

# TelehealthRS Project : Supporting teleconsulting with text mining to improve continuing professional development

Fábio Rafael Damasceno, Eliseo Reategui, Daniel Epstein  
Federal University of Rio Grande do Sul  
Postgraduate Program in Informatics on Education, Brazil  
Porto Alegre, Brazil

Carlos André Aita Schmitz, Erno Harzheim  
Federal University of Rio Grande do Sul  
Postgraduate Program in Epidemiology  
Porto Alegre, Brazil

**Abstract**— In the primary care scenario, telehealth appears as an option for continuing professional development. Questions submitted by physicians and their respective answers in the Brazilian health ministry telehealth platform were mined using a text mining tool. Graphs about the concepts present in the database were extracted using this tool, which gave a clearer view of the subjects addressed by teleconsultants. A questionnaire has been given to teleconsultants asking them about the current protocol they follow in their answering process in the TelehealthRS Project. Results showed that the graphs obtained in the mining process can make teleconsultation answering easier. A recorded interview was conducted with nine physician teleconsultants, deepening the first questionnaire topics. Their comments are now being used to guide the implementation of new features of the current Brazilian health ministry platform. This initiative should impact in the accuracy and response speed offered to professionals, improving their training and contributing to their continuing professional development.

**Keywords**— *Primary care; Text mining; Continuing professional development*

## I. INTRODUCTION

The Brazilian scenario of primary health care (PHC) includes a profile of family health teams (FHT) that often lacks qualification in terms of primary care. The evolution of scientific knowledge is turning professional training in an ongoing process to maintain the quality of services. In this context, telehealth appears as an option for continuing professional development.

Many professionals in primary health care have to deal with clinical questions in their daily routines. In this scenario, the practice of referral is quite common, in which the physician refers the patient to another colleague with a given specialty. Telehealth technology can reduce the number of unnecessary "referrals" and the possibility of iatrogenic problems<sup>1</sup>. However, in telehealth solutions, it is often the case that previous knowledge about answered questions is not

reused in answering new questions. Such limitation represents a gap that is addressed in the work proposed here.

This article is divided as follows: section two presents the concept of telehealth and its role as a tool to implement continuing professional development activities. Section three describes Brazilian epidemiological situation so that one can understand the context of this work; its first subsection presents family and community medicine as a valuable alternative for the Brazilian scenario; the second subsection shows Quaternary Prevention as an important need; the third subsection presents current challenges for the training of health professionals. Section four presents the International Classification of Primary Care (ICPC-2) and the International Statistical Classification of Diseases and Related Health Problems (ICD-10), both used in the telehealth platform. Section five introduces text mining and Sobek, the technology used in this project to help teleconsultants retrieve previous answers in the platform's database.

## II. TELEHEALTH

Health professionals have to answer clinical questions quite often during their practice. Because of that they should have at their disposal a way to express their questions. Telehealth may be considered a viable solution to the problem. Telehealth can be defined as the use of information and telecommunication in health-related activities that may be performed remotely, making it easier for professionals and patients to interact. In Brazil, telehealth includes the provision of care to support family health teams (FHT) in two ways: asynchronous (text) and synchronous (video teleconsulting). Both types are characterized as continuing professional development (CPD) activities. Teleconsulting, whether synchronous or asynchronous, is based on the best available scientific evidence, adapted to local realities and principles primed by the Brazilian Unified health System and primary care. The teleconsulting object may be clinical case discussions or topics related to work processes of the family health teams. The demand of asynchronous teleconsulting generally encompasses issues with a more general character. In synchronous teleconsulting, however, more complex and interdisciplinary

<sup>1</sup>Any patient harm (lethal or not) resulting from medical intervention.

health issues / clinical cases are involved. The interaction in this mode happens after prior appointment via traditional web conferencing tools. There is a subcategory of asynchronous requests in which the requester wants a reading material about a specific subject for the development of some activity of interest (such as update, assembling folders, groups of patients, videos and lectures) [1] [4].

The Brazilian health ministry offers a standard telehealth platform, in which the work proposed has been integrated. The teleconsulting process involves three stages. The initial request is followed by a teleregulation and a subsequent response. The requesting process is generally conducted in a virtual environment, such as the health ministry platform. It involves a professional of a particular unit and one or more health professionals linked to a telehealth center. This process is mediated by a professional who is called teleregulator. He/She mediates the process considering the profession of the requester and the content of the request, choosing the most appropriate teleconsultant. The teleregulator can select from one to five codes to classify the process according to the International Classification of Diseases and the International Classification of primary care. However, these are not search criteria available for the teleconsultants in their answering activities.

The answer to a request should be based on the best available evidence for the context of primary care. A search in the current Brazilian health ministry telehealth database is the first step in formulating the response. Checking if a similar issue has already been solved is an interesting approach, as it can speed up and enhance the whole process. The next step includes the addition of bibliographic references, links and attachments to the answer developed. In the work of [1], it has been shown that in the experience of TelehealthRS, at every two teleconsultations requested by medical professionals, one patient referral to a higher level of care is avoided.

### III. BRAZILIAN EPIDEMIOLOGICAL SITUATION AND PRIMARY CARE

This project has been developed in Brazil, so it is important to put into context the Brazilian epidemiological situation. Brazil has nowadays an ever growing aging population (demographic transition), accompanied by substantial changes in eating patterns, with increased overweight and obesity problems (nutrition transition) linked to physical inactivity and a reduction in acute conditions, as well as an increasing number of chronic health conditions (epidemiologic transition). In this scenario, health care systems have to be able to deal with the problems related to modern epidemics of chronic conditions. In the Brazilian context, the country is facing not only an epidemiological transition, but also a rapid demographic transition. The Brazilian population will continue to grow in the coming decades, with growing numbers of ageing citizens and a corresponding increase in chronic conditions .

#### A. Family and community medicine and primary health care

The analysis of health system indicates that the growth of focal specialization does not bring the expected benefits. Primary care proposed a change to this model [1], being founded on methods, evidence-based practices and socially acceptable

technologies available to all, at costs that the community and the country could afford .

In this scenario, family and community medicine became valuable alternatives. Although the implementation of these programs has been carried out relatively quickly in Brazil, a problem arose in this context: the availability of trained personnel to compose these teams, as well as the selection of people with a work profile not always suitable for the task. This problem compromised the legitimacy of the program, combined with a high number of professionals willing to join the it because of the prospect of immediate employment with salaries above average .

#### B. Quaternary Prevention

The fact that there are health professionals composing family health teams without a proper profile or training has been previously addressed. This is a worrying factor that contributes to the setting of iatrogenic problems, in a national and international level. In the United States, for example, iatrogenic problems have already held the third position among the largest cause of death in the country .

Not all medical interventions benefit people equally, and when excessive or unnecessary treatment is prescribed, they may injure patients. The increasing attention directed to iatrogenic problems led to their inclusion in quaternary prevention.

This concept, proposed in 1995 by Jamouille [2] encompasses various criteria and proposals for dealing with medicalization and intervention excess, both diagnostic and therapeutic. It was made official by the World Organization of National Colleges (WONCA) in 2003. The concept of Quaternary Prevention was developed in order to rescue the former general practice in which one single professional had the possibility to take care of a whole group of people with different types of health problems over time.

Therefore, training for quaternary prevention should become present in continuing professional development in worldwide health systems. Best practices in primary care can be developed and consolidated in the family health strategy, decreasing medicalization and iatrogenic problems. Activities of this kind are barely perceived in Brazil until the present moment .

#### C. Challenges in health professional Training

The implementation of health systems in Latin America has structural deficiencies, especially in the training and development of health professionals. In this context, the Pan American Health Organization (PAHO) began a series of studies in the 1970s to understand the logic of training and professional development of health workers, proposing strategies for approaching education in the health field to its own reality. In 1977, at the Sixth National Conference on health, continuing Education appeared explicitly for the first time, referring to the need for the sanitarian career to incorporate new knowledge into initial training programs. At the Seventh Conference in 1980, continuing Education was tied to projects related to coverage, highlighting proposals aimed at the implementation of programs of teaching-service integration . Continuing Education is defined as something that would

encompass the activities of teaching after graduation, with the purpose of keeping professionals updated, having a strong focus on the teacher-student relationship. Therefore, in a typical situation it does not involve social interaction among students and it does not really take learners' experiences into account. The teaching unit is the individual, without the focus on team structure work. The relationship is in descending order, in which the teacher holds all knowledge. It is up to him to transmit it, so that the student receives the information and this is repeated systematically in their professional lives. Some problematic factors can be observed in this continuing Education methodology such as the use of traditional classes, strongly criticized [17][18].

The creation of a differentiated human resources policy gained momentum in the discussion about the relationship between training and reality in the health scenario. In 1998 the Standing Committee on Postgraduate Medical and Dental Education (SCOPME) stated that continuing Education (CE) was no longer adequate to meet the educational and professional needs of physicians in the current health care setting. It was recommended that CE should be adjusted to a larger and broader context, constituting the continuing professional development (CPD). Still, the Brazilian scenario seems to lack in initiatives of this category. Continuing professional development is a necessity for professionals from all areas, seeking the construction of knowledge and the development of new skills and interdisciplinary work. The permanent update of health professionals is quite complex, due to the speed with which knowledge and technological knowledge are renewed in the health field, as in addition to the distribution of professionals and their services. Studies on the development of new ways to deal with personnel training problems in the health sector propagated throughout Latin America, fostering discussions and the proposal of different solutions in Brazil. In CPD an individual takes control of their own learning and development, engaging in a process of action and reflection. All professionals should have learning opportunities to maintain and enhance their skills [16][21][22].

#### IV. MEDICAL TERMINOLOGIES

In medical research, there is a need to exchange information between different researchers and groups, with the goal of independent analysis or verification of experimental results. The Internet plays an important role in this context. However, there are difficulties in the interoperability of these data - the result of the use of heterogeneous data sources by research groups. The medical terminologies contribute to standardization processes, offering an improvement in such a scenario. However, a feature of this field of research is that these terminological models are developed individually by such groups, giving rise to semantic differences, considering their respective objectives. A Terminology is a list of terms related to concepts of a particular domain. The increasing use of medical data stored in computerized patient records contributed to the need for these data to be controlled and structured.

Currently, when the Teleregulators work in the ministry's health telehealth platform, they can classify a request with the

ICD10 and ICPC-2. However, these are not being used in the answering process made by teleconsultants.

##### A. ICD-10 (*International Classification of Diseases*)

The International Classification of Diseases (ICD), being kept and controlled by the World Health Organization, is considered one of the best known medical terminologies used in medicine. The ICD is a classification of terms related to diagnoses represented in twenty-one chapters arranged according to anatomy and etiology. In the case of ICD-10 there is an alphabetical index, characterizing as a thesaurus. In ICD-10 there are no explicit relationships, each concept is defined by at least one and a maximum of two codes, one related to the etiology and an asterisk on the location of diagnosis.

##### B. ICPC-2 (*International Classification of Primary Care*)

In 1978, the World Health Organization (WHO) invited a group formed by members of the WONCA (World Organization of Colleges, Academies and Academic Associations of General Practitioners / Family Physicians) to develop a specific classification for primary care. The International Classification of Primary Care (ICPC) has opened new horizons in the world of classifications in 1987 by WONCA. The second edition of ICPC was published aiming to create a link with the 10th edition of the ICD and to add criteria for the inclusion and cross-references for most of their items.

#### V. TEXT MINING AND REPRESENTING INFORMATION EXTRACTED FROM TEXTS WITH SOBEK TOOL

Text mining has its earliest origins around 1960 when Hans Luhn and Lauren Doyle realized that the frequency and distribution of relevant words within text information were notorious for their understanding [29][30]. Text mining can be defined as a process in which a user interacts with a document collection over time using a set of analysis tools. Similarly to data mining, text mining aims for extraction of useful information from specific data sources, by identifying and exploiting patterns. In this case, data sources are collections of documents, in which the interesting patterns are found along unstructured textual data in these documents.

The text mining tool Sobek was developed using the  $n$ -simple distance graph model, in which nodes represent the main terms found in the text, and the edges used to link nodes represent adjacency information. Therefore, nodes and edges represent how the terms appear together in the text. The method relies on a parameter  $n$  to extract the compound concepts with more than one word. According to this parameter, a combination of the current word with the  $n$  subsequent words is created, attempting to identify terms represented by the most frequent group of words appearing in the text. In Sobek's graphical representation, nodes that are more relevant are presented in a larger rectangle and in darker color (e.g. the terms "Nuclear", "weapon", "atomic bomb").

#### VI. METHODOLOGY

In an initial step, Sobek has been used to identify patterns of terms in requests done from October, 2012 to January, 2014, totaling 558 requests. A small Thesaurus with medical terminology was built in order to help the tool identify terms that were expressed in the requests in different ways.

#### A. Teleconsultants' Perception of the Graphs Extracted

With the graphs of the requests and answers, TelehealthRS teleconsultants were questioned about their perceptions on how those graphical representations could support them in their daily work. Twenty-six out of thirty-three teleconsultants participated in the assessment. The questions were the following:

1. In the health ministry's Answer Protocol the teleconsultants must query the current database before they formulate an answer for the request. Do you believe this to be important?
2. At present, the database search is based on the 'description' field, where the request is written. Do you believe that the answers to these requests could benefit your work by becoming a search criteria too?
3. In the teleregulator work there is a step in which an ICPC-2/ICD-10 code have to be informed. Do you believe that these codes could benefit your work by becoming another search criteria?
4. Suppose you have to answer the following request: "I would like to know how to address smoking addiction". The following figure is a snippet from the graph extracted from the database of previous questions by medical professionals in the telehealth platform. It shows direct adjacent relationships of the concept 'smoking'.

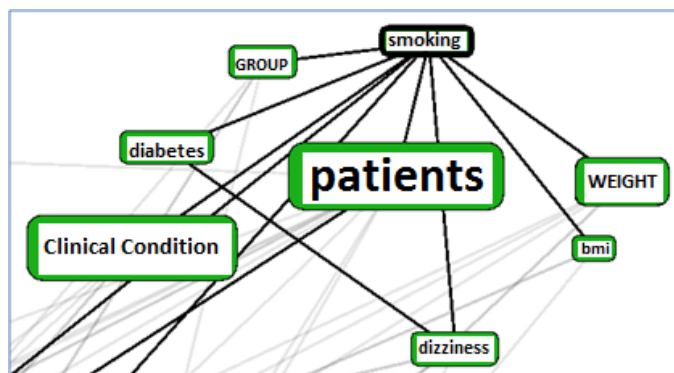


Fig. 2. Graph about the concept 'smoking' extracted from physicians' requests

A relationship between the terms 'smoking', 'diabetes', 'group', 'weight' can be seen, as well other relationships. Do you believe that this graphical representation can help your work in searching for previous requests?

5. The following figure is a snippet from the graph extracted from the database of previous answers developed for medical professionals in the telehealth platform. It shows direct adjacent relationships found for the concept 'smoking' present in the answers made for physician's requests.

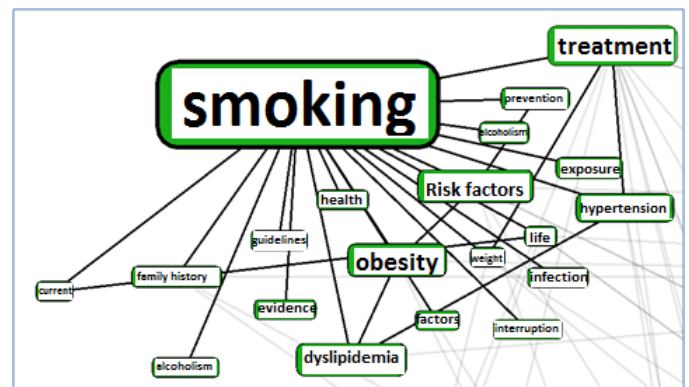


Fig. 3. Graph about the concept 'smoking' extracted from the answers for physician requests.

Currently it is possible to search for requests based on the 'description' field, which is filled out by the health professional. Do you believe that this graphical representation can benefit searching requests?

The teleconsultants could answer each question with 'Yes' or 'No', and they could write an optional commentary for each answer. Figure 4 shows the results obtained:

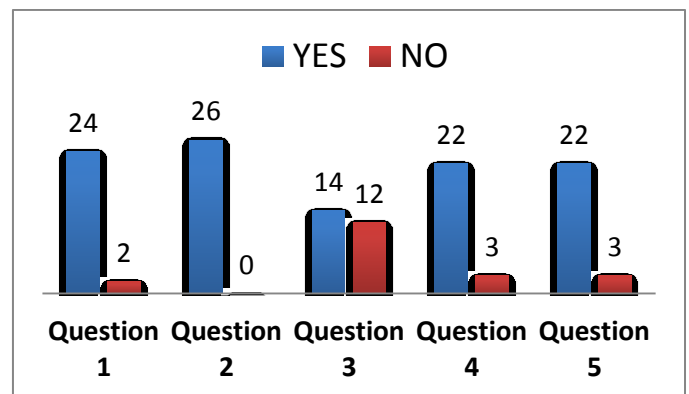


Fig. 4. Teleconsultants' answers

As reported by the teleconsultants, it is difficult to find requests already answered in the past, due to the lack of knowledge about the exact terms that were used in the description field. The relationship among terms can optimize this search according to some of the teleconsultants' comments. It was quoted as a feature that would be very useful to take advantage of the relationships present in the graphs. Besides, the features presented were seen as positive for teleconsultant training. Notice that, for question 4, there was a blank answer suggesting that the participant 'partially agreed', as stated in his commentary. The same happened with question number 5.

The teleconsultants believe, mostly, that analyzing what was previously answered can contribute to the current response being developed. There is a consensus in the sense that the use of search fields based on the answers produced previously can bring benefits to the development of the new answers. There is a clear divergence regarding the use of ICPC-2 and ICD-10 codes as search criteria in the platform and this is further discussed in the next subsection. The vast majority of

teleconsultants agreed that the visual support proposed may help them get a clearer view of what is present in the database requests/answers, and it can help them in the process of answering new requests.

#### B. Interview with Physician Teleconsultants

After answering the initial questionnaire, 9 physician teleconsultants were randomly selected to participate in a recorded interview, with the goal of deepening some topics related to the first questionnaire. The divergence quite apparent on question three was one of the reasons to carry out this subsequent step. The interview was addressed as follows:

1. *Can you describe the way you search for teleconsultations in the health ministry's telehealth platform? If possible, can you point out suggestions and/or difficulties?*

2. *In the last questionnaire a clear divergence about the ICPC-2 and ICD-10 codes as a criteria for searching teleconsultations was noticed. How do you think they can help you in addressing teleconsultations?*

3. *In the last questionnaire there were graphs about the concept 'smoking', which was present in requests made by physicians. A large amount of respondents (89%) showed their appreciation for this type of resource. How do you believe that, during your answering of teleconsultations, this visual aid could benefit you more?*

Relevant comments/statements made by the interviewees are summarized in the three topics below:

1) Question 1: Most of the interviewees pointed out that they search for previous teleconsultations using keywords, trying to remember, informally, if any colleague was working with a similar topic. One teleconsultant affirmed that he always prefers to start from scratch his work. He only refers to previous teleconsultations "when needed". Other two teleconsultants informed that they tried to use the current search in the platform but they felt disappointed with frequent non-useful listed information, considering it impractical. One teleconsultant reported that he always searches for other teleconsultations to try to elaborate more consistent answers. According to them, the text present in requests often lacks clarity, making it harder to identify the focus of the question.

2) Question 2: It was reported by some teleconsultants that ICPC-2 and ICD-10 codes were not so visible during the answering process. However, they think that the use of these codes could improve the whole process, and make search criteria clearer. One teleconsultant believes that this effectiveness depends heavily on the requested contents. Other teleconsultants shared this thought, saying that requests often do not refer to a single point, complicating the usage of the codes. It has also been highlighted that terms presented in requests are often not specific enough. They suggested teleconsultants to classify the incoming processes with ICPC and ICD codes as well.

1) Question 3: The teleconsultants see graphs as an interesting tool, believing that its usage can ease their teleconsulting task. Being able to see a concept and its

relationships with other related terms can help in the search for previous teleconsultations. According to them, if this visual help could be present in the requests/answers, it would enable them to be more accurate in the identification of search criteria. They also mentioned that the graphs would allow a teleconsultant to think about complementary information that could be important to be inserted in the answer being developed. Besides, when a first search does not return a clear set of results, the graphs could help significantly. The time-saving in teleconsultation search was mentioned as a benefit for using the graphs, in addition to being able to see answer patterns and reference materials used by other teleconsultants.

#### VII. FINAL REMARKS

This work proposed a text mining methodology to support the retrieval of previous requests and answers existing in a telehealth platform. The project should be expanded to encompass requests and their respective answers of other professionals – dentists, nurses, community health workers and other health professions present in the health ministry's platform database. We intend to obtain graphs that may be also relevant for other classes of health professionals mentioned earlier, totaling over 2566 requests and responses.

A new module for the telehealth platform will include these graphs, with the intent to present them to the teleconsultants at the time they receive a particular request. Another possibility was perceived during the implementation of the Thesaurus: some medical concepts appearing in the graphs had the same meaning or were acronyms. E.g. AIDS and Acquired Immune Deficiency Syndrome, Infarct and heart attack. Since they are considered individually in the mining process, this kind of information should improve the extraction of graphs now and in future steps of the project.

#### REFERENCES

- [1] Eno Dias Castro Filho et al., "Telessaúde para Atenção Primária," Secretaria de Atenção à Saúde, MINISTÉRIO DA SAÚDE, 2012.
- [2] Francisco Campos, Ana Haddad, Chao Wen, Maria Alkmin, and Maria Cury, "The National Telehealth Program in Brazil: an instrument of support for primary health care," *Latin American Journal of Telehealth*, 2009.
- [3] P. N. Gorman and M. Helfand, "Information seeking in primary care: how physicians choose which clinical questions to pursue and which to leave unanswered," *Med Decision Making*, 1995.
- [4] MINISTÉRIO DA SAÚDE, *Manual da Telessaúde para Atenção Básica/Atenção Primária em Saúde - Protocolo de Resposta.*, 2012a.
- [5] World Health Organization, *International Classification of Diseases, manual of the International Statistical Classification of diseases, injuries and causes of death: 10th revision.*, 1993.

- [6] Mariana Sampaio, Maria Almeida, Cláudia Coeli, Arlinda Moreno, and Kenneth Camargo Jr., "International Classification of Primary Care: A Systematic Review," *MEETING OF THE WHO COLLABORATING CENTRES FOR THE FAMILY OF INTERNATIONAL CLASSIFICATIONS*, Outubro 2009.
- [7] Eugênio Mendes, "As redes de atenção à saúde," *Ciência e Saúde Coletiva*, 2010.
- [8] Maria Anderson, Gustavo Gusso, and Eno Filho, "Medicina de Família e Comunidade: especialistas em integralidade," *Revista Atenção Primária em Saúde*, vol. 8, Janeiro/Junho 2005.
- [9] William González, *ATENCIÓN PRIMARIA DE SALUD EN ACCIÓN*: Editorial Nacional de Salud y Seguridad Social (EDNASSS), 2006.
- [10] Armando Henrique Norman and Charles Tesser, "Prevenção quaternária na atenção primária à saúde: uma necessidade do Sistema Único de Saúde," *Cad. Saúde Pública*, 2009.
- [11] B. Starfield, "Is US health really the best in the world?," *JAMA - Journal of American Medical Association*, 2000.
- [12] Marc Jamouille and M. Roland, "Quaternary prevention and the glossary of general practice/family medicine," in *WONCA congress proceedings*, Hong Kong, 1995.
- [13] Marc Jamouille. (2000) Quaternary prevention: Prevention as you never heard before. [Online]. <http://www.ulb.ac.be/esp/mfsp/quat-en.html>
- [14] Sara Lopes, Érica Piovesan, Luciana Melo, and Márcio Pereira, "Potencialidades da educação permanente para a transformação das práticas de saúde," *Comunicação em Ciências da Saúde*, 2006.
- [15] Aline Massaroli and R. Saupé, "Distinção conceitual: Educação Permanente e Educação Continuada no processo de trabalho em saúde," *II SEMINÁRIO INTERNACIONAL SOBRE O TRABALHO NA ENFERMAGEM*, 2008.
- [16] Cathy Peck, Martha McCall, Belinda McLaren, and Tai Rotem, "Continuing medical education and continuing professional development: international comparisons," *British Medical Journal*, Fevereiro 2000.
- [17] Dave Davis et al., "Impact of Formal Continuing Medical Education," *Journal of American Medical Association*, 1999.
- [18] D. Kanouse and I. Jacoby, "When does information change practitioner's behavior?," *International Journal of Technology Assess Health Care*, 1998.
- [19] Ricardo Burg Ceccim, "Educação Permanente em Saúde: Desafio Ambicioso e Necessário," *Interface - Comunicação, Saúde, Educação*, vol. 9, pp. 161-168, Fevereiro 2004.
- [20] Marluce Alves Oliveira, "Educação à Distância como estratégia para a educação permanente," *Revista Brasileira de Enfermagem - REBEn*, 2007.
- [21] T. Gibbs, D. Bridgen, and D. Hellenberg, "Continuing Professional Development," *South African Family Practice*, 2005.
- [22] J. Grant and F. Stanton, *The effectiveness of continuing professional development Joint Centre for Education in Medicine*. Reino Unido, 1988.
- [23] Yugyung Lee, Kaustubh Supekar, and James Galler, "Ontology integration: Experience with medical terminologies," *Computers in Biology and Medicine*, 2006.
- [24] R. Qamar and A. Rector, "Semantic mapping of clinical model data to biomedical terminologies to facilitate data interoperability," *Healthcare Computing*, 2007.
- [25] N. F. de Keizer, A. Abu-Hanna, and J. H. Zwetsloot-Schonk, "Understanding terminological systems. I: Terminology and typology.," *Methods of Information in Medicine*, 2000.
- [26] P. Moorman, A. van Ginneken, J. van der Lei, and J. van Bommel, "A model for structured data entry based on explicit descriptive knowledge," *Yearbook of Medical Informatics*, 1995.
- [27] Ministério da Saúde, "Classificação Internacional de Atenção Primária - Segunda Edição," Sociedade Brasileira de Medicina da Família e Comunidade (SBMFC), Tradução de: ICPC-2-R (Revised Second Edition): International Classification of Primary Care 2009.
- [28] N. F. de Keizer and A. Abu-Hanna, "Understanding Terminological Systems II :Experience with Conceptual and Formal Representation of Structure," *Methods of Information in Medicine*, 2000.
- [29] Mark Sharp, *Text Mining*.: Rutgers University, 2001.
- [30] Sue Soy. (2003) Automatic Indexing References to the Early Years of Automatic Indexing and Information Retrieval. [Online]. <http://www.gslis.utexas.edu/~ssoy/organizing/l391d2c.htm>
- [31] Ronen Feldman and James Sanger, *The Text Mining Handbook - Advanced Approaches Analyzing Unstructured Data*.: Cambridge University Press, 2007.
- [32] Eliseo Reategui, Miriam Klemann, Daniel Epstein, and A. Lorenzatti, "Sobek: a Text Mining Tool for Educational Applications," *Int'l Conf. Data Mining | DMIN'11*, 2011.
- [33] A. Schenker, "Graph-Theoretic Techniques for Web Content Mining. PhD thesis, , 2003.," University of South Florida, PhD Thesis 2003.