ISSN: 2230-9926

Available online at http://www.journalijdr.com

IJDR

International Journal of Development Research
Vol. 07, Issue, 09, pp.15648-15653, September, 2017

ORIGINAL RESEARCH ARTICLE OPEN ACCESS

ECOM - COMPUTATIONAL MODEL FOR THE AUTOMATIC SELECTION OF CANDIDATE TERMS FROM TEXT MINING TO ASSIST ONTOLOGY BUILDING

[1]Keller S. de Araújo, [2]Eduardo M. de Freitas Jorge, [1]Patricia Freitas Braga, [3]Gabriela Rezende, *[2]Hugo Saba and [1]Renelson R Sampaio

[1]Faculdades Senai Cimatec – Salvador–BA, Brasil
[2] Universidade do Estado da Bahia – UNEB, Salvador – BA, Brasil
[3]Universidade Estadual de Feira de Santana - UEFS, Salvador – BA, Brasil

ARTICLE INFO

Article History:
Received 29th June, 2017
Received in revised form
25th July, 2017
Accepted 19th August, 2017
Published online 30th September, 2017

Keywords:

Ontology,
Text mining,
Knowledge Management.

ABSTRACT

The emergence of the Web in the 90's and its constant dissemination made possible the access to an immensurable quantity of information. However, a large part of the information available on the Web has, as its main feature, content interpretation directed to people and not to software programs. Therefore, this paper presents a tool for selecting terms of an ontology from text mining techniques and knowledge management, applied over collections of unstructured documents written in Brazilian Portuguese.

Corresponding author

Citation: Keller S. de Araújo, Eduardo M. de Freitas Jorge, Patricia Freitas Braga, Gabriela Rezende, Hugo Saba and Renelson R Sampaio. 2017. "Ecom - computational model for the automatic selection of candidate terms from text mining to assist ontology building", *International Journal of Development Research*, 7, (09), 15648-15653.

INTRODUCTION

The Semantic Web, according to (Feigenbaum, 2007), is no different from the World Wide Web, as the former characterizes itself as an accessory available on the Web. The Semantic Web has great use potential because it emerges when people from different fields or competencies perform research on the most varied topics, accepting common schemes to represent the information on their studies. As many groups develop these taxonomies, the Semantic Web tools allow them to join their schemes and translate their terms. The Semantic Web, to achieve its result, must utilize techniques for understanding, classification, and recovery of information; among these factors, text mining and knowledge representation systems stand out, as do ontologies.

Based on the premise of using ontologies in the context of the Semantic Web, two fundamental processes were used to conceive the proposed model: text mining for data extraction and the ontology building of this mined content. Aranha and Passos (2006), conceptualize text mining as the search for patterns in collections of texts using the processing of natural language. To (Gruber apud Breitman, 2005), an ontology is a formal and explicit specification of a shared conceptualization. Ontology building can be seen as an important evolutionary step in the specification of knowledge (Freitas, 2003). Hence, various studies and tools exist that address the building of domain ontologies and their structures, as presented in the following chapter. However, the studies conducted on the two most popular tools in the market today, Protegé and OILEd, allowed to conclude that they require specific knowledge from their users, which leads to a higher complexity during task performance and, consequently, has an impact on the time and cost. There are several studies in the literature related to the field of ontology building, such as the investigations by (Melo *et al*, 2008), (Zancanaro, 2013), (Basegio, 2007), and (Gonçalves, 2008). However, no computational model has been found that associates the techniques of text mining, knowledge management, and ontology building to select candidate terms for domain ontology building. Considering this fact and based on the aforementioned situations, the problem examined in the present study is directed to assist in the process of domain ontology building, aiming to reduce manual intervention, process complexity, and the reliance on experts in the said construction. Based on this problematic, the present paper introduces the building of a tool named Selection of Candidates from Text Mining (Eleição de Candidatos para Ontologia a partir de Mineração de Texto - ECOM), which from a given collection of documents in Brazilian Portuguese selects terms to form an ontology of a specific domain. Furthermore, this paper describes the tool's application in two knowledge domains, mining and environmental impact, and discusses the results obtained from the experiment. Although the tool builds a list of candidate terms and does not build ontologies, the candidate terms may facilitate better mapping of the domain and the consequent building of new ontologies and, therefore, its growth. The content of the present paper is organized as follows: Section 2 describes certain studies related to this research, Section 3 presents the development process and the operation of the ECOM model, Section 4 discusses the application of the tool in two knowledge domains and analyzes the results, and finally, Section 5 states the final considerations.

Related Studies

Several studies have focused on ontologies and ontology building from the application of techniques for text mining and knowledge management. However, the investigations by (Melo *et al*, 2008), (Zancanaro, 2013), (Basegio, 2007), and (Gonçalves, 2008) were selected due to a greater proximity with the problematic discussed in the present study and, among these, due to the page limit for the elaboration of this paper. Only the study by (Basegio, 2007), which was selected because it deals more specifically with the building of ontologies in Brazilian Portuguese, will be discussed. In his study, (Basegio, 2007) proposes the building of ontological structures from texts in Brazilian Portuguese. For this purpose, an approach is built, comprising the following steps: identification of terms, identification of taxonomy relations, and identification of the ontological structure. It should be noted that the author did not have a reliable tool for selecting the corpus and, therefore, has defined beforehand the collection with the following information associated with each word in the document: the word in its original form; the lemma of the original word, that is, the word in its single masculine form; and the grammatical tag of the word (for example, noun, adjective, pronoun, definite article, indefinite article, verb, and adverb, among other parts of speech). To validate this approach, the prototype, created to generate the ontological structures, was applied in two case studies: one with the expert's validation and amendment of the lists of simple and composed terms and the other without the said validation step. After this application, the author concluded that the entirely automated solution would not lead to a high degree of precision, when using only statistical techniques to identify terms and taxonomic relations.

ECOM - Computational Model for the Automatic Selection of Terms

The Selection of Ontology Candidates from Text Mining (Eleição de Candidatos para Ontologia a partir de Mineração de Texto – ECOM) model is divided into two modules: text mining and terms selection. The first module addresses the application of natural language processing (NLP) and mining techniques to collections of texts in Brazilian Portuguese. This module uses NLP and forms of *Clustering* of the most frequent terms in the documents (K-means).The second module comprises steps for terms selection to assist in the process of building domain ontologies, using the results obtained by the first module as an input for the process.

For the conception of ECOM, an analysis and development phase was performed, described hereafter.

ECOM Analysis

In the analysis phase, the functional requirements were defined and collected from the identification of the problem under study from the investigations conducted, with a focus on text mining and ontology building processes; a solution to this problem was then proposed. These requirements can be observed in diagram UML (*Unified Modeling Language*) of the use case represented in Figure 1. Its main use cases will be detailed hereafter.

Create directories

If the domain is valid (the domain is considered valid if it has more than five characters and no blank spaces), the system will create the ontology folder and inside it the folders for each new domain required by the user, where their respective documents will be saved.

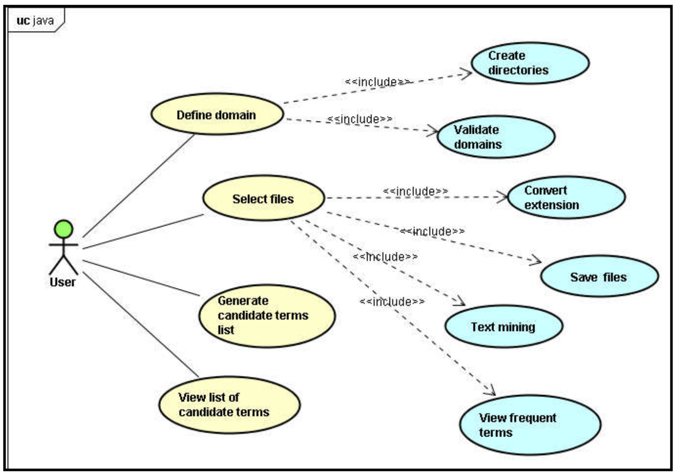

Figure 1. Diagram of the use case

Select files

After validating the domain, the system will show the upload files screen, where the user can add the documents of the collection that can only have a.pdf extension. After the files are uploaded, the use cases included will be performed.

Convert Extension

All documents in the collection are converted to .txt. This conversion was defined due to easy processing and for extension standardization.

Save files

After converting the files, they are saved in the folder created with the domain's name.

Text mining

In this use case, the techniques of text mining are applied, as described in Section 3.2.

View frequent terms

After completing the text mining in the collection of documents made available by the user, a list with the most frequent terms is returned and will be used as a reference through the process of selecting terms for the ontology of the specified domain.

Specification and Algorithm of the Modules of Text Mining and Terms Selection

After completing the analysis of requirements, described in the previous section, the next step was specifying and codifying the text mining and terms selection modules. Notably, all requirements were not known. Therefore, during the development, other functional requirements were identified, such as the restriction of the file extension to .pdf and the display of the candidates terms list in the system itself. For this purpose, the incremental methodology was used in the present project, where, according to Sommerville (2011),each phase is developed as a functionality and, in the end, all phases are integrated because if there is any problem, it will be possible to identify and solve it during the course of the project and not only at the end. The ECOM was developed in the .NET platform, using its language C# and Visual Studio 2010 developing environment, all from Microsoft Corporation. The ECOM has three phases, composed of four steps each, as shown in Figure 2, described hereafter.

Phase 1: Document preparation

The goals of the first phase are to prepare the documents in the collection that the user uploaded to apply mining techniques and generate the BOW (Bag of Words). This phase is subdivided into four steps—combination, saving documents in.txt, indexing documents, and removal of *StopWords*—as described hereafter.

Combination: This step executes one of the knowledge conversion modes, combination, which according to (Takeuch and Nonaka, 2008) is the conversion of a certain type of explicit knowledge generated by an individual to aggregate it with the explicit knowledge of the organization. This type of conversion is also addressed by the theories connected to information processing and occurs through the grouping (classification or summarization) and processing of different types of explicit knowledge. At this moment, the documents are selected by the domain expert, who employs his expertise to analyze the documents that are candidates to belong to the collection. Only then, after reading the documents, will the expert determine the documents that will form the collection that will be analyzed by the model. This way, there is a conversion of explicit to explicit knowledge.

Save documents in .txt - After the texts are uploaded, they are converted to the .txt format and saved in the folder Original Files, created inside the directory tree. This process is conducted to establish a standard for the other phases and for performance gains because with the manipulation of only .txt, there are some savings in the processing.

Document indexing: The indexing process is necessary to support the process performed in phase two, building of the distances matrix. In this step, an indexDoc.txt file is created, with a list of all the files in the collection associated with an index, defined in ascending order and starting at zero. Through this index, each file becomes identified by a number, which facilitates the manipulation of the files in the collection.

Remove *Stop Words*: To obtain a list of words in each document of the collection with a meaning capable of establishing semantic and syntactic relations between them, it is necessary to exclude prepositions, numbers, and special characters (,(,),),f,&,()%,), among other terms. To support this process, in this step, NLP techniques, such as *tokenize*, are applied to substitute special characters with blank spaces; after the *tokenize* process, the *Stop Words* are removed, eliminating all prepositions and articles.

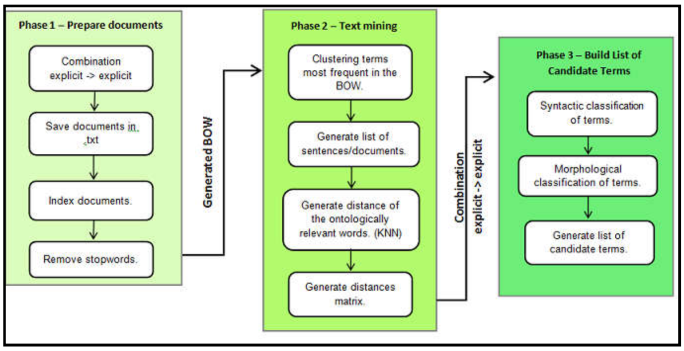

Figure 2. ECOM phases

Each term removed in the processes of *tokenize* and *StopWords* is substituted by a blank space. This substitution was necessary because, in many cases, there were terms separated by special characters (such as in the example below) and, when the character was removed, word joining would occur, which would prevent the word from being identified as a valid term in Brazilian Portuguese, and it would end up being disregarded.

Example

mineração(Descoberta do conhecimento) [mining(Knowledge Discovery)] mineração Descoberta conhecimento. [mining Knowledge Discovery] No inclusion of spaces Mineração Descoberta conhecimento. [Mining Knowledge Discovery] With the inclusion of spaces.

Phase 2: Text mining

The goal of this phase is to submit the collection of documents to the text mining pre-processing and processing steps, through the application of techniques and algorithms such as *stemming*, *tokenize*, *k*-means, and clustering. The output of this phase is the list of more frequent terms in each document and in the collection, in addition to a matrix of distances between the most frequent words and the respective terms associated with distance 1 or 2.As in the example below, the most frequent term in the analysis is "mineração" (mining); then, the term with distance 1 is "dados" (data) and with distance 2 is "recebe" (receive). It is worth noting that the preposition "de" (of) was not considered because in this phase, the *StopWords* have already been removed. There is also a new combination in this phase because the explicit knowledge pre-processed in phase 1 is again refined, but this time automatically.

Example:

A mineração de Dados recebe influências das áreas de Processamento de Linguagem Natural. (Data mining is influenced by the NLP areas.)

***Cluster* the most frequent terms in the documents (K-means)**: In this step, the algorithm k-means is applied over a list of terms in the collection, but before this application, a unique list of all terms is generated, which has already eliminated repeated words and plural words, keeping only its form without plural inflexion that corresponds to the following endings: s, ~oes, ~aes, ~aos, es, is, and e is. This treatment became necessary because it was observed that the algorithm of *clustering* returns groups of words summarized by frequency in the text, without making any type of analysis of the terms. It was established that the algorithm k-means will always work with five clusters performing 500 iterations. These values were defined based on tests with collections of documents of up to 60 files, and it was verified that working with more than five clusters or 1,000 iterations would cause the risk of relevant terms being excluded from the analysis, as they would be grouped with less relevant terms. After *clustering*, the most relevant cluster is defined by calculating the average between the sum of the frequencies and the quantity of words in the cluster. Even by treating the list of terms before *clustering* and performing the algorithm itself, certain repeated terms and flaws were still observed, such as words beginning with <, <,or ", among other characters. Thus, a method to verify each character in the terms returned was created, thus preventing the retention of irrelevant terms in the corpus and, consequently, proceeding to the analysis step. After this step, it was observed that, of the 13,938 terms in the collection, 102 remained after the cluster, and 25 remained after application of the verification method. This observation was possible through the comparative analysis of each list and the calculation of their differences.

Generate list of sentences/documents: In this step, the documents in their complete form were returned, to create a list of sentences of all documents in the collection. The sentences are identified by collections of words, period (.) to period, that is, each period found was understood to signal the end of one sentence and the beginning of another. This way, each set of words between periods was stored in a position in the list.

Generate the distance of words with ontological relevance: In summary, an ontology of a given domain is formed by the terms that have relations that can be semantic or hierarchic in character, that is, for an ontology about mining, it is understood that the terms "filtra" (filter) and "descobrir_Conhecimento" (knowledge_Discovery) have a semantic relation with the domain, whereas data and text are types of mining, which characterize the relation as hierarchic. From this search of relations to form a list of candidate terms to build the domain ontologies, this step seeks the relation between terms based on their proximity. To this end, the list of most frequent terms from step one is matched with the list of sentences from step 2,and a matrix of words and sentences is generated, where each term is associated with the sentences where it appears, maintaining the two words that precede and follow the term.

Generate a distances matrix - From the matrix generated in the previous phase, a second matrix is generated, where column 1 presents the terms already brought up from the previous step and column 2shows the respective matrix with the most frequent words in each term or collection of sentences and the distance, 1 or 2.

Phase 3: Build a list of candidate terms

Syntactic classification of terms - In this step, each term in the matrix of matrixes generated in phase 2 is classified by subject, verb, and direct or indirect object. For this purpose, initially, a *stemming* technique is used, not in its totality, but only when verbs are identified. For the recognition of terms, a list is created with 10,000 verbs; if, when reducing the word to its root, it is found in the referred list, then it is classified as a verb. After identifying the verbs, the words that precede and follow it are verified. In addition to the list of verbs, a list with 45,000 nouns was also created. This way, to identify a subject, the grammatical rule of positioning the terms was used, which has some exceptions that are not covered in its totality in the present study. Thus, what precedes the verb and is found in the list of nouns is considered the subject; the following term is classified as a direct object if it is at distance 1 from the verb, as an indirect object if it is separated from the verb by a preposition, and as a predicate if the verb is in the list of linking verbs.

Morphological classification of terms: This step is performed simultaneously with the previous step, as the nouns and verbs are already identified. In this step, the presence of adjectives in the sentences are verified. Adjectives are terms that add characteristics to the noun, such as “casa bonita” [pretty house](bonita [pretty] = adjective) and “processo rápido” [rapid process] (rápido [rapid] = adjective), among others. There are two ways to recognize an adjective in the sentence: an adjective is linked directly to a noun or is linked to a noun by a linking verb. However, there are many exceptions and rules that, currently, can only be applied by humans, as the example below, where the verb is replaced with a linking verb and it is verified whether the context indicates a term that adds a characteristic—in which case it is classified as an adjective.

Example: "Eram exercícios difíceis." (The exercises were difficult). One asks: What were the exercises? The answer is, "difícil" (difficult). Therefore, "difícil" is an adjective. Thus, to identify the adjectives in a computational algorithm, a list of 100 adjectives was created. One word is considered an adjective when it is linked to a noun either directly or by a linking verb and it is present in the referred list.

Generate list of candidate terms: After the classifications in the previous steps, the algorithm analyses the most frequent terms related with the main terms brought from *clustering* in step 1 of phase 2and its classification. Thus, an execution order is defined. First, all terms related more to distance 1 are selected. Then, the classification of each term is analyzed. If the term is tagged as a verb, it is elected as a candidate term, together with its successor that will be, in the morphological classification, a noun or an adjective. If the term is not a verb, it will be elected only if, in the original text, it is separated from the main term by a preposition because it will signify a complement of the main term or even a type of the latter, as in data mining and text mining, where the data and text confer more meaning to the main term "mining" as they specialize the term. After the analysis of the terms with distance 1, the terms with distance 2 are analyzed, when only terms tagged as nouns or adjectives are considered.

ECOM Validation

To validate the model proposed in the present study, two collections of texts were used with a distinctive main topic, environmental impact and mining, from the areas of environment and information technology, respectively.

Model Application to the Mining Domain

The collection of documents was validated by a professional with a Master's degree who was a PhD candidate in computational modeling; this professional will be referred to throughout the text as validator A. To conduct the tests, the author and validator A selected a sample of 30 documents consisting of dissertations and papers. After the validation of the document collection, the documents were mined by the tool, which resulted in a list of 45 words, represented by the table in the left side of Figure 3. This list was validated by validator A, in which 73.33% of the terms, highlighted in yellow in the table in Figure 3, were considered irrelevant, resulting in 26.67% confidence of the results of ECOM mining in this domain. However, all the words were kept in the process because the final objective was to demonstrate the model's ability to elect terms for building domain ontologies.
After mining, the 45 resultant terms were subjected to the process of syntactic, morphologic classification and the subsequent selection of candidate terms, as described in Section 3.2.This yielded a list of 19 candidate terms for building an ontology of the mining knowledge domain, as shown in the right side table in Figure 3. After the automatic selection of candidate terms, the list generated was analyzed by validator A, and it was observed that the terms selected by the algorithm are, in its totality, related to the domain. However, some English terms, even though strongly associated with the domain, were disregarded because the scope of the present study encompasses only the Brazilian Portuguese language. In addition to the aforementioned analysis, a comparison was performed between the final list from the mining process and the list of elected terms, where it was possible to observe that the terms initially considered irrelevant by the validator were excluded during the process of selecting the candidate terms and, therefore, do not constitute the final list of candidate terms. It should be noted that the figures 3 and 4 were kept in the original language, Portuguese of Brazil, because the purpose of the article was limited to use documents in that language and, therefore, the results were obtained in the same language.

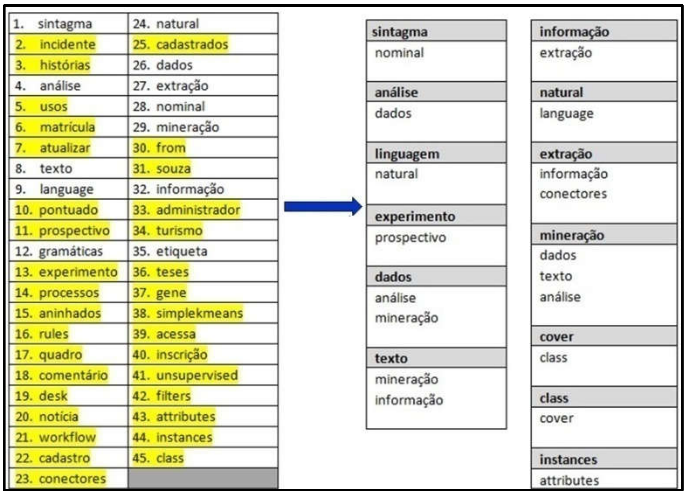

Figure 3. Comparison of the list from mining and the list of candidate terms

Application of the Model in the Environmental domain

The environmental impact domain was chosen because it is outside the context of information technology (IT) and because it was considered important to apply the model to a topic outside the IT context. It should be stressed that the environmental impact domain is one of the domains that has fewer terms in other languages, such as English. The collection of documents was defined by a professional with a Master's degree in urban environmental engineering, who will be referred throughout the text as validator B. To conduct the tests, validator B selected a sample with 20 documents consisting of articles. After the validation of the document collection, the documents were mined by the tool, which yielded a list of 123 words, represented by the table to the left in Figure 4.This list was validated by validator B, where 44.7% of the terms (highlighted in yellow in the table in Figure 4) were considered irrelevant, resulting in 55.3% confidence, for the results of ECOM mining in this domain. However, all of the words were maintained in the process because the final objective was to demonstrate the model's ability to select candidate terms for domain ontology building. After mining, the 123 resultant terms were subjected to the process of syntactic, morphologic classification and the subsequent selection of candidate terms, as described in section 3.2.This procedure yielded a list of 31 candidate terms, represented by the table on the right side in Figure 3, to build an environmental impact domain ontology. After the automatic selection of candidate terms, the list generated was analyzed by validator B; it was observed that the terms selected by the algorithm are, in totality, related to the domain. In addition to the aforementioned analysis and according to Figure 4, a comparison was performed between the final list from the mining process and the list of selected terms, where it was possible to observe that the terms initially considered irrelevant by the validator were excluded during the process of selecting candidate terms and, therefore, do not constitute the final list of candidate terms.

Final Considerations

Considering the analysis performed in the model results, the level of complexity in the process for extracting textual data, and the relationship between them, the result of the mining process and the subsequent selection of candidate terms for ontology building were verified and considered correct.

1.drenagem	25.questões	49.energia	73.serrote	97.brasileira
2.consumo	26.econômicos	50.planeta	74.todos	98.pluviais
3.helmintíases	27.emprego	51.braço	75.social	99.industrial
4.atendida	28.cerrado	52.acesse	76.tempo	100.redução
5.bairros	29.causal	53.etapas	77.proliferação	101.emissão
6.doença	30.condições	54.expert	78.ministério	102.regiões
7.ambiental	31.direto	55.agricultura	79.participação	103.terra
8.floresta	32.incidências	56.agrícola	80.dezembro	104.crescimento
9.degradação	33.incubadora	57.hidrelétricas	81.resíduos	105.cerca
10.rios	34.pública	58.perda	82.município	106.ambientes
11.agronegócio	35.sistema	59.instituto	83.mata	107.metano
12.efeito	36.aquecimento	60.econômica	84.Paulo	108.febre
13.emissões	37.malária	61.turbinas	85.cabral	109.expansão
14.amazônia	38.classes	62.barra	86.classificação	110.modelo
15.água	39.tonelada	63.lançamento	87.média	111.produção
16.pressão	40.hídrica	64.alagadiço	88.barreiro	112.poluição
17.municipal	41.causais	65.planejamento	89.esgotos	113.milhões
18.gases	42.contaminação	66.reservatório	90.relação	114.matos
19.infecciosas	43.serviço	67.sólidos	91.empoçamento	115.saúde
20.plano	44.saneamento	68.disponibilidade	92.conflito	116.casos
21.desmatamento	45.superfície	69.amarela	93.pelas	117.concentraçã o
22.renda	46.ingepro	70.dengue	94.fonte	118.atmosfera
23.ocorrência	47.quais	71.sociedade	95.ferreira	119.carbono
24.alagoinhas	48.vetores	72.tucuruí	96.urbana	120.estufa
				121.civil
				122.valor
				123.contato

drenagem
águas pluviais

helmintíases
intestinais

ambiental
classificação
saneamento
municipal

efeito estufa
gases

água
abastecimento
sistema

municipal
plano
saneamento
ambiental

infecciosas
intestinais
helmintíases

saneamento
ambiental
municipal
alagoinhas
plano

lançamento
resíduos

resíduos
sólidos

classificação
ambiental

urbana
amarela
febre
dengue

pluviais
águas
drenagem

febre
amarela

saúde
pública

estufa
efeito
gases

doença
amarela
dengue

sistema
abastecimento
água

Figure 1. Comparison of mining list and candidate terms list

This allows for the conclusion that the model proposed in the present study, the ECOM, achieved its main goal, which is to facilitate the process of building domain ontologies, aiming to reduce manual intervention, process complexity, and the reliance on experts in the referred building process. Considering this goal, although the validation of the collection of documents and the list of terms was performed manually by the validators, the mining process and posterior selection of candidate terms occurred without their intervention. The present study is not yet final because it opens future possibilities for research, including the following goals:1) to allow the feedback of terms, where the user, from the recurrence of the most frequent terms already presented in the model, can choose to exclude certain terms and establish relationships between them, as it is known that the ontology-building process strongly depends on the relationships established between the terms, which are still too complex to be defined by automated means and2) to perform the identification of hierarchic relationships between the terms to build semantic nets so that the user can view the relationships generated from the results of mining the document collection. These networks may allow the display of relationships between the documents in the collection and their terms.

REFERENCES

Aranha, C., Passos, E. 2006. "A tecnologia de mineração de textos" [Text-mining technology], RESI-Revista Eletrônica de Sistemas de Informação, n. 2.

Baségio, T. L. Uma Abordagem Semiautomática para Identificação de Estruturas Ontológicas a partir de Textos na Língua Portuguesa do Brasil [A Semi-automatic Approach for Identifying Ontological Structures from Texts in Brazilian Portuguese]. Thesis (Master's Degree) Rio Grande do Sul Catholic University, 2007.

Berners-Lee, T., Hendler, J., Lassila, O. 2001. The semantic web.

Braga, F. Patrícia. Um Modelo Computacional para Extração Textual e Construção de Redes Sociais e Complexas [A Computational Model for Textual Extraction and Building of Social and Complex Networks]. Thesis (Master's Degree) SENAI CIMATEC, 2010

Breitman, K. 2005. Web Semântica a Internet do Futuro [Semantic Web: the Future Internet]. Rio de Janeiro: LTC

Feigenbaum, Lee, Ivan Herman, T. H. E. N. & Stephens, S. 2007. "The Web Semantic in action. Scientific American.

Freitas, de Frederico Luiz Gonçalves. 2003. "Ontologias e a Web Semântica" [Ontologies and the Semantic Web].

Melo de Gerard, Suchanek, F. M, Pease. A.2008. Integrating YAGO into the Suggested Upper Merged Ontology. In ICTAI.

Smith, M. K; Welty, C; Mcguiness, D. L. 2004. "OWL Web Ontology Language Guide". Available at: http://www.w3.org/TR/2004/REC-owl-guide-20040210/, accessed in July of 2012.

Sommerville, I. Engenharia de Software [Software Engineering] 9. ed. [S.l.: s.n.], 2011.

Nonaka, I. and H. Takeuchi, 2008. Gestão do conhecimento. Porto Alegre, Bookman.

Zancanaro, A., Erpen, G. J., Santos, S.L.J., Steil, V. A. and Todesco, L. J. 2013. "Mapeamento da Produção Científica sobre Memoria Organizacional e Científica" [Scientific Production Mapping over Organizational and Scientific Memory]. Perspectivas em Ciência da Informação [Perspectives in Information Science].

****