Designing an information retrieval system for the STT/SC

Andrei de Souza Inácio
Brazilian Institute for Digital Convergence - INCoD
Federal University of Santa Catarina - UFSC
Florianópolis, SC - Brazil
andrei@inf.ufsc.br

Aldo von Wangenheim
Brazilian Institute for Digital Convergence - INCoD
Federal University of Santa Catarina - UFSC
Florianópolis, SC - Brazil
awangenh@inf.ufsc.br

Rafael Andrade
Informatics Department
Federal Institute Catarinense – Câmpus Ibirama
Ibirama, SC - Brazil
rafael.andrade@ibirama.ifc.edu.br

Douglas D. J. de Macedo
Department of Computer Science
Federal University of Sergipe - UFS
Aracaju, SE - Brazil
dmacedo@ufs.br

Abstract— **The Santa Catarina State Telemedicine and Telehealth System - STT/SC stores well over 2 million examinations and every month about 20 thousand new imaging exams are sent to the system database. As a significant part of the findings associated to these medical data are stored as text documents, the precise extraction of information is a difficult task. In the healthcare domain it is common to find different terms, some of them appearing as composed expressions, used to represent the same concept, whereas the existence of simple modifiers can turn an expression into another, different concept. This work presents an information retrieval architecture developed for the STT/SC. It consists in a module integrated into the STT/SC and was developed in order to support both experienced and inexperienced users to perform queries. Experiments were performed to evaluate the accuracy of the proposed system in real situations. Results show that the search can be performed more much faster and with acceptable precision.**

Keywords— Retrieval Information, Telemedicine, DeCS, Controlled Vocabularies.

I. INTRODUCTION

After more than two decades of crescent worldwide digitalization of healthcare, medical records in hospitals and public healthcare organizations store vast unexplored quantities of data. Processing and exploring these data in order to gather better epidemiological information offers great opportunities for researchers, clinicians and public healthcare policymakers.

The existence of so much data and the necessity to extract information for epidemiological research and the definition of more informed public healthcare policies motivated researchers from the areas of Artificial Intelligence and Natural Language Processing to develop tailored information retrieval and knowledge extraction techniques [1][2][3]. As a significant part of these medical data are stored as unstructured text documents, without much of standardization, the precise extraction of information is, however, a difficult task. In the healthcare domain it is common to find different terms, some of them appearing as composed expressions, used to represent the same concept, whereas the existence of simple modifiers can turn an expression into another, different concept.

In order to represent these concepts in a formal manner, domain ontologies are often used [4]. In the healthcare domain various standardized vocabularies that can be seen as ontologies already exist, such as MeSH [5], SNOMED-CT [6] and DeCS [7]. Ontologies help the exchange of information between systems and also the development of medical applications that make use of structured information. On the other side, there exist also international standards that specify structure and semantics of patient records, findings reports and other documents, such as the DICOM Structured Reporting (DICOM SR [8]) and HL7/CDA[9] standards.

An information retrieval architecture based upon structure and semantics standards, and also controlled vocabularies, consists in a system capable of providing more detailed information and more reliable and precise results, such as detailed correlations between morbidities, patient history and radiological findings. These data are vital for public healthcare policymakers, since they can provide evidence to efficiently direct healthcare resources and professionals based upon concrete morbidity landscapes and relationships.

This work presents an information retrieval architecture developed for the Santa Catarina State Telemedicine and Telehealth System – STT/SC. This architecture employs controlled vocabularies for data catalogization and a specially developed search engine for data indexing and storing. Even though the services offered by the STT/SC cover almost all medical areas where Telemedicine makes sense, this study focuses on Telecardiology and Teledermatology examinations, where the exclusive use of DICOM SR and controlled vocabularies is presently fully integrated in the clinical routine.

This paper is organized as follows: Section II presents the background to the work; Section III presents related work that deals with solutions or techniques for information retrieval. In Section IV we present the architecture devised for our solution and Section V presents the experiments performed for the validation of our prototype and the results obtained. Conclusions and future work are presented in Section VI.

978-1-4799-6644-8/14/$31.00 ©2014 IEEE

II. Background

A. STT/RCTM

In 2005 the Telemedicine Lab (LABTELEMED) of the Federal University of Santa Catarina – UFSC, together with the Santa Catarina State Health Department – SES/SC created the STT/SC. The main objective was to offer asynchronous telemedicine services, mainly for medical imaging examinations performed in the context of the Brazilian Public Health System – (SUS in Portuguese language). In 2010 a new customized web-based infrastructure was launched, offering a large palette of new services, including web- and smartphone-based access to examination results to patients and continuous education for medical, nursing and technical staff. In 2012 the STT/SC was available in all 295 municipalities of the State and in 2013 the mark of 2 million examinations was achieved.

B. Information Retrieval

The Information Retrieval – IR area, often considered synonymous with document retrieval or text retrieval, aims the development of search techniques that allow the efficient and effective finding of documents whose information content is relevant for the information needs of a given user [10]. One of the most important steps during the search process in IR systems is the prediction of which documents should be considered relevant for some search task and which ones should be discarded. This selection task is performed by an algorithm that, based upon previously defined heuristics, decides which documents should be considered relevant for retrieval and sorts them accordingly to the relevance criteria established by these heuristics [11].

There are three classic IR models: Boolean, vector space and probabilistic. The Boolean model is one of the most used in commercial systems and is based upon logic operators. The vector space model is mostly employed in text retrieval methods and represents terms, documents and queries through vectors, allowing the ranking of the searched documents accordingly to relevance criteria. The probabilistic model employs statistic techniques in order to estimate if a document is relevant to some search accordingly to the presence of search terms and other, statistically related, terms [11].

C. Controlled Vocabularies

Controlled vocabularies are lists of terms organized accordingly to a structure and allow organizing knowledge for subsequent retrieval. They are commonly employed in indexing schemes, thesauri, taxonomies and ontologies and are based upon predefined, authorized terms that have been preselected by the designer of the vocabulary. The structure of such vocabularies allows the specification of relationships between concepts that can help when they are employed in tasks where the identification of semantic relationships between terms is important.

i. DeCS

The DeCS - *Descritores em Ciências da Saúde*, is a multilanguage (Portuguese, Spanish and English) controlled vocabulary created by the Pan-American Health Organization (PAHO). It is a superset of the Medical Subject Headings (MeSH [5]) from the NLM – U.S. National Library of Medicine and extends it through new subject areas, mainly in the fields of Public Health and alternative medicines. It presents a fully trilingual structure, whilst maintaining the indexing schemas and indexes of MeSH where they already existed. DeCS was developed in order to allow the use of a unified terminology for simultaneous search in all three languages, Portuguese, Spanish and English, providing a means for language-independent text retrieval and, at the same time, allowing backward compatibility with all MeSH-based indexes. As DeCS is hierarchically structured, searches can be performed on more general or more specific terms, as well as in a whole category of the hierarchy. DeCS indexes 26,664 terms, where 20,895 are common with MeSH [7].

ii. SBC/CSR

In 2009 the Brazilian Cardiology Society – SBC published guidelines for the analysis of electrocardiographic (EKG) examinations and the issuing of EKG findings reports in a document called *Diretrizes da Sociedade Brasileira de Cardiologia sobre Análise e Emissão de laudos Eletrocardiográficos* [12].

Although Telecardiology is one of the most widespread application areas of Telemedicine, terms employed in EKG findings reports are neither covered in DeCS, nor in Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT), Unified Medical Language System (UMLS [13]), Logical Observation Identifiers Names and Codes (LOINC [14]) nor any other well-established medical controlled vocabulary.

For this reason, during the conception phase of the STT/SC, SES/SC created an expert panel, composed mainly of cardiologists that were long-time users of the RCTM, with the task to develop a controlled vocabulary for EKG findings reports based upon the SBC guidelines [15]. It resulted in a vocabulary aligned to the SBC guidelines called *Sociedade Brasileira de Cardiologia/Cyclops Structured Reporting* - SBC/CSR [16], which contains 93 descriptors [15].

III. RELATED WORK

Recently various solutions for the information retrieval in medical systems and on the web in general have been developed. Solutions that relate to the objectives of this work are presented below.

Chen e Papakonstantinou (2011) present a search model where the experienced users specify the query context [17]. The ranking of the results is computed accordingly to keyword statistics collected based upon this specialized context. The search consists of two parts: a context specification, which defines the document set that will be the search context, and a keyword set for the query. The result is a ranked document listing. Experiments were performed using PubMed documents, which were annotated with MeSH terms. The authors demonstrated that context-based searches enhanced the quality of the results.

Korkontzelos *et al.* (2011). present ASCOT, a customized search tool for clinical trials [18]. Clinical documents and observations, which are stored in a database as XML documents, are annotated with UMML terms using the MetaMap UIMA annotator. The Apache Solr search tool [19] is used for data indexing.

Mendonca *et al.* (2012) propose a workflow for the building of a knowledge extraction tool for Internet documents available in Portuguese [20]. The tool has three components: knowledge base, search engine and search interface. This platform uses a domain ontology for document annotation and retrieval, and a prototype in the field of neurological diseases was implemented as a proof-of-concept. The main advantage presented for this approach is the shortening of the search time.

Luo *et al.* (2008) developed MedSearch, a prototype of a search engine for the retrieval of medical documents on the web [21]. The MeSH ontology is employed to index and retrieve documents. During the search process, large queries are transformed into a set of queries of moderated size employing data mining techniques for the sifting out of irrelevant terms and stopwords, while trying to avoid information loss.

Each of the approaches described above, however, meets only partially the requirements that we identified regarding information retrieval at the STT/SC. STT/SC documents exist as various sets of independent data that are stored in three different forms: (a) free text findings reports, (b) DICOM Structured Reporting (DICOM SR) findings reports containing free text as comments, and (c) structured information ranging from DICOM header metadata to patient data and laboratory examination results. The need to perform searches encompassing this heterogeneous data universe motivated the development of a hybrid information retrieval architecture.

Interviews performed with users of the STT/SC identified a set of additional requirements that were also not present in previous works: (r1) the interface should provide customized filters for each examination modality; (r2) the user should be able to select which information is presented in the search results; (r3) statistical information concerning the search results should be presented; and (r4) the search interface should be intuitive in order to allow unexperienced users to perform searches, without knowing the usage of logical operators.

For this purpose we adapted some of the concepts developed in earlier works. Besides the three components and the document annotation process suggested by Mendonca *et al.* [20] we included: a DICOM metadata parser, an automated data indexer and a new method for query processing. As most of the data, being DICOM SR documents or structured information, are already implicitly semantically annotated, the search engine operates mainly as an indexing aid and a retrieval mechanism. Context-based information retrieval is also employed, but differently from Chen and Papakonstantinou [17], in our approach the context is selected automatically since all users possess assigned roles registered in the system, which provide a context for each user. During indexing, stemming techniques suggested by Luo. *et al.* [21] were employed and, as a search engine, Apache Solr was used, as in Korkontzelos *et al.* [18].

IV. Methods

A. Architecture

The architecture is depicted in Figure 1 and is composed of seven main components: a user interface for searching; a query processing mechanism; a search engine; the knowledge bases; a DICOM parser and a data source.

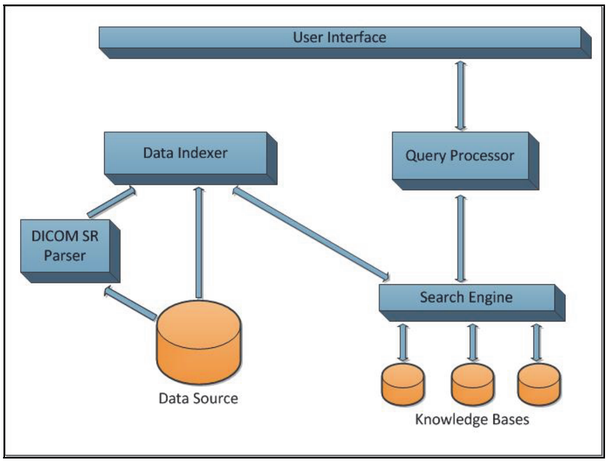


Figure 1. The proposed Architecture

i. Data Source

The totality of the data from the STT/SC, from clinical data through examination and image data, which are all stored in a PostgreSQL database, can be used as a data source. For the retrieval system, indexing occurs only on clinical and demographic information and findings reports.

Clinical informations and patient history for each examination are registered at the moment an examination or a report is sent to the system. This examination submission process is based upon standard protocols and online forms. This provides these data with a predefined semantics. There are, however, a few fields that are free text, mainly concerning previous illnesses, drugs and treatments.

All findings reports being presently provided are entered in DICOM SR format, employing descriptors from DeCS and SBC/CSR. Here there are also free text fields, registered as DICOM SR Comments, encompassing conclusion and additional observations about the patient and the examination, that are also indexed by the system. All reports are stored in a DICOM SR-compliant XML format.

ii. DICOM SR Parser

DICOM SR is a document standard designed to code and interchange clinical information employing the DICOM standard hierarchical structure [8]. It employs IODs, abstractions of the real world that represent patients, examinations, images and reports that are related to a specific DICOM study. DICOM studies are generated or acquired by different modules throughout the STT/SC subsystems. In order to centrally parse and interpret these DICOM objects in our retrieval system, a special parser, based upon tools available at the *dcmtk* suite [22], was developed. It converts DICOM SR content and DICOM metadata into XML documents for further indexing and storage.

iii. Data Indexer

This component is responsible for accessing the data source in order to obtain patient and clinical data and to request data of a particular examination from the DICOM SR parser. Since clinical information are stored in a semi-structured way, some flags stand for concrete information with a well-defined semantics, such as the "tobacco smoking" flag, which can have the values 1, 2 or 3, representing respectively "smoker", "ex-smoker" and "never smoked". For such data with a predefined semantics, special indexing routines were implemented. After all data are obtained and normalized, they are packaged into a XML document and sent to the search engine for indexing.

iv. Search Engine and Knowledge Bases

The search engine, based upon Solr [19], is responsible for indexing, storing the XML documents at the knowledge bases and ranking the results to be presented to the user. Techniques such as stemming, stopwords elimination and synonym detection are applied at this level. Besides indexes, also XML-formatted examination and findings reports data are stored. On each query data totalizers are generated, which are then used for the statistical graphs presented to the user.

v. Query Processor

The query processor is an intermediate component between the user interface and the search engine. It is responsible for normalizing the query accordingly to a predefined syntax, and to submit it to the search engine. It also prepares the statistical data for visualization on the interface using the documents recovered by the search engine.

vi. User Interface

The interface, depicted in figure 2, was developed in order to support both experienced and unexperienced users to perform queries. It consists in a module integrated into the STT/SC and divided into five sections: query terms, knowledge base selection, results, statistics, selected columns and filters.

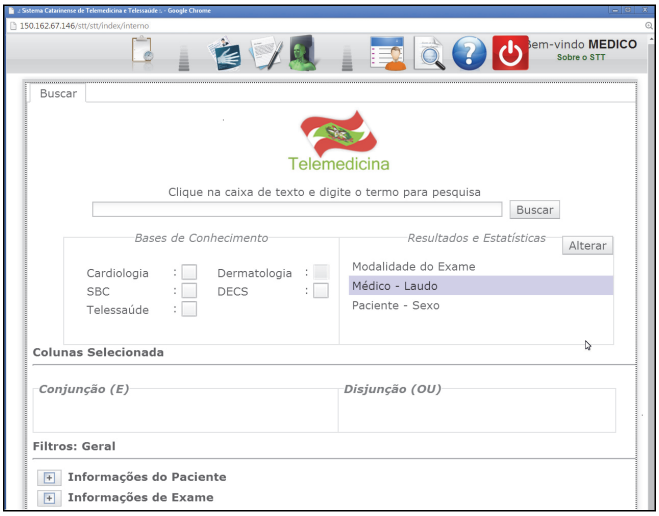
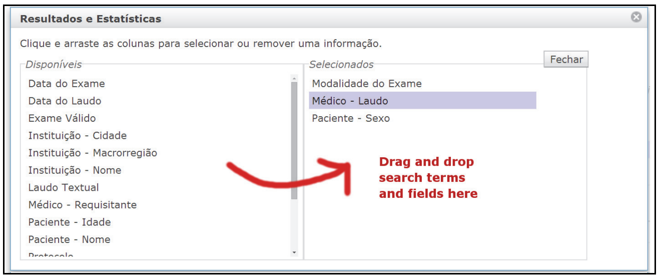


Figure 2. User interface for searching

During knowledge base selection, the system presents only the knowledge bases compatible with the context of the user profile. E.g. a user that is a "reporting physician" with expertise area "Cardiology" will have access only to the cardiology examinations and the SBC/CSR knowledge bases.

In the *Results and Statistics* section, depicted in figure 3, the system presents a drag and drop interface with the information that are available for searching. The user then drags the selected information to the *Selected Fields* window.

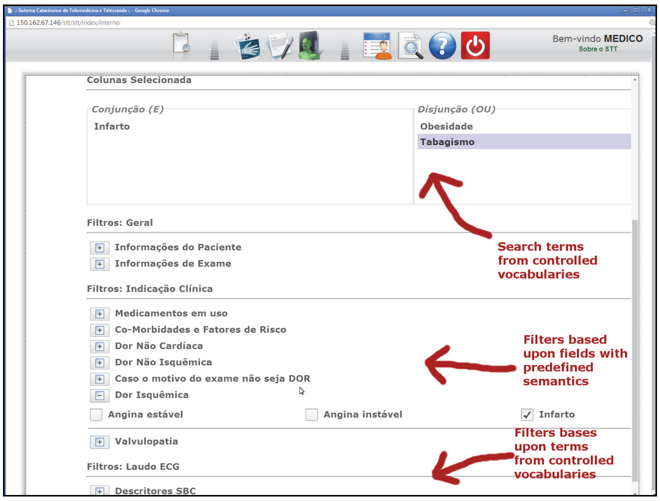
The user can select terms from the controlled vocabularies as well as classifying and filtering variables such as examination modality, patient sex, and pain classification or examination date.



Drag and drop search terms and fields here

Figure 3. Drag and drop search interface

According to the selected knowledge bases and user profile, the system makes filters and search options available for the search refining. Filters are available for all fields with predefined semantics and controlled terms.



Search terms from controlled vocabularies

Filters based upon fields with predefined semantics

Filters bases upon terms from controlled vocabularies

Figure 4. Query Interface –showing a situation where the Cardiology knowledge base was selected.

In order to allow logical operations between selected terms and filters, all selected items appear in the *Selected Columns* section (Figure 4), where the user can define between which terms and filters the logical operations "AND" and "OR" are to be performed. After specification of informations to be searched for and filters, the user is presented with a ranked list of the retrieved documents. For each document, the user can visualize all informations, from findings texts to associated images. Figure 5 shows graphics generated for a given user query.

Through these graphics it is possible to obtain morbidity statistics and their relationship to clinical background information.

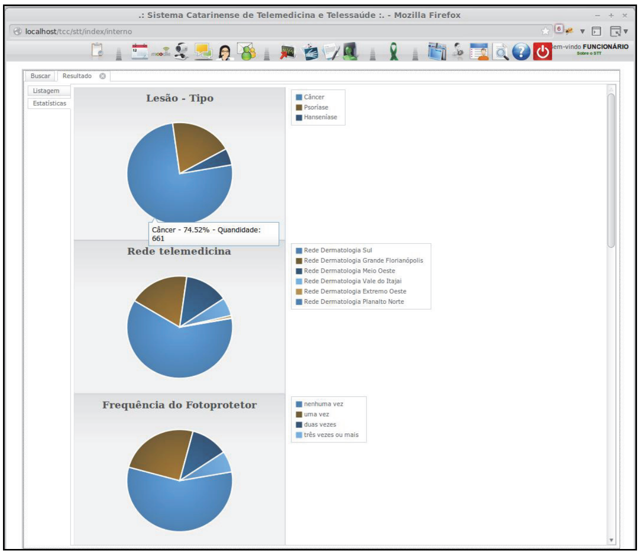

.: Sistema Catarinense de Telemedicina e Telessaúde :. - Mozilla Firefox
localhost/tcc/stt/index/interno
em-vindo FUNCIONÁRIO
Sobre o STT
Buscar
Resultado
Listagem
Estatísticas
Lesão - Tipo
Câncer
Psoriase
Hanseníase
Câncer - 74.52% - Quandidade: 661
Rede telemedicina
Rede Dermatologia Sul
Rede Dermatologia Grande Florianópolis
Rede Dermatologia Meio Oeste
Rede Dermatologia Vale do Itajai
Rede Dermatologia Extremo Oeste
Rede Dermatologia Planalto Norte
Frequência do Fotoprotetor
nenhuma vez
uma vez
duas vezes
três vezes ou mais

Figure 5. Statistic data retrieved for a user query

B. *Validation Strategy*

Validation experiments were performed based upon the GQM – Goals Questions Metrics approach, a goal-oriented method for the measurement of software products and processes[23]. We defined the following evaluation goals:

G1. Adequacy of search performance and

G2. Adequacy of retrieval results.

Both goals were defined from the point of view of medical users and government administration staff. From these goals and points of view, we derived the following questions:

Q1. What is the impact of the tool on the time the user takes to execute a query? and

Q2. What is the precision of results when using the tool?

For Q1 the used metric was response time in minutes and for Q2 the chosen metric was *Precision at k (P@k)*, with k=15. This metric expresses the precision of retrieved documents when the totality of relevant documents is not known[24]. A case study with five searches was defined, with two in the Dermatology and three searches in the Cardiology domain:

S1. Reddish abdominal lesion compatible with Hansen's disease (HD);

S2. Neoplasia located on the scalp;

S3. Angina, hyperhidrosis, and palpitation within the limits of normality;

S4. Hypertension with sinus arrhythmia and

S5. Obese smokers bradycardia with sinus.

C. *Materials*

For the validation study we selected 50,000 EKG and 889 Dermatoscopy examinations from the STT/SC, which were indexed with their findings reports and associated images. These were 50,000 randomly selected EKG examinations from a universe of 490,097 EKG examinations with findings reports written in DICOM SR and the totality of existing Dermatology examinations. All examinations had at least one findings report with CSR/SBC and/or DECS terms inserted by the responsible physician at the moment of the findings report issuing.

V. EXPERIMENTAL RESULTS AND ANALISYS

G1. Query Processing Times

Searches performed with the tool were compared against times necessary to perform database searches employing queries at the database interface. Times were measured in seconds but are expressed in minutes. Since the tool presents statistical graphics of the results, the response time measured for the tool included time necessary for the extraction and export of query data in .csv format, and generation of the graphics in OpenOffice spreadsheet format.

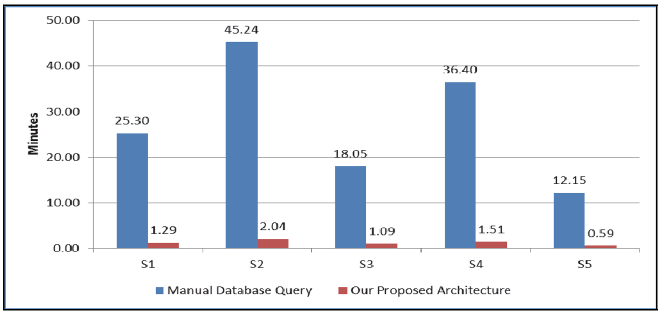


Figure 6. Query times.

The mean time for a manual database query was 27.4 minutes, vastly superior to the mean time to perform a search with the tool, which was of 1.3 minutes. This occurs because, for each database search, at least 15 tables had to be joined in order to obtain the query data, besides various sub queries necessary to obtain statistical data.

G2. Precison of the Results

We conducted experiments to evaluate the accuracy of the proposed architecture in real situations using the 5 searches of the case study defined for Q2.

The mean P@15 precision was 76%. Figure 7 shows that searches S1 and S2 presented results that were not satisfactory. This occurred because many results containing the conclusion “Not compatible with neoplasia” were retrieved. This could be avoided analyzing if search terms are negated or not.

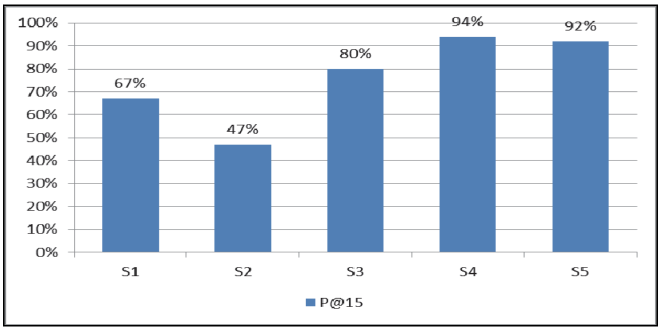


Figure 7. P@15 results for all searches.

VI. Conclusions and Future Work

This work presented a medical information indexing and retrieval architecture and its web-based search interface. With this search mechanism medical staff, public health policymakers and researchers will be able to comfortably extract morbidity information and correlate them with different clinical and demographic data from the database of the STT/SC, which contains more than 2 million examinations and is still growing. Results of our case study show that searches can be performed much faster with the proposed search mechanism and that the precision of results is acceptable in most cases.

In some searches, irrelevant items within the 15 first results were identified. This occurred partially because search terms found in the additional free text observations inserted into the findings reports were treated with the same relevance as formally diagnostically relevant items of the DICOM SR structure and, partially because the semantics of negations associated to search terms in the findings reports were not taken into consideration.

As future work, in order to enhance the precision of results, we intend to implement a mechanism for the detection of negated terms and phrases and a ranking mechanism that treats terms differently, according to their location in the structure of the DICOM SR document, enabling to treat a specific term differently e.g. if it is found in the diagnostic statement or in the clinical observations. As now exists a standardized procedure for findings reports issuing at the STT/SC, employing DICOM SR, which is gradually being extended to all examination modalities presently covered, applicability of this search mechanism will be gradually possible in all areas of Telemedicine practiced in Santa Catarina.

References

[1] Tianyong Hao; Yingying Qu; Fang Xia, "Domain knowledge acquisition by automatic semantic annotating and pattern mining," Information Retrieval& Knowledge Management(CAMP), 2012 International Conference on, vol., no., pp.34,38, 13-15 March2012.

[2] Friedlin, J.; Mahoui, M.; Jones, J.; Jamieson, P., "Knowledge Discovery and Data Mining of Free Text Radiology Reports," Healthcare Informatics, Imaging and Systems Biology(HISB), 2011 First IEEE International Conference on, vol., no., pp.89,96, 26-29 July2011 .

[3] Chi-Huang Chen; Xiao-Ou Ping; Zi-Jun Wang; Sheau-Ling Hsieh; Li-Chin Chen; Yi-Ju Tseng; Ching-Wei Hsu; Feipei Lai, "The keyword-based and semantic-driven data matching approach for assisting structuralizing the textual clinical documents," Biomedical Engineering and Informatics(BMEI), 2010 3rd International Conference on, vol.6, no., pp.2532,2535, 16-18 Oct. 2010.

[4] Hakan Bulu, Adil Alpkocak, Pinar Balci. "Uncertainty modeling for ontology-based mammography annotation with intelligent BI-RADS scoring Computers" in Biology and Medicine, Volume43, Issue4, 1 May2013, Pages301–311.

[5] MeSH. "Medical Subject Headings". Avaliable in: http://www.ncbi.nlm.nih.gov/mesh. Accessing in 2013-12-12.

[6] SNOMED CT. "Systematized Nomenclature of Medicine--Clinical Terms". Avaliable in: http://www.ihtsdo.org/snomed-ct. Accessing in 2013-12-12.

[7] Bireme. "Decs - descritores em ciências da saúde". Avaliable in: http://decs.bvs.br/P/decsweb2012.htm. Acessing in 2013-02-13.

[8] Clunie, D. DICOM structured reporting. PixelMed Publishing, 2000.

[9] Boone K. W. "The HL7 clinical document architecture, The CDA TM Book". Springer, London (2011) pp.17-21.

[10] Guevara-Mendez, D.; Bedoya, O., "Information retrieval on ontology based in keywords", Computing Congress (CCC), 2012 7th Colombian, vol., no., pp.1,5, 1-5 Oct. 2012.

[11] Baeza-Yates, R.; Ribeiro-Neto, B. "Modern Information Retrieval". New York: ACM Press, 1999. 511p.

[12] Pastore, CA. Pinho, C. Germiniani, H. Samesima, N. Mano, R. "Diretrizes da Sociedade Brasileira de Cardiologia sobre Análise e Emissão de Laudos Eletrocardiográficos". Arquivos Brasileiros de Cardiologia. V93, ..3, supl. 2. 2009. ISSN 0066-782X.

[13] Bodenreider, O.; Burgun, A. "Aligning knowledge sources in the UMLS: methods, quantitative results, and applications".Medinfo. 2004.

[14] Huff SM, Rocha Ra, Mcdonald CJ, de Moor GJ, et al. "Development of the Logical Observation Identifier Names and Codes (LOINC) vocabulary". JAMIA. 1998, 5(3): pp. 276-292.

[15] Barcellos CL Jr, von Wangenheim A, Andrade R. "A reliable approach for applying DICOM structured reporting in a large-scale telemedicine network". Presented at the 24th International Symposium on Computer-Based Medical Systems (CBMS), June 27–30, 2011.

[16] Andrade, Rafael. "Vocabulário SBC/CSR". Avaliable in: http://www.incod.ufsc.br/src. Accessing in 2013-12-14.

[17] Chen, L. J.; Papakonstantinou, Y. "Context-sensitive ranking for document retrieval". In: Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. New York, NY, USA: ACM, 2011. (SIGMOD '11), p. 757–768. ISBN 978-1-4503-0661-4.

[18] Korkontzelos, I. et al. "Text mining for efficient search and assisted creation of clinical trials". In: Proceedings of the ACM fifth international workshop on Data and text mining in biomedical informatics. New York, NY, USA: ACM, 2011. (DTMBIO '11), p. 43–50. ISBN 978-1-4503-0960-8.

[19] Lucene. "Apache Solr Lucence". Avaliable in http://lucene.apache.org/solr/. Acessing in 2014-03-16

[20] Mendonca, R. et al. "Towards ontology based health information search in portuguese - a case study in neurologic diseases". In: Information Systems and Technologies (CISTI), 2012 7th Iberian Conference on. [S.l.: s.n.], 2012. p. 1–4. ISSN 2166-0727.

[21] Luo, G. et al. "Medsearch: a specialized search engine for medical information retrieval". In: Proceedings of the 17th ACM conference on Information and knowledge management. New York, NY, USA: ACM, 2008. (CIKM '08), p. 143–152. ISBN 978-1-59593-991-3.

[22] Marco Eichelberg ; Joerg Riesmeier ; Thomas Wilkens ; Andrew J. Hewett ; Andreas Barth ; Peter Jensch. "Ten years of medical imaging standardization and prototypical implementation: the DICOM standard and the OFFIS DICOM toolkit (DCMTK)". Proc. SPIE 5371, Medical Imaging 2004: PACS and Imaging Informatics, 57 (April 19, 2004); doi:10.1117/12.534853.

[23] Basili, V. R.; Caldiera, G.; Rombach, H. D. "The goal question metric approach". In: Encyclopedia of Software Engineering. [S.l.]: Wiley, 1994.

[24] Van Rijsbergen, C.J. "Information retrieval" (2nd. Ed.). Butterworth-Heinemann, Newton, MA, USA, 1979.