure of 0.85 and 0.79 respectively.**


## I. Introduction

A large percentage of clinically relevant radiologic patient information is represented in unstructured formats such as free text reports. Measurements represent important information documented in reports. On the one hand clinicians measure only things of importance, and on the other hand measurements are comparable and thus provide valuable insights into the change of the patient's health status over time. In radiology we mainly have size measurements describing the spatial extent of anatomical entities. For instance, radiologists measure the size of tumors and metastatic lesions (characteristic changes of parenchyma in different organs, enlarged lymph nodes) in consecutive examinations to evaluate response to treatment.
Currently, radiologists and clinicians need to manually collect measurements from different reports in order to compare the respective values. Sometimes they even need to go back to the original image and measure the entities again.
We present a mechanism to extract measurement-entity relations automatically from text. The aim is to facilitate and speed up the comparison of measurements from consecutive reports. To illustrate the challenge of extracting measurement-entity relations, consider the sentence "Enlarged lymph node right paraaortal below the renal pedicle now 23 mm". There are established Information Extraction (IE) techniques to detect and extract the measurement value "23" and unit "mm". Further ontologies are used in IE, e.g., to recognize and extract ontology concepts representing anatomical entities such as "lymph node", "inferior para-aortic lymph node", "renal pedicle", "kidney" and others. It is however difficult to extract the binary relation between the measurement and the corresponding anatomical entity the measurement is about, i.e. to resolve `isAbout(23 mm, paraaortal lymph node)`. This is especially challenging in long sentences where many different entities occur with the measurement and the measurement is not close to the entity described.
Thus the idea is to abstract from the sentence structure and use a knowledge model containing information about the typical size of anatomical entities in combination with hierarchical information of medical ontologies. Our knowledge-based approach incorporates the following three components:

- **Knowledge Model:** Contains typical size specifications for anatomical entities commonly measured by radiologists such as lesions, lymph nodes or organs. The model is linked to existing medical ontologies.
- **Annotator:** Responsible for the linguistic preprocessing of the texts. In particular sentence splitting, measurement extraction and named entity recognition (annotation of ontology concepts).
- **Resolution Algorithm:** Builds on top of the knowledge model and the annotator. The measurement-entity relations are resolved, combining the output of the annotator with the knowledge model and also the structure of the ontology used for annotation.

The remainder is organized as follows: In Section II we give an overview of measurements typically found in radiology reports. Then we present the knowledge model in Section III and describe the annotator in Section IV. The main contribution is the relation extraction described in Section V where we present the resolution algorithm. Our approach is evaluated in Section VI before we give an overview of related work in Section VII and conclude with a discussion in Section VIII.

## II. Measurements in Radiology

In radiology reports we mainly have measurements specifying the size of anatomical entities in terms of volume, area or length. Our analysis of radiology reports yielded that *length* measurements are most frequent, so we concentrate on them. A length measurement describes the extension of an anatomical entity or structure into one dimension. For instance "mediastinal lymph node with diameter 21 mm", "hepatic duct dilated up to 1.1 cm”, “wall of gallbladder 12 mm” or “liver with anterior-posterior diameter of 15.5 cm”. Simplified a length measurement comprises the following components:

- **Anatomical entity:** lymph node, liver, wall of gall bladder, lesion etc.
- **Value specification:** 21 mm, 1.1 cm etc.
- **Measured quality (optional):** width, diameter, anterior-posterior diameter, height, thickness etc.

Two or three length measurements might be grouped together to describe the extension of a certain entity along orthogonal axes: e.g. “lesion in segment 7/8 with 1.4 cm x 1.1 cm” or “spleen with 3.8 x 9 x 10.5 cm not enlarged.” For organs these measurements are mainly taken in parallel to the main body axes to specify height, width or depth. For smaller entities the axes are mostly defined by the form of the entity itself: For the evaluation of tumors or metastatic lesions in computed tomography (CT) or magnetic resonance imaging (MRI) the radiologist firstly measures the longest diameter in the axial slices and then the longest perpendicular extension [1]. This form of standardized measuring procedure allows to compare measurements from consecutive examinations.

## III. KNOWLEDGE MODEL

The medical literature contains much information about the normal size of anatomical entities as well as descriptions of typical abnormal or pathological structures like, e.g., cysts, lesions or enlarged lymph nodes. The following types of size specifications are commonly used:

- **Interval:** e.g. “anterior-posterior diameter of liver 10-13 cm” or “enlarged lymph node 1-5 cm”.
- **Normal value with deviation:** e.g. “truncus pulmonalis: $1.4$ cm $\pm 0.4$ cm”.
- **Upper bound:** e.g. “normal lymph node $\leq 1$ cm”.
- **Lower bound:** e.g. “aorta diameter $\geq 4$ cm at root”.

The main function of these specifications is to define which size of some anatomical entity is considered to be normal and which not. Obviously lower bounds are problematic for that purpose since they include unreasonably high values (e.g. an aorta diameter of 10 cm is *not* normal). In cases where we could find only a lower bound, we asked a radiologist to define a clinically reasonable upper bound to avoid this problem. Upper bounds are interpreted as the interval $[0, x]$. Thus the resulting model contains only interval specifications. We say that a value $x$ is *in the range* of some size specification if $x$ is contained within the respective interval.

### A. Representation of the Knowledge Model

The knowledge model is formalized in a semantic model represented in RDFS [2]. Similar to measurements, we have three components as shown in Figure 1[1]: the anatomical entity, the quality described and the size specification. The knowledge model is based on our Model for Clinical Information (MCI) described in [3], which is an information model based on well defined upper ontologies from the OBO Foundry [4]. It is used in combinations with Radiological Lexicon RadLex [5] and the Foundational Model of Anatomy (FMA) [6] to reference anatomical entities. To represent the quality described by a specification (e.g. diameter, length, thickness etc.) we use concepts from the Phenotypic quality ontology (PATO) [7]. For size specification we take concepts from the Ontology for Biomedical Investigations (OBI) [8].

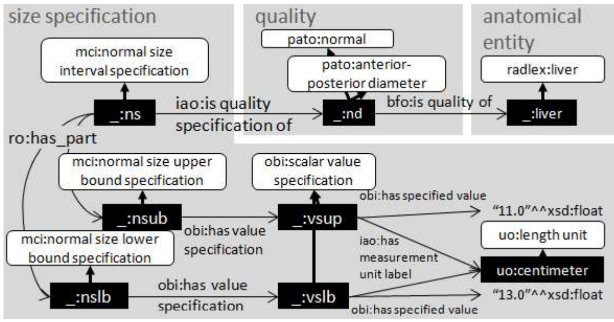


Fig. 1. Example representation of normal interval specification “the anterior-posterior diameter of a liver is normally 11-13 cm”. The white rounded rectangles are classes, the black rectangles instances. Thick arrows are rdf:type relations.

The current model was created manually using a clinical book about normal findings [9] and Radiopedia [10]. It contains 50 size specifications about 38 different anatomical entities of clinical interest and typically measured in radiological examinations. For instance we have size descriptions for spleen, kidney, gallbladder, pancreas, lymph nodes, aorta, lesion, bile ducts etc. Size specifications might depend on the patient (e.g. on age and gender) or at least on some reference population. Inclusion of this information is ongoing work.

## IV. ANNOTATOR

The component that takes care of the extraction of relevant information from unstructured clinical texts is the *annotator*. Information Extraction (IE), as a task of Natural Language Processing (NLP), is a technique to find important information pieces in unstructured texts and to extract them as structured information [11]. For instance, IE is used to detect semantic entities such as date values, names, measurements, etc., in texts. Our annotator extracts two main information pieces: measurements as well as ontology concepts such as anatomical entities, morphological structures and clinical findings.

While measurements are extracted using pattern-based techniques (i.e. through detection of utterances by using a set of regular expressions that express predefined combinations of measurement values and units), we use ontology-based IE techniques for the extraction of anatomical entities. The latter technique facilitates the controlled vocabulary of an ontology to map the ontological concepts to the corresponding words in the text. This task is also referred to as named entity recognition or *semantic annotation*.

For this work we use RadLex, which contains about 38,000 anatomical entities and other concepts like clinical findings or imaging observations relevant for the radiology domain. More precisely we map the concepts listed in the ontology to the

[1]For better readability we write any concepts by prefixed annotation properties, i.e. we write ‘radlex:liver’ instead of radlex:RID58.

words in clinical reports. There are freely available annotators like, e.g., the annotator of the BioPortal [12] for annotation of text with concepts from biomedical ontologies. However, we need a more specific annotator tailored for German texts.

### A. Functional Scope of the Annotator

There are three features that our annotator supports:

- It detects multiword terms independent from the ordering of the individual tokens.
- The annotator respects the sentence boundaries and maps multiword terms only when they occur within these boundaries.
- It recognizes inflected forms of ontological concepts in the text such as detection of plural form or other grammatical inflections based on stemmed forms.

### B. Technical Realization of the Annotator

The implementation of the annotation pipeline builds on the UIMA framework[2]. The annotator itself is an adapted version of the UIMA Concept Mapper, which annotates texts preprocessed by an own medical text preprocessing pipeline including sentence splitting and tokenization, and subsequently normalizes and stems the medical language tokens. The output of the annotator is a set of measurement, anatomical entities' and other ontology concepts' annotations in the form of RDF triples. Thus annotations can be easily integrated with the knowledge model for subsequent analysis.

## V. Resolution Algorithm

The resolution of measurement-entity relations is based on the annotations, the knowledge model and the structure of RadLex. We illustrate the different steps of the resolution algorithm along the *example sentence*: "Enlarged lymph node right paraaortal below the renal pedicle now 23 mm" (original in German: "Vergrößerter Lymphknoten rechts paraaortal unterhalb des Nierenstiels jetzt 23 mm."). The algorithm has to extract that "23 mm" specifies the size of a lymph node (or better a paraaortal lymph node), i.e. the relation `isAbout(23 mm, paraaortal lymph node)`. Using the annotator described in the previous section the following set of RadLex annotations are obtained: 'radlex:lateral aortic lymph node','radlex:right', 'radlex:lymphadenopathy', 'radlex:enlarged', 'radlex:lymph node', 'radlex:inferior paraaortic lymph node', 'radlex:renal pedicle', 'radlex:inferior', 'radlex:paraaortic' and 'radlex:kidney'.
For each sentence containing measurements we analyze the set of annotations to infer the anatomical entity the measurement is about. The algorithm relies on eight steps: First, we *check whether the sentence in our scope* (1). For all sentences in scope we *filter* (2) and *extend* (3) the set of annotations using the ontology structure of RadLex. We then *create a spanning tree* (4) covering the annotations, *attach corresponding size specifications* (5) from the knowledge model to it, *compare them to the measurement values* (6) and *compute a ranking* (7) of entities. Finally *the best entity is selected* (8) in dependence of some threshold criteria.

[2]http://uima.apache.org/

*1) Scope:* The classification of sentences being in or out of our scope is a preprocessing step before we get to the core of the resolution algorithm. Most of the sentences of radiology reports contain one or two measurements, but we also have sentences with up to 14 measurements. For simplicity the current algorithm resolves only one measurement-entity relation per sentence. Thus we restrict the scope of the algorithm to sentences with one or two measurements. Sentences with more than two measurements are *out of scope* of the current implementation. The underlying assumption is that sentences with more measurements can be resolved using the same technique looping over all contained measurements. For sentences with two measurements we require that they represent a size comparison of the same entity to different times. For instance in "Spleen now with 10.5 x 4.5 cm slightly smaller than in previous examination with 13.3 x 6.7 cm." both measurements are about the same anatomical entity (the spleen), while in "Splenomegaly with 23.0 x 14.5 x 8.5 cm and approx. 1.0 cm lesion." the measurements are about different entities. Using simple heuristic we check whether both measurements have the same dimension and not too different values. Further certain keywords like 'previous examination' or 'progressive' have to occur in the sentence.

*2) Filter Annotations:* Not all annotations are good candidates for the measurement-entity resolution: For instance, while 'radlex:lymph node' is a good candidate 'radlex:inferior' is not. In RadLex we do expect to find good candidates only under 'radlex:anatomical entity', 'radlex:imaging observation' and 'radlex:clinical finding'. Annotations under these classes are referred to as *relevant annotations*. All other annotations, for instance those under 'radlex:imaging modality', 'radlex:procedure' or 'radlex:Radlex descriptor' are filtered out. This removes 'radlex:right', 'radlex:enlarged', 'radlex:inferior', and 'radlex:paraaortic' from the list of annotations of the example sentence.

*3) Extend Annotations:* In RadLex, clinical findings are linked by the property radlex:Anatomical_Site to respective anatomical entities. E.g. 'radlex:hepatomegaly' is linked to 'radlex:liver' by this property. We take advantage of these links and extend the initial set of annotations. I.e., for each initial annotation we query RadLex for related anatomical entities add them to the set of annotations. Thus we add annotations for of anatomical concepts the annotator could not detect directly.

*4) Create Spanning Tree:* Using the set of relevant annotations we create a minimal spanning tree from the RadLex subclass hierarchy. The spanning tree for the annotations of the example sentence is shown in Figure 2. The spanning tree is represented in RDF, like the knowledge model.

*5) Attach Size Specifications:* For all concepts of the spanning tree we check the knowledge model for size information about the respective concept and attach it to the spanning tree. Regarding our example we find size specifications for 'radlex:kidney' and 'radlex:lymph node'. Using the hierarchy of the ontology in form of the spanning tree has two advantages: Firstly, we can propagate size information down to subclasses. Thus, as shown in the next step, each size assertion in the knowledge model implicitly applies to *many* concepts. Secondly, the spanning tree with subclass paths enhances the chance to find a matching concept in the knowledge model.

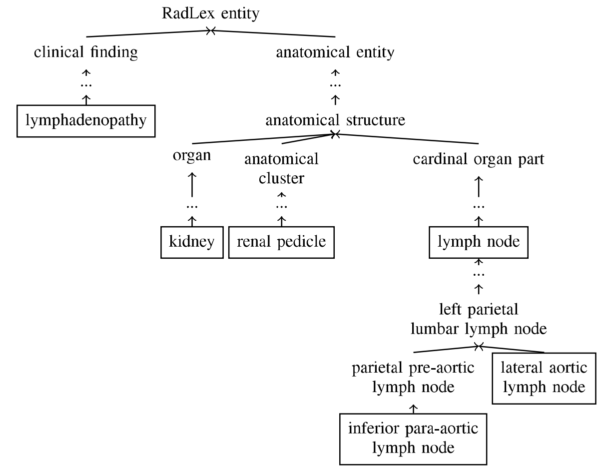


Fig. 2. Minimal spanning tree for RadLex annotations relevant for relation resolution for the sentence "Enlarged lymph node right paraaortal below the renal pedicle now 23 mm". Arrows represent a subclass relationship, dots indicate that a subclass path is omitted.

*6) Compare Size Specifications:* For all concepts $c$ of the spanning tree with size specifications we compute a comparison value indicating how well the measurement value $x$ fits into typical size range $[m, M]$ of $c$:

$$compValue(c, x) := \begin{cases} \frac{m-x}{x} & , x < m. \\ 0 & , x \in [m, M]. \\ \frac{x-M}{M} & , x > M. \end{cases} \quad (1)$$

I.e. if the measurement value is within the range we assign a comparison value 0, otherwise we compute a value $> 0$. If we have multiple attached size specifications, the lowest (best) comparison value is saved. Other nodes of the spanning tree then get the comparison value of the closest superclass assigned if available. E.g. 'radlex:lymph node' has comparison values $1.3$ ("normal lymph nodes are $[0, 1]$ cm") and $0$ ("enlarged lymph nodes are $[1, 5]$ cm") so we assign the value $0$. 'radlex:kidney' gets comparison value $0.73$ ("anterior-posterior diameter of kidney is normally $4$ cm"). Then information is propagated: 'radlex:lateral aortic lymph node' and 'radlex:inferior para-aortic lymph node' get comparison value $0$ assigned since 'radlex:lymph node' is their closest superclass with size specification. The knowledge model does not cover 'radlex:lymphadenopathy' and 'radlex:renal pedicle', so these concepts will not be further processed.

*7) Compute Ranking:* The final ranking value includes also the position of the concept $c$ within the RadLex hierarchy:

$$rankValue(c) := compValue(c, x) + \frac{1}{depth(c)} \quad (2)$$

Thus, in case of equal comparison values more special concepts (deeper in the hierarchy) are preferred. This again shows the advantage of using the ontology hierarchy.

*8) Select:* Finally we select the RadLex concept with the lowest ranking value if it is below some predefined threshold. In our example case the concept 'radlex:inferior para-aortic lymph node' is selected and the resulting relation `isAbout(23 mm, inferior para-aortic lymph node)` is obtained. Not selecting any entity can have two reasons: Either the knowledge model does not contain information about the annotated entities or all entities have bad comparison values and thus do not fit. Using the threshold, we avoid to select wrong entities and thus enhance precision.

## VI. EVALUATION

We evaluate our approach on two different data sets of German radiology reports. We analyzed the sections for finding descriptions and assessment. Both data sets are described in the following subsection, before we present the evaluation results.

### A. Data Sets

*1) Lymphoma Patients:* This data set consists of 2584 German radiology reports (27 different readers) of 377 lymphoma patients. The imaging modality was mainly computed tomography (CT), but also magnetic resonance imaging (MRI) and ultrasound (US). The inspected body regions were mainly abdomen, thorax and head, but includes also various other regions from the whole body. In total the reports contain 5200 sentences with 6790 length measurements. The lymphoma data set was used during the development of our knowledge model.

*2) Diverse Internistic Patients:* This dataset consists of 6007 German radiology reports (27 different readers), where imaging modality was computed tomography (CT). The reports contain 14225 sentences with 22063 length measurements.

### B. Evaluation of Resolution Results

As explained above in the current implementation resolution is restricted to sentences in scope as defined in Section V-1. In the lymphoma data set 8.25% and in the internistic data set 16.15% of all sentences with measurements are out of scope. These sentences are not regarded in the following analysis, since we did not attempt to resolve them.

*1) Evaluation schema:* Resolution results were evaluated by a radiologist based on a randomly selected subset of 500 sentences for each data set. We generated a list containing all sentences and the resolved entity (if available). Evaluation was done according to the following schema:

- **correct:** The entity resolved is exactly what the measurement of the sentence is about. The radiologist cannot find a better entity.
- **(correct):** The entity resolved is correct however it could be more specific (e.g., "lymph node in jaw angle with 1 cm" with resolution to 'lymph node') or the measurement is *not only* about the resolved entity (e.g., "Mediastinal and axillary lymph nodes up to 1.5 cm." with resolution 'mediastinal lymph node').
- **false:** The resolved entity is false or the entity was not resolved, but the radiologist can identify the correct entity within the sentence.
- **unresolvable:** The sentence does not allow a resolution (e.g. "The biggest is now 2.7 cm.") or the measurement does not represent a size description

(e.g. “Tracheal tube 4 cm above the carina.”) and the algorithm did not resolve to a false entity.

Sentences evaluated as **correct** and **(correct)** are regarded as *true positive*, **unresolvable** as *true negative* (since the algorithm did not resolve to a false entity). Sentences evaluated as **false** are regarded as *false positive* if the algorithm resolved to a false entity and as *false negative* if nothing was resolved.

*2) Results:* As shown in Table I our approach has a recall of 0.87 and 0.79 respectively for the sentences within our scope and a precision of 0.84 and 0.79 respectively. Accordingly the F1-measure is 0.85 and 0.79. As expected we see a drop in recall and precision for the internistic dataset since we used only the lymphoma data set during the development of the knowledge model. As explained below in the evaluation of the annotator, both, precision and recall, can be further enhanced with small adaptations to RadLex.

TABLE I. EVALUATION RESULTS FOR 500 RANDOMLY SELECTED SENTENCES FOR EACH DATA SET.

| data set | recall | precision | F1-measure |
|---|---|---|---|
| lymphoma | 0.8698 | 0.8389 | 0.8540 |
| internistic | 0.7904 | 0.7864 | 0.7884 |

*3) Evaluation by Anatomical Entity:* In sum we resolve relations to 86 different anatomical entities. Figure 3 illustrates the evaluation results for 55 of these entities which occurred in the randomly selected subset. That is, in using the RadLex hierarchy we extend the coverage of the knowledge model to more concepts. For instance size descriptions for ‘lymph node’ apply for all 250 subclasses. Figure 3 shows that we achieve good results for lesion and various lymph nodes, even though resolution could be more precise. This is due to the fact that RadLex has a relatively detailed subclass hierarchy for lymph nodes, however not all of the subclasses have German labels. Thus in many cases the annotator detects only ‘lymph node’ even though the sentence contains a more precise description. We notice that false resolution is concentrated mostly on mass related entities. Further we see that the algorithm has problems with aorta, liver, and cyst, but performs very good for spleen, renal cyst or ascending aorta. From all sentences 25 from the lymphoma dataset and 19 from the internistic dataset were classified as unresolvable as explained above.

*4) Evaluation of Annotator:* The annotator was able to annotate multiword terms and recognize inflected forms. This was of high relevance since ontology concept labels often consist of many terms while in reports these terms do not occur in the same order and form. For instance the sentence “Some enlarged lymph nodes up to 1.8 cm in mediastinum” gets the annotation ‘radlex:mediastinal lymph node’ which is correct. The annotator creates $0-22$ relevant annotations per sentence (average 2.93 ). In general more annotations enhance recall, however wrong annotations reduce precision. So the quality of the annotator strongly depends on the quality of the ontology. Using RadLex brings the following two issues:

- **Missing annotation:** RadLex does not cover all anatomical entities occurring in reports (and only about 25% of all RadLex concepts have German labels). Thus the correct entity gets not always annotated: About 6.59% of all sentences have no relevant annotation. Further, in 50% of the false resolutions, the correct entity was not annotated. Thus the resolution algorithm had no chance to pick the right entity.
- **Wrong annotation:** The extensive usage of synonyms in RadLex for mass related classes leads to wrong annotations and in consequence to false resolutions. For instance ‘radlex:breast mass’ has synonyms ‘mass’, ‘nodule’, ‘lesion’, ‘nodular enhancement’ and ‘area of enhancement’. Thus each time a ‘mass’ or ‘lesion’ is mentioned in a report the annotator assigns ‘radlex:breast mass’ and then the resolution algorithm falsely resolves to ‘breast mass’.

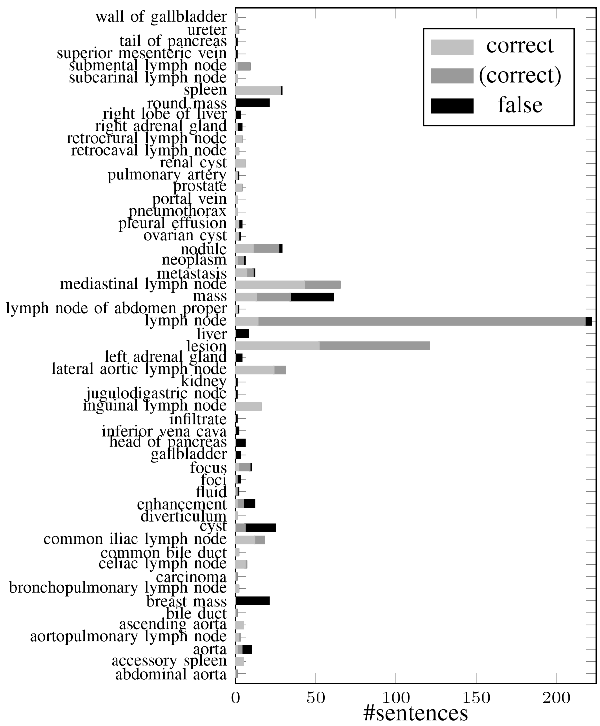


Fig. 3. Resolution results by resolved anatomical entities based on 500 randomly selected sentences from the lymphoma data set and 500 from the internistic data set respectively.

The quality of the vocabulary provided by the ontology is critical for our approach. Simple adaptations of the ontology can significantly enhance precision and recall results.

## VII. Related Work

Medical text analytics has been conducted in the context of the cTAKES project [13], which is also based on the UIMA framework, and MedLee [14]. While cTAKES approach to measurement detection is based on finite state machines, the MedLee system is rule-based. Relations between medical entities like conditions (e.g. diseases or disorders), symptoms and indicated treatment are extracted e.g. in [15] or [16]. The vocabulary of medical ontologies like those from the Unified Medical Language System are commonly used for the extraction of medical entities (e.g. in [16]). Existing approaches for relation extraction mostly use machine learning (ML) techniques in combination with linguistic parse trees [13].

While we concentrate on measurement-entity relations these methods are more general. However we could not find evaluations of these approaches for the extraction of measurement-entity relations. Within the i2b2 Relations Challenge in 2010 the supervised ML system described in [17] showed the best results for the extraction of relations between medical problems, treatments and tests with an F-measure of 0.74. The i2b2 Challenge on Temporal Relation extraction showed that ML outperforms other systems in *detection* of events, while hybrid approaches were better at *classification* of temporal relations [18]. Similarly in [16] a hybrid approach (combine relation patterns and ML) yields an F-measure of 0.94 for the extraction of relations between disease and treatment. It is however difficult to compare these results to the specific relation extracted with our approach.

## VIII. Conclusion

We demonstrated that a rather simple knowledge-based approach is able to resolve measurement-entity relations. In combination with a large ontology, a small knowledge model has a great effect. We expect that our results can be further enhanced using more than one ontology and extending the knowledge model. We further emphasize that even though here we concentrated on length measurements, our approach can be applied to other measurements (e.g. density) as well. But a pure knowledge-based approach has two limitations. Firstly, intervals overlap: An enlarged lymph node might be 3.2 cm - the aorta diameter as well. That is why our model fails for certain concepts, which are often used for location description, but rarely measured itself, such as the aorta. The overlap is especially problematic for concepts with a wide variation in size like lesions. Secondly, in reports we often have deviations from normal range. In "Splenomegaly with 23.0 x 14.5 x 8.5 cm." the size measurement of the spleen is *above* the normal range, but within the typical range for an enlarged spleen. We suggest that through incorporation of the measurements' context in the resolution algorithm and the usage of additional classifiers these limitation can be overcome.

### A. Outlook

In future work we will address the mentioned limitations: (1) Adjust the comparison function, in using annotations such as 'enlarged', 'thickened', 'normal' etc., which are currently filtered out. (2) Include other classifiers such as the measurement-entity distance within the sentence. (3) Use context information: E.g. DICOM headers of reports contain information about the examined body regions. (4) Enhance annotation coverage by adding more German labels to RadLex. A promising next step is the usage of the resolved relations. We can link findings from consecutive reports and thus make the change transparent. In general, extraction and structured presentation of relevant findings described in unstructured free text reports potentially facilitates the radiological reading process. In combination with automatic detection and segmentation methods it provides the basis for automatic navigation between corresponding image and text information. Investigation of these issues is subject of our ongoing work.

## Acknowledgment

This work was partially supported by EIT ICT Labs in the MedicalCPS activity.

## References

[1] E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, and J. Verweij, "New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)," *European Journal of Cancer*, vol. 45, no. 2, pp. 228–247, 2009.

[2] (2014) RDF Schema 1.1 - W3C Recommendation 25 February 2014. [Online]. Available: http://www.w3.org/TR/rdf-schema/

[3] H. Oberkampf, S. Zillner, B. Bauer, and M. Hammon, "An OGMS-based Model for Clinical Information (MCI)," in *ICBO*, 2013, pp. 97–100.

[4] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall *et al.*, "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration," *Nature biotechnology*, vol. 25, no. 11, pp. 1251–1255, 2007.

[5] (2014) Radlex version 3.11. [Online]. Available: http://radlex.org/

[6] C. Rosse and L. V. M. Jr., "The Foundational Model of Anatomy Ontology," *Esophagus*, vol. 2007, pp. 59–117, 2007.

[7] (2014) Phenotypic quality (PATO). [Online]. Available: http://purl.obolibrary.org/obo/pato.owl

[8] (2014) Ontology for Biomedical Investigations (OBI). [Online]. Available: http://purl.obofoundry.org/obo/obi.owl

[9] T. B. Möller and E. Reif, *CT- und MRT-Normalbefunde*. Thieme, 1998.

[10] (2014) Radiopedia. [Online]. Available: http://radiopaedia.org/

[11] S. M. Meystre, G. K. Savova, K. C. Kipper-Schuler, and J. F. Hurdle, "Extracting information from textual documents in the electronic health record: a review of recent research." *Yearbook of medical informatics*, pp. 128–44, Jan. 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18660887

[12] P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, and M. A. Musen, "BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications." *Nucleic acids research*, vol. 39, no. Web Server issue, pp. W541–W545, 2011.

[13] G. K. Savova, J. J. Masanz, P. V. Ogren, J. Zheng, S. Sohn, K. C. Kipper-Schuler, and C. G. Chute, "Mayo clinical text analysis and knowledge extraction system (cTAKES): architecture, component evaluation and applications." *Journal of the American Medical Informatics Association : JAMIA*, vol. 17, no. 5, pp. 507–513, Sep. 2010. [Online]. Available: http://dx.doi.org/10.1136/jamia.2009.001560

[14] C. Friedman, "Towards a comprehensive medical language processing system: methods and issues." in *Proceedings of the AMIA annual fall symposium*, American Medical Informatics Association, 1997, p. 595. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2233560/pdf/procamiaafs00001-0630.pdf

[15] A. Roberts, R. Gaizauskas, and M. Hepple, "Extracting clinical relationships from patient narratives," in *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, ser. BioNLP '08. Stroudsburg, PA, USA: Association for Computational Linguistics, 2008, pp. 10–18. [Online]. Available: http://dl.acm.org/citation.cfm?id=1572306.1572309

[16] A. Ben Abacha and P. Zweigenbaum, "Automatic extraction of semantic relations between medical entities: a rule based approach." *Journal of biomedical semantics*, vol. 2 Suppl 5, no. Suppl 5, p. S4, Jan. 2011. [Online]. Available: http://www.jbiomedsem.com/content/2/S5/S4

[17] B. Rink, S. Harabagiu, and K. Roberts, "Automatic extraction of relations between medical concepts in clinical texts." *Journal of the American Medical Informatics Association : JAMIA*, vol. 18, no. 5, pp. 594–600, 2011.

[18] W. Sun, A. Rumshisky, and O. Uzuner, "Evaluating temporal relations in clinical text: 2012 i2b2 Challenge." *Journal of the American Medical Informatics Association : JAMIA*, vol. 20, no. 5, pp. 806–13, 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23564629