

Meta-generalis: A novel method for structuring information from radiology reports

Flavio Barbosa¹; Agma Jucci Traina²; Valdair Francisco Muglia¹

¹Universidade de Sao Paulo Ribeirao Preto School of Medicine, Internal Medicine, Ribeirao Preto, Sao Paulo, Brazil; ²Universidade de Sao Paulo Instituto de Ciencias Matematicas e de Computacao, Sao Carlos, SP, Brazil

Keywords

Radiological report, free-text, structured report, ontology, terminology.

Summary

Background: A structured report for imaging exams aims at increasing the precision in information retrieval and communication between physicians. However, it is more concise than free text and may limit specialists' descriptions of important findings not covered by pre-defined structures. A computational ontological structure derived from free texts designed by specialists may be a solution for this problem. Therefore, the goal of our study was to develop a methodology for structuring information in radiology reports covering specifications required for the Brazilian Portuguese language, including the terminology to be used.

Methods: We gathered 1,701 radiological reports of magnetic resonance imaging (MRI) studies of the lumbosacral spine from three different institutions. Techniques of text mining and ontological conceptualization of lexical units extracted were used to structure information. Ten radiologists, specialists in lumbosacral MRI, evaluated the textual superstructure and terminology extracted using an electronic questionnaire.

Results: The established methodology consists of six steps: 1) collection of radiology reports of a specific MRI examination; 2) textual decomposition; 3) normalization of lexical units; 4) identification of textual superstructures; 5) conceptualization of candidate-terms; and 6) evaluation of superstructures and extracted terminology by experts using an electronic questionnaire. Three different textual superstructures were identified, with terminological variations in the names of their textual categories. The number of candidate-terms conceptualized was 4,183, yielding 727 concepts. There were a total of 13,963 relationships between candidate-terms and concepts and 789 relationships among concepts.

Conclusions: The proposed methodology allowed structuring information in a more intuitive and practical way. Indications of three textual superstructures, extraction of lexicon units and the normalization and ontologically conceptualization were achieved while maintaining references to their respective categories and free text radiology reports.

Correspondence to:

Valdair Muglia, MD., Ph.D.
Universidade de Sao Paulo Ribeirao Preto School of
Medicine
Internal Medicine
Av Bandeirantes 3900
Campus Monte Alegre
Ribeirao Preto
Sao Paulo 14049900
Brazil
Email: fmuglia@fmrp.usp.br

Appl Clin Inform 2016; 7: 803–816

<http://dx.doi.org/10.4338/ACI-2016-03-RA-0037>

received: March 31, 2016

accepted: July 22, 2016

published: August 24, 2016

Citation: Barbosa F, Traina AJ, Muglia VF. Meta-generalis: A novel method for structuring information from radiology reports. *Appl Clin Inform* 2016; 7: 803–816
<http://dx.doi.org/10.4338/ACI-2016-03-RA-0037>

Background and significance

A language takes advantage of a system of signs appropriate to a specific notation [1]. In the cultural universe of a linguistic community, signs are equivalent to the set of words available in the language and form its lexicon. The lexicon of a language consists of a system of units, known as lexical units, that convey the general knowledge of this community [2]. Therefore, in a general context, the common or general language is the oral and written communication system used by a linguistic community on a daily basis [3].

Usually a language specialization occurs to perform a specific activity. Communication needs to be differentiated, i.e., its content needs to be accurately and unambiguously described, with reference to a particular scientific or other established normative structure [4]. The concept is the mental organization of knowledge or of a unity of either thought or knowledge. It can be expressed – and consequently shared – in the specialized community through a specialized lexical unit called „term“ [5]. However, a term is not the concept itself, but rather the access key, the reference, that allows the human mind to retrieve the mental organization of particular knowledge [6]. A specialized language is widely used in the healthcare system because it has its own terms, describing the medical observations associated with clinical data and additional examinations of a particular patient [7].

Medical observations are written in reports that often have a characteristic overall schematic structure known as *textual superstructure*. Textual superstructure consists of well-defined areas of text designated as textual categories, i.e., constituent parts of the text [8].

In the case of an imaging examination, the radiologist or imaging technician is responsible for coordinating the execution of the diagnostic imaging examination and recording its analysis and interpretation in a so called “radiology report” [9]. In general, the radiology report contains the following constituent parts: the technique used in the examination, patient data, body region examined, description of the image findings and, when a confident diagnosis is made, a conclusion or diagnostic impression [10, 11]. This report is sent to the requesting physician and will be used in the investigation of the patient’s medical condition.

The radiology report is the means of communication by which the requesting physician becomes aware of the interpretations of the performing physician; the communication should be accurate and objective. The radiological community pursues terminology harmonization – a consensus to use specific terms in a specific situation [3] – and created several terminological products over time, including the American College of Radiology’s Index for Radiological Diagnosis (ACR score) [12], the Radiology Lexicon (RadLex) [13], the Radiological Society of North America (RSNA), and the Breast Imaging Reporting and Data System (BIRADS) [14]. In addition to the terminology, the latter has proposed the standardization of procedures based on the examination results and is globally accepted [15]. However, adherence to accepted guidelines is highly variable between radiologists within and between institutions [16].

The radiology reports are generally structured empirically by asking each of the medical teams, both the requesting and performing teams, what constituent parts of the text are necessary to compose the radiology report [17, 18]. Therefore, information or terms that are often used may no longer be considered, leading to the rejection of the use of the structured report. Structured data entry ensures accuracy in information retrieval, but is less comprehensive than the text-free structure and may prevent the performing physician from reporting important findings not contemplated in a document with a predefined structure. Therefore, a possible solution would be to interpret the text written by the specialist and to transfer the information to an ontological computational structure.

Objectives

The objective of this study was to propose a methodology for structuring information from radiology reports that takes into consideration regional specificities, including the terminology to be used.

Methods

The sample consisted of 1,701 consecutive magnetic resonance imaging (MRI) reports of the lumbosacral spine obtained from three Institutions: a) university hospital, named (A), where we collected 600 examinations; b) a private clinic, named (B), with 600 examinations; and c) a private clinic from other state, named (C) with 501 examinations. For inclusion in the study, the reports had to be complete (i.e., containing information on the examination technique, description, and conclusion) and had to refer only to the examination of the lumbosacral spine. Repeated examinations of the same patient were excluded, regardless of the time interval between them.

This study was submitted to and approved by the Research Ethics Committee of our institution on July 16, 2012 under Certificate No. 00593812.2.1001.5440. The flowchart of study is described in ► Figure 1.

To accommodate the textual corpus – the set of MRI reports of the lumbosacral spine in this case—and to perform text mining and ontological organization, a computational tool called *Meta-generalis* was created, which was built using the C# language and designed using data access layers, business rules, and unit tests (► Figure 2).

Using text-mining techniques, the textual corpus was decomposed into paragraphs, sentences, and lexical units, called ‘tokens’ in the specialized language [19]. Lexical units were subjected to seven normalization procedures:

1. conversion to lowercase letters;
2. removal of accent marks;
3. removal of special characters and punctuation;
4. removal of stop words;
5. conversion from plural to singular;
6. conversion of nominal forms to verbs; and
7. conversion of suffixes.

To assess the impact of each normalization procedure, two indicators were used: lexical density (LD), defined as the ratio between the number of normalized lexical units (N_{Lex}) and the total number of lexical units in the textual corpus (N) [20] and the lexical agglutination (LA), given by the complementary of lexical density.

Equation 1: Lexical density (LD) Index

$$LD = \frac{N_{Lex}}{N}$$

Equation 2: Lexical agglutination (LA) index

$$LA = (1 - LD)$$

The inverted file, which consists of the data structure that stores the lexical units into their respective documents, was stored in the relational database Microsoft SQL Server 2008 R2 Express (free). Compound candidate terms, which are those composed of more than one lexical unit (nGram>1), were identified by horizontal contextual ontological decision (► Figure 3), which involves the identification of the candidate term, taking into account the previous or next concept. The quantified results of this study refer to simple terms (nGram = 1) while the manual conceptualizations consider the satellites lexical units of a term to determine the concept to be associated.

Superstructures and analysis of candidate terms by Meta-generalis

The process of analysis and conceptualization of the candidate terms of a specific textual corpus in an ontological structure is known as conceptualization. The concept of a candidate term was established automatically by searching for it in an electronic terminological product (ETP) or manually when the term was not found in the ETP [21]. In the manual conceptualization, the concept was determined via consultation with specialists in the field, dictionaries, and the literature. The relation-

ship between the candidate term and the concept was established using the relationships “is”, “associated”, “indicates”, and “part of”.

The list of candidate terms (for example “disk bulging”, “hernia”) associated with the textual corpus of the three Institutions was fully conceptualized (► Supplementary Material – Appendix C).

To identify the textual superstructure, the candidate terms with frequencies between 95% and 100% of the number of documents included in the study were listed because the terms with frequencies equal or similar to the number of documents of the textual corpus can potentially represent the constituent parts of the textual superstructure. The candidate terms listed under this condition were analyzed within the context that originated them, and when it involved the identification of a constituent part of the text, they were considered objects of the concept “textual category” in the ontology.

The candidate terms were retrieved from the ETP of Health Sciences Descriptors [22] through requests via Hypertext Transfer Protocol (HTTP) and interpretation of the definition key (“<definition>”) in the markup language file (XML) retrieved. After searching for definitions of the candidate terms in the DECS (Descriptors of Health Sciences, in Portuguese language), the concepts retrieved were evaluated and identified as either compatible or incompatible. The candidate terms that had no reference in the ETP were manually conceptualized. The sources consulted included the Michaelis Online Dictionary [23] because, in this reference, the concepts are associated with the term in several specialized languages, including health sciences; the Dorland Illustrated Medical Dictionary [24]; the Unified Medical Language System (UMLS) [25]; scientific reports; international weights and measures agencies; and the human anatomy atlas [26].

The conceptualization of the candidate terms of a textual corpus in the computational tool *Meta-generalis* transformed the textual corpus into an ETP allowing retrieval of the concept of a particular candidate term or query in formalized ontology.

In the methodology proposed, the questionnaire used was an exploratory tool aimed at evaluating the superstructure and terminology extracted using this methodology. The evaluation questionnaire was sent to 10 specialist radiologists from the participating Institutions without identifying them; therefore, all responses were anonymous.

Suggested models and questionnaire validation

To assess the superstructures identified in the radiology reports studied, three textual superstructures were presented and designated as Models A, B, and C. The superstructure of Model A was composed of two constituent parts, designated ‘Description’ and ‘Conclusion’, and the ‘Description’ part could potentially include a constituent part designated ‘Observations’ to be filled as needed. In Model B, the superstructure consisted of two explicit constituent parts, designated ‘Report description’ and ‘Diagnostic impression’, and an implicit constituent part designated ‘Examination technique’, which was informed in the first paragraph of the ‘Report description’. Model C consisted of three explicit constituent parts designated ‘Examination technique’, ‘Description’, and ‘Result’.

Ten specialist radiologists with at least three years of experience in this type of examination evaluated the textual superstructure and the extracted terminology using an electronic questionnaire. These professionals were allocated as follows: six from Institution A, two from Institution B, and two from Institution C.

The list of terms to be evaluated by the radiologists was created using the following criteria:

1. A list of terms with lexical differences between the Institutions was generated, designated ‘List 1’ and contained 372 (three hundred seventy-two) terms.
2. Based on List 1, a specialist radiologist excluded the terms referring to human anatomy (e.g., artery, liver, etc.), those common to natural language (e.g., correlation, complex, etc.), descriptors of minor diseases or pathologies (e.g., cyst, lipoma, etc.), and procedures (e.g., ultrasound, computed tomography, etc.). A total of 95 terms were excluded; therefore, List 2 was composed of 277 terms.
3. Using the “RANDBETWEEN” function in Microsoft Excel, a random number was assigned to each of the 277 terms in List 2, and the terms were then sorted from highest to lowest, based on those numbers.
4. After this process, the first 30 terms in the evaluation questionnaire were selected.

Results

Quantitative analysis of the results showed that the number of raw lexical units extracted from Institutions A, B, and C were 2,633, 1,680, and 1,816, respectively, totaling 6,129 units. After the seventh and last normalization procedure, the numbers of lexical units normalized in Institutions A, B, and C were 1,949, 1,193, and 1,426, respectively, totaling 4,568 units (► Figure 4), of which 4,183 were conceptualized.

After the application of the seventh and last normalization procedure, the lexical densities (LDs) of Institutions A, B, and C were 74.0%, 71.0%, and 78.5%, respectively, and the lexical agglutinations (LAs) accumulated were 26.0%, 29.0%, and 21.5%, respectively (► Figure 5).

In Institution A, which follow HL7 standards (Health Level Seven International) [27], the candidate terms were automatically submitted to DECS, 22.1% of the terms were automatically conceptualized, and 77.9% were manually conceptualized. For Institution B, 77.8% and 22.2% and for Institution C, 83.1% and 16.9% of the terms were conceptualized automatically and manually, respectively (► Figure 6). The manual conceptualization decreased because the candidate terms were automatically submitted to ETP generated from the study of the previous Institution. These results indicate a trend of decreasing need for manual conceptualization as the amount of ETP used increased, particularly the terms extracted from the specialized language studied.

Three textual superstructures were identified in the radiology report, one for each Institution, containing five constituent parts of the textual super structure, with terminological variation in their designations, as follows: 'Examination technique' (explicit), 'Description', 'Examination technique in the first paragraph of the Description' (implicit), 'Conclusion', and 'Observations' (► Figure 7).

The textual superstructure of the radiology report of Institution A comprised the constituent parts 'Description', 'Conclusion', and 'Examination technique' in the first paragraph of the 'Description'. However, the latter was used in 19% of the radiology reports of the sample.

In Institution B, the constituent parts 'Description', 'Diagnostic impression', and 'Examination technique' were identified in the first paragraph of 'Description'.

In Institution C, the constituent parts 'Result', 'Conclusion', 'Examination technique', and 'Observations' were identified, and all of these constituent parts were highlighted in the text. The constituent part 'Result' was used to describe the imaging examination results, as was the constituent part 'Description' identified in Institutions A and B. For the constituent part 'Conclusion', the term 'Diagnostic impression' was used in 5% of the sample reports of this Institution, i.e., terminological variation, a case of synonymy, occurred. The constituent part of the text designated "Observations" was identified by the abbreviation "OBS:" and occurred in 8.5% of the sample reports.

During the conceptualization process, out of the 4,568 candidate terms initially listed, 4,183 terms were conceptualized and yielded 727 unique concepts in Portuguese. Institutions A, B, and C shared 606, 387, and 431 concepts, respectively (► Supplementary Material – Appendix C).

The candidate terms were related to the concepts through 13,963 relationships, of which 6,018 relationships were used in Institution A, 2,875 were used in Institution B, and 5,070 were used in Institution C. Owing to the sharing of concepts among the Institutions, the distribution was not uniform.

To each evaluator was given a sequential identification code so that the first evaluator was named A1 and the last was named A10. The full questionnaire can be found in ► Supplementary Material – Appendix A.

For qualitative analysis of the results, three models based on the superstructures observed in the sample were presented. However, they were not equivalent in the three Institutions, showing differences in terms of the position of the constituent parts of the text and of the term that described each of the constituent parts associated with the superstructure.

No evaluator chose Model A, as noted in the justification presented in question #2, because this model did not contain any information on the examination technique. Additionally, the term 'Diagnostic impression' is more appropriate than the terms 'Conclusion' or 'Result'.

Most evaluators (60%) preferred Model B because it contained the constituent parts necessary to prepare the radiology report [10] and the term 'Diagnostic impression' instead of 'Conclusion' or 'Result'.

The evaluators (40% of the total) who chose Model C considered it more important to explain the constituent part 'Examination technique' than the terms used to indicate the concluding constituent part of the radiology report, i.e., 'Diagnostic impression', 'Result', or 'Conclusion'.

More than half of the evaluators (60%) preferred the term 'Diagnostic impression' instead of the term 'Result' when considering those who chose Model B.

When inquired about whether the terms presented were suitable for describing the findings, none of the evaluators disagreed. However, most evaluators disagreed about the definitions of some terms, as detailed below. Also, when inquired about the number of terms used in their institution to prepare the reports, all respondents reported the use of 16 to 30 terms. If inquired about whether some of the terms and concepts presented did not correspond to the meaning used in their respective Institution, out of the 33 terms presented, three were considered inadequate: 'field of view', 'hydromyelia', and 'spondylosis'.

Moreover, most of the evaluators fully agreed that a common terminological product between the Institutions would facilitate the process of writing a radiology reports. None of the evaluators disagreed that the incorporation of a terminological product common to various Institutions would be a facilitator of communication (► Supplementary Material – Appendix B).

Discussion

The process of building the *Meta-generalis* computational tool was motivated primarily by the lack of a tool capable of mining texts in Portuguese, which allows, in an integrated manner, the contextual conceptualization of candidate terms using an ontological structure that could be stored in a relational database (RDB). Most of the software available either mine texts or create ontologies, but rarely use RDBs, which allow for inferences to be made at any time. Furthermore, we did not find any software that considered or identified the textual superstructures present in the textual corpus.

Modeling a data structure capable of accommodating terminology records [3] by language, formalizing this structure ontologically, creating algorithms in C# to contemplate the inverted files, converting texts, anonymizing, and storing texts in RDB were the main technical challenges of this study.

Although we have used existing techniques as mining and relational database that already, information retrieval is only one aspect of this study. The purpose was to describe a method to create reusable electronic terminological products through conceptualization, listing and associating terms to their concepts into an ontological structure able to assign semantics and using external terminology products and, also, the own terminology products generated by studies made by *Meta-Generalis* tool.

The important concept behind the *Meta-generalis* tool and methodology is that the knowledge is not in the lexical units but in what they represent in the real world. Terminological variations exist because each term refers to a concept that may not exist or may be incomplete, depending on the personal experience and language of those who read it. Therefore, knowledge extraction does not occur by the extraction of lexical units, but by their conceptualization using an ontological structure. Thus, this computable structure allows all the members of the specialized community to which the term belongs to know the same description of the object, i.e., allows them to have the same knowledge about the concept.

In a similar study, Hripcsak et al [28] classified nearly 900,000 reports of chest x-ray after defining some attributes for classification based on 150 reports that have some kind of coding and specific references in the text as „possible pneumonia“. But by using a guided approach in the processing of natural language they experienced difficulties with synonyms, language expressions, denial and intensity. In addition, this approach is language specific, preventing its generalization to other languages. Once the knowledge is formalized, you can conceptualize what is intensity and denial, for examples, and which candidate terms are synonymous, including those with typo errors, regardless of language. Hassanpour & Langlotz [29], published a study with similar purpose: extract semantics of radiological reports in free text. However, unlike these authors, we did not opt for simplification of the concepts in classes, which actually resulted in a huge effort to formalize the concept of each candidate terms, including classes included in its simplifications. However, the machine learning ef-

fort were concentrated in the first institution, where 77.9% of the terms were conceptualized and related compared to only 16.9% for the third institution; demonstrating that learning initial effort tends to decrease as more knowledge is formalized, because they have the same kind of technical language. It is important to note that, in our study, 1,701 reports were analyzed, while Hassanpour & Langlotz used only 150. To evaluate the results, they randomized 15 radiological reports 5 of each organization and asked for 2 radiologists verify if the classes they created (information model) were present in these reports. In the Meta-generalis, we asked for 10 radiologists to assess 30 randomized concepts and 3 textual superstructures in order to check whether they found representative of original free text reports.

The seven normalization procedures used herein were able to agglutinate the textual corpus composed by the radiology reports of the three institutions studied at an average rate of 25.5%, conserving intelligible lexical units. The first normalization procedure, the conversion to lowercase letters, obviously did not agglutinate the text, but allowed the computer to perform exact comparisons, increasing the processing speed because there was no concern about differences between uppercase and lowercase letters. The normalization procedure with the highest rate of agglutination among the institutions was the conversion from plural to singular, with an average of 12%.

The textual superstructure of radiology reports was similar in the three institutions evaluated. However, there was terminological variation (synonymy) between the names of the constituent parts within the same Institution and between different institutions. The constituent part of the text attributed to the description of the clinical findings was denominated 'Description' in Institutions A and B and 'Result' in Institution C.

Textual superstructures have been suggested automatically by the META-Generalis based on ontology created by the association of the term with the concept 'Textual Category' including the ETP generated in the previous study. However, precision and recall tests were not obtained in this study using the created ontology. The conceptualization process is the most labor-intensive for constructing the ETP because, despite the automatic conceptualization, candidate terms needed to be analyzed within their context to determine the compatibility of the associated concept and the normalized lexical unit. Concepts representing classes of objects, such as 'lumbar vertebra', and the objects associated with these classes, such as 'L1' is a 'lumbar vertebra', were created.

The use of ETPs that are specific of the specialized language allows greater conceptualization because the difference between automatic retrieval of concepts from DECS and the ETPs of Institutions A and B was 61% higher. There were no major terminological differences between the Institutions, and the number of unique concepts between them accounted for an average of 6% of the concepts.

Three different textual superstructures were identified, originating 4,183 candidate terms, all of which were conceptualized, yielding 727 concepts in Portuguese.

The analysis of the responses of the evaluators when considering the first two questions makes it clear that the evaluation was based on the personal experience of each evaluator and not in a formalized knowledge. However, the evaluators were unanimous in the topic about the minimum information required for the radiology report, confirming the results of previous studies [11, 30], i.e. the examination technique, description of findings, and diagnostic impression. The preference for "Diagnostic impression" rather than "Results" probably is related to the inherent uncertainty of diagnostic in radiology and medicine as well.

Three terms were considered inadequate by the evaluators: 'spondylosis', 'visual field' [23], and 'syringohydromyelia' [31], of which 'spondylosis' was considered inadequate by two evaluators.

The methodology developed consisted of six steps:

1. collection of radiology reports of a specific MRI examination;
2. textual decomposition;
3. normalization of lexical units;
4. identification of textual superstructures;
5. conceptualization of the candidate terms; and
6. evaluation of the identified superstructures and of the terminology extracted by specialists in the specialized language studied.

This methodology allows identifying textual superstructures, a result similar to that proposed by BIRADS [14], an ontological conceptualization of terminology considering the language and the context, and was also similar to SNOMED-CT [32].

This study aimed at proposing a methodology for structuring the data contained in radiology reports, because these types of reports are usually written in specialized language and free text. For this reason, although the computational structure was constructed considering the possibility of information retrieval, i.e., retrieval of reports using a query vector, these types of studies have not been conducted. Therefore, future studies in this field will be of great importance to assess how much ontology improves the retrieval of information in the textual corpus studied. Export and import of ontology to computational tools such as Protégé [33] would also be an interesting topic for future studies to promote interoperability between different systems.

The limitations of this study included the small number of institutions evaluated (despite the significant number of examinations and reports), which can compromise the variation of terms and concepts that can exist within the same language due to regional variations. The references used for conceptualization of the medical terms are all commonly used in medicine, but their selection was random and may represent a selection bias of the authors. Furthermore, reports of a single anatomical region and a single type of examination were evaluated. Also, we acknowledge that evaluation of the process by radiologists from the same institution is a limitation. However, there were radiologists from three different institutions, which minimize this effect. This approach was necessary because the inclusion of various types of reports could jeopardize the manipulation and proper assessment of data. However, the proposed method can allow the management of clinical knowledge using free text that poses no barrier to physicians doctor similar to a previous described experience [34].

Conclusions

In conclusion, a methodology was developed and proposed for structuring information in the native language, allowing extraction of the lexical units while maintaining the references to their respective constituent parts and conceptualizing these units ontologically. Furthermore, the superstructures identified were considered adequate by the evaluators.

Clinical relevance

It was possible to develop a methodology for structuring reports using ontologic concepts, creating a software to accelerate the process of structuring a report. The system was able to standardize a MRI report using and mixing terms from different communities.

Conflict of Interest

Authors declare no conflict of interest.

Human Subjects Protection

Authors declare this research was conducted in compliance with ethical standards of our Institutional Review Bureau, as stated in the manuscript.

Acknowledgements

We would like to thank Prof. Maria Cristiane Galvao for helping us with her ontology and terminology expertise. We also thank Dr. Jose Daniel Oliveira for his diligence to provide MRI reports.

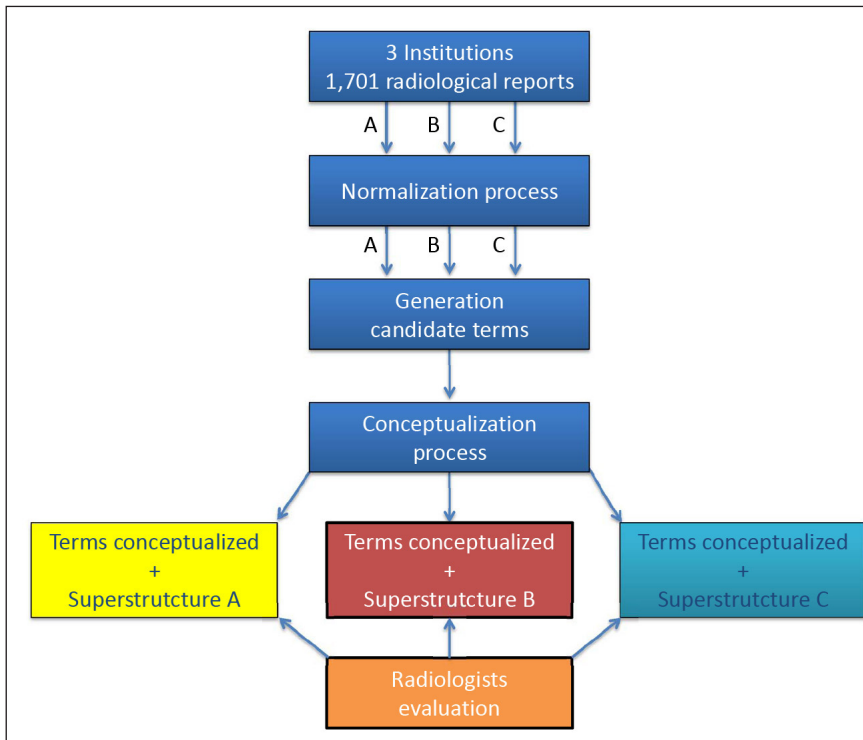


Fig. 1 Flowchart of study.

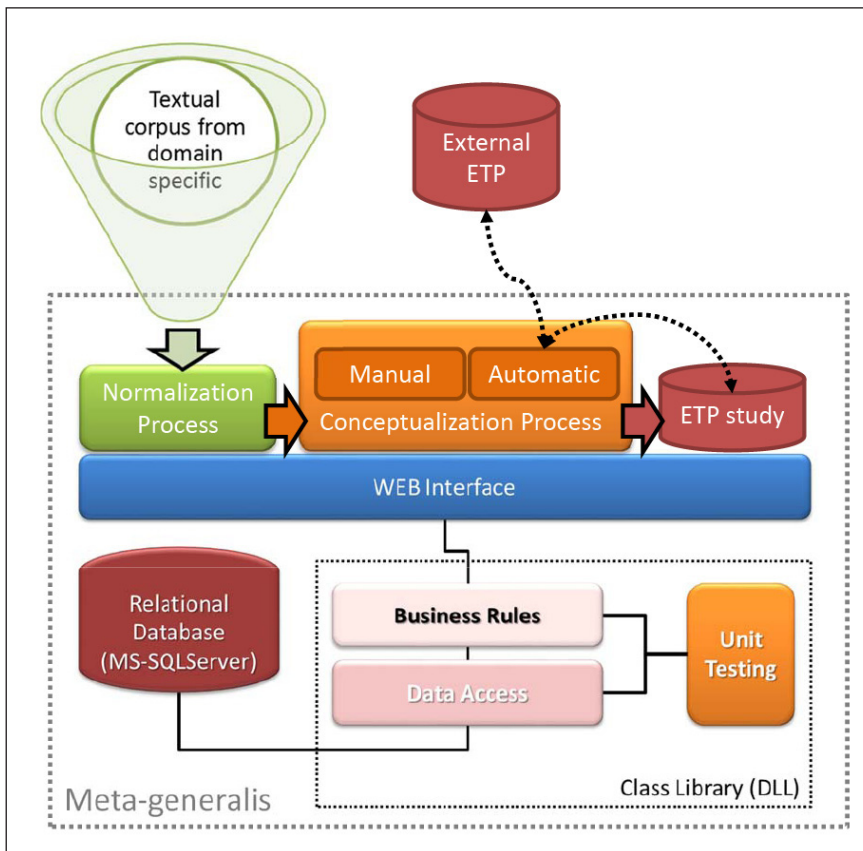


Fig. 2 Schematic drawing of Meta-generalis software.

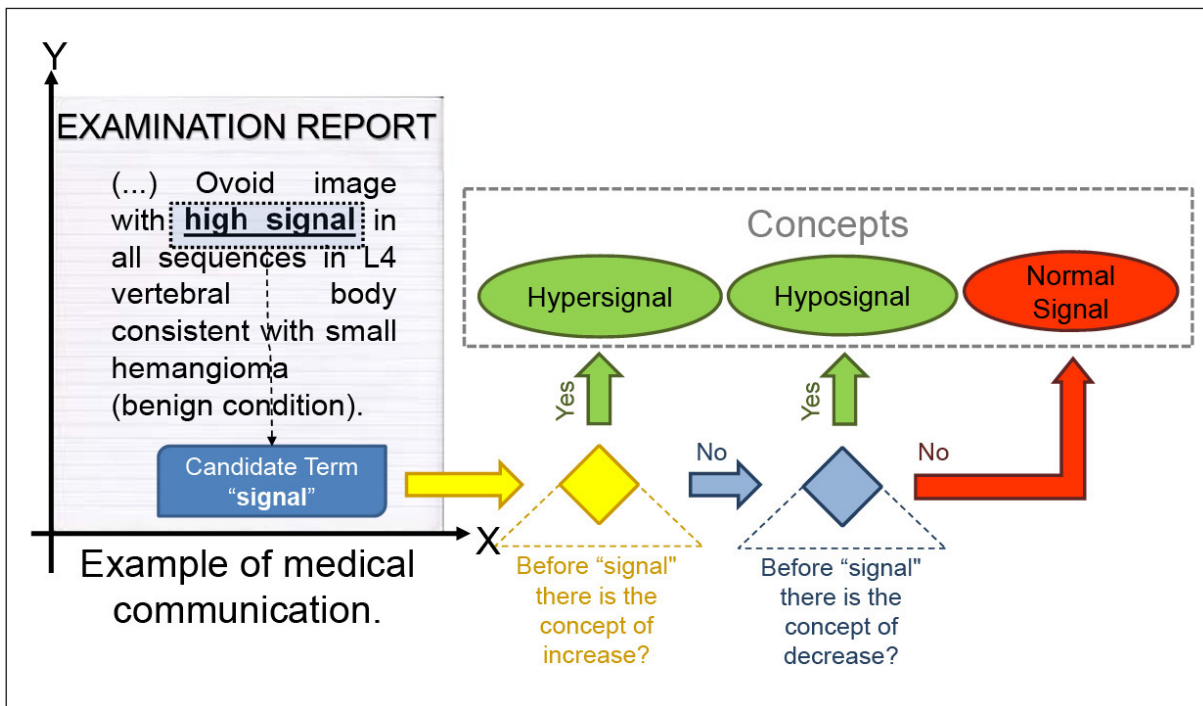


Fig. 3 Example of horizontal ontological context decision usage. A concept is extracted from lumbar spine MRI report.

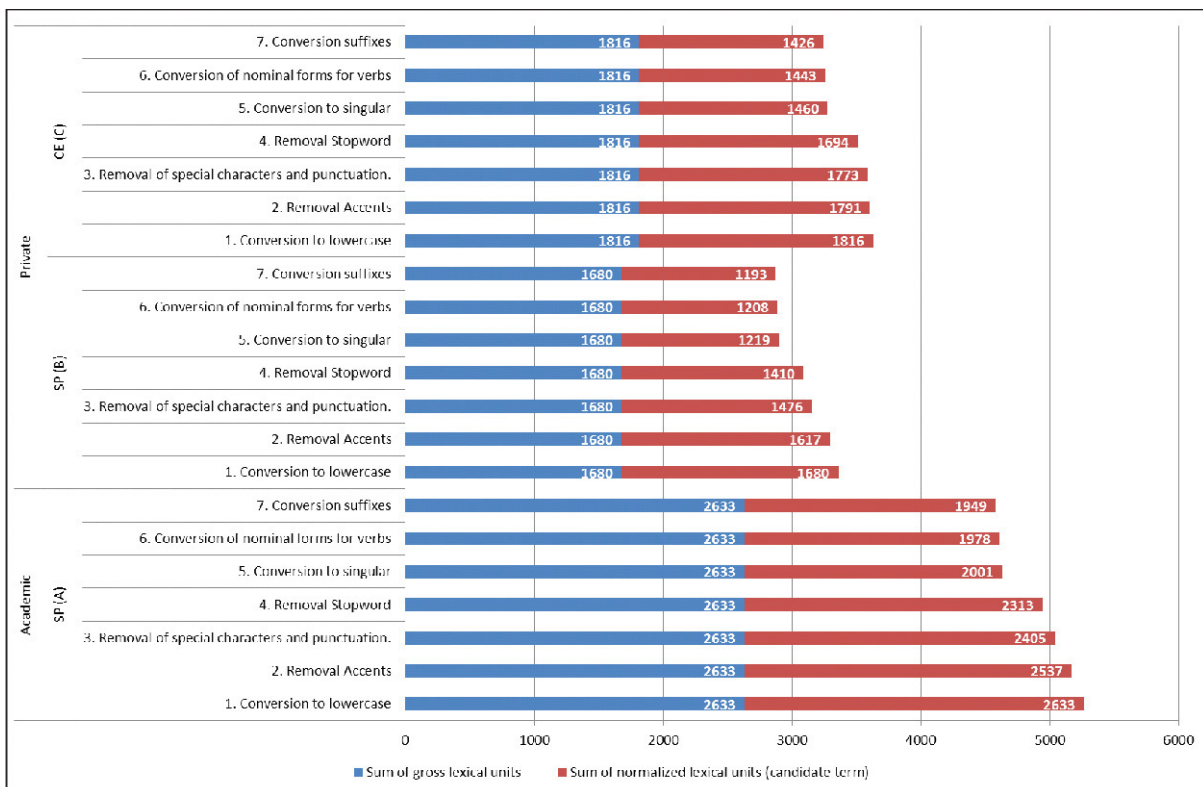


Fig. 4 Normalizations results by institution and type of normalization.

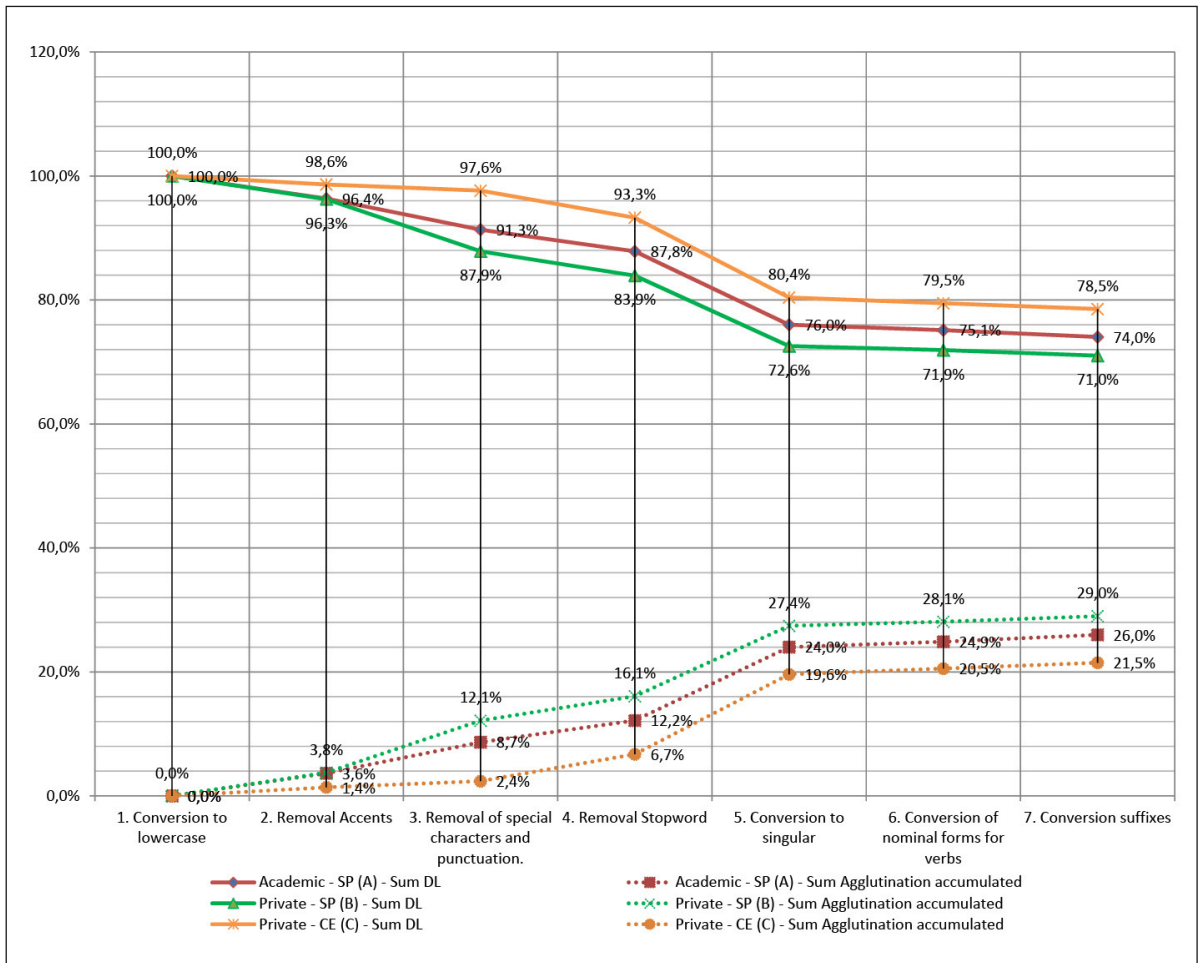


Fig. 5 Percentage of lexicon density and lexicon agglutination accumulated for institution and type of normalization

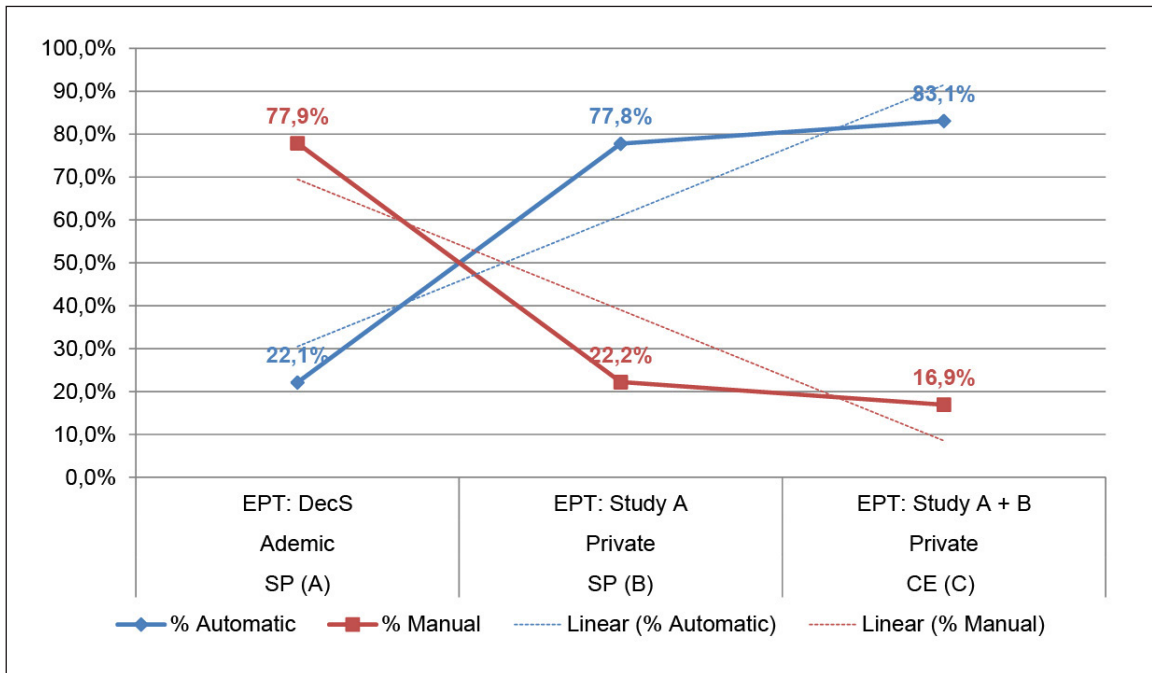


Fig. 6 Trend decreasing for the manual conceptualization when using the EPT generated from the study of the previous Institution

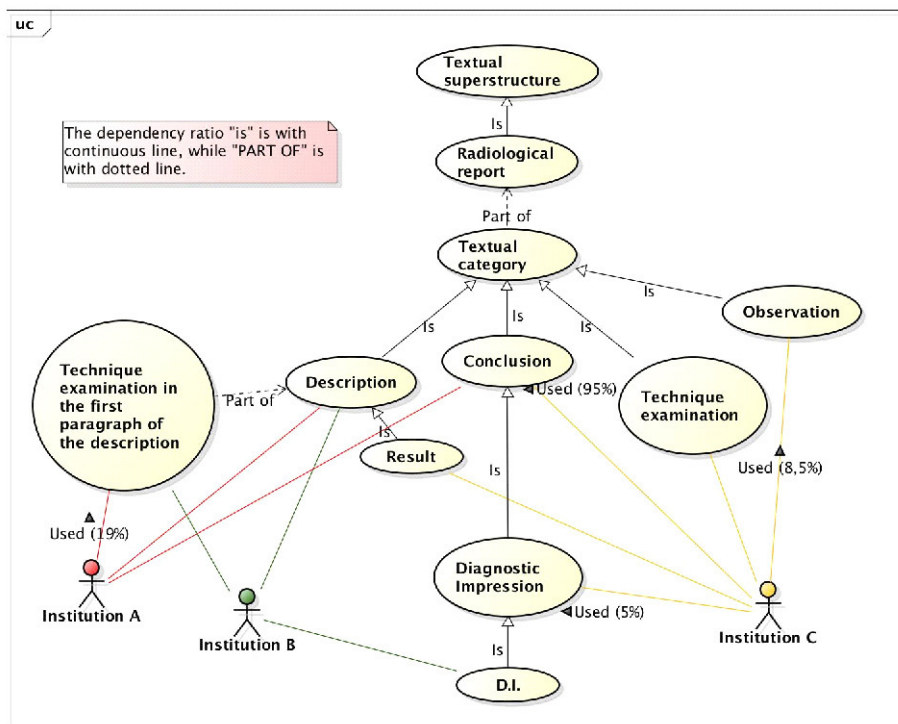


Fig. 7 Textual superstructures identified in all 3 institutions

References

1. Brentari D. A prosodic model of sign language phonology. First edition, MIT Press, Cambridge, 1998.
2. Reis AKO. The interfaces of lexicon disciplines. *Dialogia* 2010; 9: 173–179.
3. Pavel S, Nolet D. Terminology Manual. Public Works and Government Services. Translation Bureau. Government of Canada, 2002. <https://linguisticadocumentaria.files.wordpress.com/2011/03/pavel-terminologia.pdf>. Accessed 11 Dec 2012.
4. Temmerman R. Towards new ways of Terminology description: the Sociocognitive approach. First edition, John Benjamins Publishing Company Amsterdam, 2000, p. 258.
5. Silva-Filho SC. Concept/Term: The variation of a relationship. *Trama* 2010;6:75–86. <http://e-revista.unioeste.br/index.php/trama/article/view/5121/4125>. Accessed 12 Sep 2012.
6. Francelin MM, Kobashi NY. Conceptions about the concept in information and knowledge organizations. *Cien Inf* 2011;40:207–228. <http://core.ac.uk/download/pdf/11889801.pdf>. Accessed 12 Sep 2012.
7. Spackman KA, Campbell KE, Côté RA. SNOMED RT: a reference terminology for health care. *Proc AMIA Annu Fall Symp* 1997: 640–644.
8. Vieira JML, Correa RE. Information visualization to user-friendly interface construction for information retrieval systems. *R Eletr Bibliotecon Ci Inf* 2012; 16: 73–93.
9. Maciel-Lima SM, Rasia JM. Images do not speak for themselves: a sociological study of a liver transplantation team. *Saude Soc* 2012; 1: 42–51.
10. Fenelon S. Ethical and legal aspects in imaging. *Radiol Bras* 2003; 36: 3–6.
11. Prestes-Jr LCL, Tourinho EK, Rangel M. Medico-legal analysis of imaging demands. *Radiol Bras* 2010; 45: 98–100.
12. American College of Radiology – ACR score. Available at <http://www.acr.org/~media/ACR/Documents/PDF/QualitySafety/Radpeer/ScoringWhitePaper.pdf>. Accessed on September 28th, 2015.
13. Langlotz CP. RadLex: a new method for indexing online educational materials. *RadioGraphics* 2006; 26(6): 1595–1597.
14. ACR Breast Imaging and Radiological Data System – BIRADS. Fourth edition, American College of Radiology, Reston, 2003.
15. Passman LJ, Farias AM, Tomazelli JG, de Abreu DM, de Assis M, de Almeida PF, da Silva RC, Santini LA. SISMAMA – Implementation of an information system for breast cancer early detection programs in Brazil. *Breast* 2011; 20 (Suppl. 2): S35–S39.
16. Collins SA, Bakken S, Vawdrey DK, Coiera E, Currie L. Clinician preferences for verbal communication compared to EHR documentation in the ICU. *Applied Clinical Informatics* 2011; 2: 190–201.
17. Andreisek G, Deyo RA, Jarvik JG, Porchet F, Winklhofer SF, Steurer J. Consensus conference on core radiological parameters to describe lumbar stenosis – an initiative for structured reporting. *Eur Radiol* 2014; 24: 3224–3232. doi: 10.1007/s00330–014–3346-z
18. Rubin DL. Creating and curating a terminology for radiology: ontology modeling and analysis. *J Digit Imaging* 2008; 21(4): 355–362.
19. Lavelli A, Califf ME, Ciravegna F, Freitag D, Giuliano C, Kushmerick N, Romano L, Ireson N. Evaluation of machine learning-based information extraction algorithms: criticisms and recommendations. *Language Resources and Evaluation* 2008; 42(2): 361–393.
20. Roberto JA, Martí MA, Llorente MS. Analysis of lexical richness in the context of latent demographic user attributes classification. *Procesamiento de Lenguaje Natural Journal* 2012; 48: 97–104.
21. Friedman C, Alderson PO, Austin JHM, Cimino JJ, Johnson SB (1994) A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc* 1994; 1: 161–174.
22. DECS – Descritores em ciências da saúde – Health science descriptors [online]. Bireme. <http://DECS.bvs.br>. Accessed 14 Jun 2012.
23. Michaelis Online Dictionary. <http://michaelis.uol.com.br/escolar/ingles/index.php>. Accessed 06 Jun 2012
24. Dorland's. Illustrated Medical Dictionary. Thirty-second edition, Saunders, Philadelphia, 2011.
25. Unified Medical Language System (UMLS). <https://uts.nlm.nih.gov/home.html>. Accessed 19 Jun 2013
26. Netter FH. Human Anatomy Atlas. Sixth edition, Saunders, Philadelphia, 2014, pp.142–144.
27. Health Level Seven International – HL7. Available at www.hl7.org. Accessed on september 12th, 2015
28. Hripsack G, Austin JH, Anderson PO, Friedman C. Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports. *Radiology* 2002; 224: 157–163.
29. Hassanpour S, Langlots CP. Information extraction from multi-institutional radiology reports. *Artificial Intelligence in Medicine* 2016; 66: 29–39.
30. Fenoy AJ, Menezes AH, Fenoy KA. Craniocervical junction fusions in patients with hindbrain herniation and syringohydromyelia. *J Neurosurg Spine* 2008; 9: 1–9.

31. Gerstmair A, Daumke P, Simon K, Langer M, Kotter E. Intelligent image retrieval based on radiology reports. *Eur Radiol* 2012; 22(12): 2750-2758.
32. Snomed Healthcare terminology. <http://www.ihtsdo.org/snomed-ct>. Accessed 14 Jun 2012.
33. Protégé software. Available at protege.stanford.edu. Accessed 12 Sep 2012.
34. Plazzotta F, Luna D, Quirós FGB. Health Information Systems: Integrating clinical data in different scenarios and users. *Rev peru med exp salud publica* 2015, 32: 343-351.