Towards A Model Of Knowledge Extraction Of Text Mining For Palliative Care Patients In Panama.

Denis Cedeno Moreno, Miguel Vargas-Lombardo

Abstract: Solutions using information technology is an innovative way to manage the information hospice patients in hospitals in Panama. The application of techniques of text mining for the domain of medicine, especially information from electronic health records of patients in palliative care is one of the most recent and promising research areas for the analysis of textual data. Text mining is based on new knowledge extraction from unstructured natural language data. We may also create ontologies to describe the terminology and knowledge in a given domain. In an ontology, conceptualization of a domain that may be general or specific formalized. Knowledge can be used for decision making by health specialists or can help in research topics for improving the health system.

Index Terms: electronic health records, knowledge, ontology, palliative care, text mining

1 Introduction

The man from the early centuries of its existence has managed to save, seeds, fruits, tools, nowadays stores information. This information is from different sources and in different formats, electronically or on paper. It is known that knowledge [1] is a treasure for humans and have it means to have control of the situation. To have this knowledge depends on the ability of human beings in certain tasks with information, know where to look, to summarize large volumes of information and make knowledge [2]. With the development of innovative technological tools related to information processing it is possible to access and analyze large volumes of data related to patients in a health institution, many of these tools are based on the extraction of knowledge from textual information sources by applying computational linguistics type. Computational linguistics [3,4] focuses primarily on the design of mechanisms that allow computers to understand natural language, as well as various information processing tasks. Health Information Technology (HIT) along with other innovations in computer technology have the potential to improve the quality of modern health information systems. Text mining (TM) is an application of computational linguistics and word processing that seeks to facilitate the identification and extraction of new knowledge from collections of documents or text corpora. The main objective of this paper is to present the design of a model for knowledge extraction, process based text mining and development of an ontology that allows find patterns and obtain information from medical records of patients units of palliative care. The remainder of this manuscript is structured as follows: The following section describes the methodology to do this job. In the next section, the experiences of the design of the proposed model are shown. After that, a discussion is presented. Finally, we draw the main conclusions of this work.

- *Denis Cedeno Moreno*
- *Doctoral Students of Engineering Projects*
- *Research Group Electronics Health and Supercomputing, Technological University of Panama, Panama City Panama E-mail: denis.cedeno@utp.ac.pa*

2. Literature Review

2.1 Text Mining

Enormous amount of data generated [5,6], in health organizations [7, 8], it is necessary the application of powerful new methods of processing and access to information [9], with the advancement of technologies related to the information you can access and analyze this data related to health and disease [10]. Is common to record patient data electronically [11], this includes general information, but also those related to diagnosis, severity, analytical results, functional tests and medication, and the characteristics of the contacts of the patient [12], this wealth of information in digital format have three major advantages: improved quality, reduced working time and use information through automated systems, such as Text Mining [13, 14]. TM [15, 16], it is defined as the process of discovering interesting patterns and new knowledge in a collection of texts. Is the most recent area of research word processing, is responsible knowledge discovery [17], process that does not explicitly exist in the texts, but that arise relate the content of several of them. It is through the process of analysis that value is added to the information to turn it into knowledge [18], only computers can quickly manipulate large amounts of text, the process of TM consists of two main stages: the preprocessing and discovery phase [19]. In the first phase, the texts become some sort of representation structured or semi-structured then facilitates analysis, while in the second phase of the intermediate representations are analyzed in order to discover in them some interesting patterns or new knowledge. TM is a multidisciplinary area [20], which has experienced an exponential increase in the production of information, along with information technology have led to complex systems management and provision of information for various tasks. Like most of the information (over 80%) is currently stored as text, it is believed that TM [21], has great commercial value. TM is a technical analysis of information that is identified with knowledge discovery in texts, processing large volumes of unstructured free text to extract knowledge requires the application of a number of techniques including analysis include Information Retrieval (IR), the Natural Language Processing (NLP) [22,23,24,25], Information Extraction (IE). IR systems that identify documents in a collection that matches a user's query. NLP is the analysis of human language to achieve that computers understand natural

language as humans do. The role of the NLP in TM is to provide systems in the phase information extraction of linguistic data they need to perform their task. IE is the process of automatically obtaining structured document from a structured natural language data.

2.2 Text Mining and Electronic Health Record

Medical valuable information entered in the Electronic Health Record (EHR). There is an important source of data that can be used for research and quality of health services, EHR [26,27,28]. The data elements of an EHR vary from one system to another, however, doctors usually store the following data: age, sex, diagnosis, medical history, prescription drugs, laboratory, clinical procedures, results, allergies, immunizations, vital signs and observations [29], palliative care specialist write in EHR annotations such as observations of patients on medical treatment, signs and symptoms. The information in the log may be in narrative form or semi-structured way [30,31,32]. With them, you can perform analysis and extract information in order to improve both the tasks of medical and scientific research. In addition to its clinical use by the doctor, the EHR, can be used as a repository for medical information [33, 34] about many patients and provides rich data waiting to be analyzed and mined for clinical discovery. The methods of text analysis are intended to extract useful information from the wide range of unstructured documents and free format generated daily in hospitals. Medical researchers [35, 36] are on the threshold of a new era in which the EHR are gaining an important role in supporting daily activities. New technology and emerging innovative infrastructure tools are being developed to ensure that the EHR can be used for secondary purposes such as medical research, including the design and implementation of clinical trials of new drugs. We are currently on the edge of a golden era of understanding of medical documents, with a wealth of information available to support the processes of health [36]. Computer and information management [37], tools are now part of the world of biomedical science [38]. Platforms and computing infrastructures allow new types of experiments that were impossible to do ten years ago. Progress in the area of HIT [39], undoubtedly has enabled advances in patient documentation [40]. This generation of electronic health data is a great promise for contributing significantly to the health of patients [41], but also to transform biomedical research [42]. The main challenge for the medical TM in the coming years is to make such systems useful for researchers. TM has emerged as a potential solution for EHR based on patients suffering from diseases such as cancer systems [43].

2.3 Text Mining and Cancer

TM can help to gain knowledge of a mountain of text and its use is now widely applied in biomedical research [44]. Many researchers have used the TM technology to discover new knowledge and improve the development of biomedical research, especially those relating to malignant diseases such as cancer [45]. Cancer is a deadly disease it caused 7.4 million deaths in 2008 [46]. For this reason, cancer is one of the most important areas of study for biomedical researchers. With so much text on this disease, it is almost impossible for doctors to investigate all these documents and discovered significant new knowledge. TM can help researchers to complete this difficult task [47]. The researcher can realize the advantages of the TM and facilitating research and helps to find new knowledge [48], to the diagnosis, treatment and prevention of cancer [49], TM employs many computing technologies such as machine learning, natural language processing, biostatistics, information technology, and pattern recognition to find new hidden results in biomedical structured text [50, 51]. In recent years, TM and statistical analysis [52] has been applied to large areas of medicine, due to the existence of large amounts of data. Because TM provides a mechanism to transform text into knowledge, it is becoming a way to explore, analyze, consult and manage information from a large number of medical fields, including in some cases data from patients of palliative care with cancer [51].

2.4 Knowledge Representation

Convergence various areas of knowledge [53], has led to the design and implementation of computer systems that support the integration of heterogeneous databases with medical information. TM has been to contribute to the development of technologies and includes different techniques made in the field of recovery and textual computational linguistics [3, 54]. The TM oriented knowledge extraction from unstructured in natural language stored in textual databases data, identifies with knowledge discovery in texts and is commonly referred to as Knowledge-Discovery in Text (KDT)[55,56,57].

3. Results and Discussion

3.1 Palliative Care specialists

Currently palliative care medical specialists have a complete record of the activities and interactions of both patients and other physicians involved in the assessment process [58]. Further analysis of this set of activities and interactions allows us to understand what happened in a certain discussion or activity. However, when you have a considerable amount of manual analysis interactions is practically impossible due to the time and effort required task. The application of TM techniques [59, 60], to the domain of cancer is one of the newest and most promising areas of research for the analysis of the data. In the country, diagnoses, procedures and investigations and the outcome of all is published in text format they are made [61]. It has also increased the distribution of medical information in different types of documents, not only in scientific articles, also in electronic medical records. Because most of the information about patient data palliative care is in text documents, the application of powerful new methods of processing and access to information is necessary. TM is presented as emerging technologies that support for the discovery of knowledge possessed stored data.

3.2 Propose Model

This tool allows you to analyze text elements in order to identify and expand knowledge deduct from any organization of documents, the functions that should primarily meet textual mining tool include: Identify "facts" and point from the text of the document data, to which we refer. Group similar documents (clustering). Determine the topic or topics covered in the documents by automatically categorizing the texts. Identify the concepts discussed in the documents and networking concepts. Facilitate access to the distributed information in the documents in the collection, using automatic abstracting, and visualization of the relationships between the concepts covered in the collection. In our project to discover

knowledge in text documents, must go through several important stages (see table 1) that will induce this process, we define these stages as pre-processing through this process will give an intermediate form text that allows it to be treated computationally then apply some text mining technique and finally display the result [62,63]. These stages can be seen in Figure 1.

Table 1. *Stage of TM*

Stage	Description
Pre process	Select those terms that best represent the objects of study documents also eliminates irrelevant information and will carry out operations or transformations on the text, to generate a representation or semi-structured then facilitates analysis.
Discovery	This is the stage where the intermediate representations will be analyzed in order to discover in them some interesting patterns or new knowledge.
View the results	Step exploration results that are presented to users.

Several processes or stages on which works the TM, consider structuring the content of the texts is an intermediate representation model and an essential process in TM [64] on this basis is that the algorithms or methods of discovery as ontologies may apply. The most commonly used today for structuring the content models are the vector space model, as well as taxonomies of concepts and graphs [65, 66].

Figure 1. *Stages of Text Mining*

We consider addressing in this project to the process of structuring the content handling it with the application of a new model, since finding new alternatives for representation or structuring of texts [67] allows innovation in information retrieval, and achieves that facilitate the processes that lead us to discover knowledge. The structuring process [68, 69] will handle it in several phases; the first phase of the new model will be based on an application developed in Java programming language. With the use of specialized classes to separate or break the elements of a text string, tokenization, which receives as input the text file and produces an output consisting of tokens or symbols. In this regard, a first approximation of a word representation model and a method for its construction is defined automatically. These proposed model can be seen in Figure 2. In the model, input is given by a text file with the corpus of what you want to process, this text file will contain the comments made by medical specialists of palliative care, specifically the area where the observations of the patient are specified in EHR. As it noted in the EHR of each patient attending the palliative care unit and is staffed by a physician, detailing each of these observations. We want to gather a large sample of a group of patients to be stored in this file (corpus) that function as input element to the process.

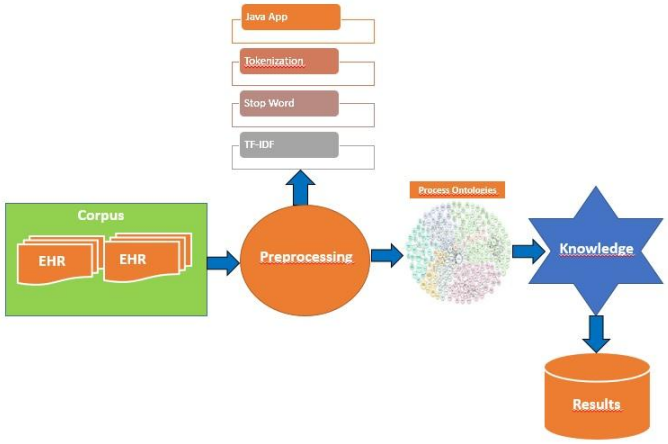

Figure 2. *Proposed Model*

Once you obtained the file through an application in Java programming language and using tools such as tokenization classes, we will read the file to which we go apart in basic units (word for word) to form a new structure. A new structure that will contain the different elements of our corpus store it in a text file. Then it passes through a phase that achieves eliminate concepts or elements with minor or relevance for us, because not all the words of our EHR are equally representative of the corpus that we will evaluate. We are doing is essentially a lexical analysis of our initial text, whose purpose is the treatment of numbers, hyphens, punctuation, capitalization or lowercase words, proper names. They are then eliminated the concepts that have less relevance, in their contextual link with other concepts, it is what is known as empty words. What we have is a list of relevant words for our project; each word will assign different weights. To do this we will use the techniques weighting model Term Frequency-Inverse document frequency (TF-IDF) [70, 71] to assess the importance of each word in the corpus of documents. In the second phase, we will focus on knowledge deduced by identifying patterns and information. Revealing implicit knowledge that is not noticeable from a single document, but from a series of documents. It is a reality that a lot of TM tools used ontologies [4, 72] for the pursuit of knowledge. In ontologies, the words describe concepts that define formally the relationships and rules that specify dependencies between concepts. Ontologies are used to structure and categorize the information specific to a domain. Many TM applications have implemented ontologies [73] in their workflows to structure your search strategy, visualization and classification of information. We believe that developing a specific ontology with information EHR palliative care patients achieve recovery of this corpus of knowledge [74]. Ontologies enable query expansion, reformulation of a query to improve retrieval performance keyword. The use of an ontology in the project, will focus on systematically organize knowledge from a set of terms, concepts and relationships between them. Because ontologies define and establish complex relationships, incorporating rules and axioms that no other linguistic elements naturalizes, obtaining a formal representation of concepts and relationships between them.

4. Conclusion

True knowledge extraction from large volumes of text demand using innovative mechanisms in information processing and text mining or the application of artificial intelligence techniques. The information associated with a context and an experience as it is the process of palliative care specialists, knowledge becomes becoming an intangible resource that provides real value to the organization. The knowledge base that will manage the model can be used for future research in the field of palliative care. The development of ontology proposal provides a basis for other domain ontologies and palliative care can be very beneficial for future research conducted in the region by other specialists in the area of palliative care.

5. Acknowledgement

The author wishes to acknowledge the support of the Group Electronics Research and Health Supercomputing of Technological University of Panama, especially its director Dr. Miguel Vargas-Lombardo for his advice.

6. References

[1] S. M. Allameh and S. M. Zare, “Examining the impact of KM enablers on knowledge management processes,” Procedia Comput. Sci., vol. 3, pp. 1211–1223, 2011.

[2] M. Terzieva, “Project Knowledge Management: How Organizations Learn from Experience,” Procedia Technol., vol. 16, pp. 1086–1095, 2014.

[3] K. a. Dill, A. Lucas, J. Hockenmaier, L. Huang, D. Chiang, and A. K. Joshi, “Computational linguistics: A new tool for exploring biopolymer structures and statistical mechanics,” Polymer (Guildf)., vol. 48, no. 15, pp. 4289–4300, 2007.

[4] I. Peñalver-Martinez, F. Garcia-Sanchez, R. Valencia-Garcia, M. Á. Rodríguez-García, V. Moreno, A. Fraga, and J. L. Sánchez-Cervantes, “Feature-based opinion mining through ontologies,” Expert Syst. Appl., vol. 41, no. 13, pp. 5995–6008, 2014.

[5] N. Zhong, S. Member, Y. Li, and S. Wu, “Effective pattern discovery for text mining,” 2010.

[6] H. Hashimi, A. Hafez, and H. Mathkour, “Selection criteria for text mining approaches,” Comput. Human Behav., vol. 51, pp. 729–733, 2015.

[7] T. Suzuki, H. Yokoi, S. Fujita, and K. Takabayashi, “Automatic DPC code selection from electronic medical records: Text mining trial of discharge summary,” Methods Inf. Med., vol. 47, pp. 541–548, 2008.

[8] M. Krallinger, A. Morgan, L. Smith, F. Leitner, L. Tanabe, J. Wilbur, L. Hirschman, and A. Valencia, “Evaluation of text-mining systems for biology: overview of the Second BioCreative community challenge.,” Genome Biol., vol. 9 Suppl 2, no. Suppl 2, p. S1, 2008.

[9] M. Krallinger, A. Valencia, and L. Hirschman, “Linking genes to literature: text mining, information extraction, and retrieval applications for biology.,” Genome Biol., vol. 9 Suppl 2, no. Suppl 2, p. S8, 2008.

[10] Y. H. Tseng, C. J. Lin, and Y. I. Lin, “Text mining techniques for patent analysis,” Inf. Process. Manag., vol. 43, no. September, pp. 1216–1247, 2007.

[11] S. Ananiadou, D. B. Kell, and J. Tsujii, “Text mining and its potential applications in systems biology.,” Trends Biotechnol., vol. 24, no. 12, pp. 571–579, 2006.

[12] S. Ananiadou and J. McNaught, “Text Mining for Biology and Biomedicine,” Comput. Linguist., vol. 33, pp. 135–140, 2006.

[13] G. Akçapınar, “How automated feedback through text mining changes plagiaristic behavior in online assignments,” Comput. Educ., vol. 87, pp. 123–130, 2015.

[14] V. Gupta and G. S. Lehal, “A survey of text mining techniques and applications,” J. Emerg. Technol. Web Intell., vol. 1, no. 1, pp. 60–76, 2009.

[15] Q. Mei and C. Zhai, “Discovering evolutionary theme patterns from text: an exploration of temporal text mining,” Proc. Elev. ACM SIGKDD Int. Conf. Knowl. Discov. data Min., pp. 198–207, 2005.

[16] H.-C. Yang, C.-H. Lee, and H.-W. Hsiao, “Incorporating Self-Organizing Map with Text Mining Techniques for Text Hierarchy Generation,” Appl. Soft Comput., vol. 34, pp. 251–259, 2015.

[17] J. I. Guerrero, C. León, I. Monedero, F. Biscarri, and J. Biscarri, “Improving Knowledge-Based Systems with statistical techniques, text mining, and neural networks for non-technical loss detection,” Knowledge-Based Syst., vol. 71, pp. 376–388, 2014.

[18] R. J. Mooney and R. Bunescu, “Mining knowledge from text using information extraction,” ACM SIGKDD Explor. Newsl., vol. 7, no. 1, pp. 3–10, 2005.

[19] H. Mahgoub, H. Mahgoub, N. Ismail, N. Ismail, F. Torkey, and F. Torkey, “A Text Mining Technique Using Association Rules Extraction,” World Health, pp. 21–28, 2008.

[20] H. (National U. of S. Liu, H. (Osaka U. Motoda, R. Setiono, and Z. Zhao, “Feature Selection : An Ever Evolving Frontier in Data Mining,” J. Mach. Learn. Res. Work. Conf. Proc. 10 Fourth Work. Featur. Sel. Data Min., pp. 4–13, 2010.

[21] M. Reinberger and P. Spyns, “Unsupervised text mining for the learning of dogma-inspired ontologies,” Ontol. Learn. from Text Methods, Appl. Eval., no. September, pp. 29–43, 2005.

[22] M. Krallinger, R. a a Erhardt, and A. Valencia, “Text-mining approaches in molecular biology and biomedicine,” Drug Discov. Today, vol. 10, no. 6, pp. 439–445, 2005.

[23] T. Baldwin, P. Cook, B. Han, A. Harwood, S. Karunasekera, and M. Moshtaghi, A Support Platform for Event Detection using Social Intelligence. 2012.

[24] Y. Yang, L. Akers, T. Klose, and C. Barcelon Yang, “Text mining and visualization tools - Impressions of emerging capabilities,” World Pat. Inf., vol. 30, no. September 2015, pp. 280–293, 2008.

[25] A. Stavrianou, P. Andritsos, and N. Nicoloyannis, “Overview and semantic issues of text mining,” ACM SIGMOD Rec., vol. 36, no. 3, p. 23, 2007.

[26] D. J. Berndt, J. a. McCart, D. K. Finch, and S. L. Luther, “A Case Study of Data Quality in Text Mining Clinical Progress Notes,” ACM Trans. Manag. Inf. Syst., vol. 6, no. FEBRUARY, pp. 1–21, 2015.

[27] P. Lependu, S. V Iyer, C. Fairon, and N. H. Shah, “Annotation Analysis for Testing Drug Safety Signals using Unstructured Clinical Notes.,” J. Biomed. Semantics, vol. 3 Suppl 1, no. Suppl 1, p. S5, 2012.

[28] D. R. Murphy, A. Laxmisan, B. a Reis, E. J. Thomas, A. Esquivel, S. N. Forjuoh, R. Parikh, M. M. Khan, and H. Singh, “Electronic health record-based triggers to detect potential delays in cancer diagnosis.,” BMJ Qual. Saf., vol. 23, no. September 2015, pp. 8–16, 2014.

[29] S. L. West, W. Johnson, W. Visscher, M. Kluckman, Y. Qin, and A. Larsen, “The challenges of linking health insurer claims with electronic medical records.,” Health Informatics J., vol. 20, pp. 22–34, 2014.

[30] R. Cohen, M. Elhadad, and N. Elhadad, “Redundancy in electronic health record corpora: analysis, impact on text mining performance and mitigation strategies.,” BMC Bioinformatics, vol. 14, p. 10, 2013.

[31] S. Tanuja, D. Acharya, and K. R. Shailesh, “Comparison of different data mining techniques to predict hospital length of stay,” J. Pharm. Biomed. Sci., vol. 07, no. 07, 2011.

[32] T. S. Cole, J. Frankovich, S. Iyer, P. Lependu, A. Bauer-Mehren, and N. H. Shah, “Profiling risk factors for chronic uveitis in juvenile idiopathic arthritis: a new model for EHR-based research.,” Pediatr. Rheumatol. Online J., vol. 11, p. 45, 2013.

[33] N. Ramakrishnan, D. Hanauer, B. Keller, and B. Ramakrishnan, N., Hanauer, D., & Keller, “Mining electronic health records.,” Computer (Long. Beach. Calif)., vol. 43, no. October, pp. 77–81, 2010.

[34] P. Daumke, S. Schulz, M. L. Müller, W. Dzeyk, L. Prinzen, E. J. Pacheco, P. S. Cancian, P. Nohama, and K. Markó, “Subword-based semantic retrieval of clinical and bibliographic documents,” Methods Inf. Med., vol. 49, no. FEBRUARY, pp. 141–147, 2010.

[35] M. Kvist, M. Skeppstedt, S. Velupillai, and H. Dalianis, “Modeling human comprehension of Swedish medical records for intelligent access and summarization systems - Future vision , a physician ’ s perspective,” 9th Scand. Conf. Heal. Informatics, 2011.

[36] P. Coorevits, M. Sundgren, G. O. Klein, a. Bahr, B. Claerhout, C. Daniel, M. Dugas, D. Dupont, a. Schmidt, P. Singleton, G. De Moor, and D. Kalra, “Electronic health records: New opportunities for clinical research,” J. Intern. Med., vol. 274, no. September 2015, pp. 547–560, 2013.

[37] I. Fatima, M. Fahim, D. Guan, Y.-K. Lee, and S. Lee, “Socially interactive CDSS for u-life care,” Proc. 5th Int. Confernece ubiquitous Inf. Manag. Commun., no. September 2015, pp. 1–8, 2011.

[38] R. Harpaz, A. Callahan, S. Tamang, Y. Low, D. Odgers, S. Finlayson, K. Jung, P. LePendu, and N. H. Shah, “Text Mining for Adverse Drug Events: the Promise, Challenges, and State of the Art,” Drug Saf., no. September 2015, 2014.

[39] A Holzinger, R. Geierhofer, F. Mödritscher, and R. Tatzl, “Semantic Information in Medical Information Systems: Utilization of Text Mining Techniques to Analyze Medical Diagnoses,” J. Univers. Comput. Sci., vol. 14, no. 22, pp. 3781–3795, 2008.

[40] M. Andrade-Navarro and C. Perez-Iratxeta, “Text mining of biomedical literature: Doing well, but we could be doing better,” Methods, vol. 74, pp. 1–2, 2015.

[41] F. Rinaldi, K. Kaljurand, and R. Sætre, “Terminological resources for text mining over biomedical scientific literature,” Artif. Intell. Med., vol. 52, no. 2, pp. 107–114, 2011.

[42] R. a A. Seoud and M. S. Mabrouk, “TMT-HCC: A tool for text mining the biomedical literature for hepatocellular carcinoma (HCC) biomarkers identification,” Comput. Methods Programs Biomed., vol. 112, no. 3, pp. 640–648, 2013.

[43] I. Spasić, J. Livsey, J. a Keane, and G. Nenadić, “Text mining of cancer-related information: Review of

current status and future directions.," Int. J. Med. Inform., vol. 83, pp. 605–623, 2014.

[44] B. Xie, Q. Ding, H. Han, and D. Wu, "MiRCancer: A microRNA-cancer association database constructed by text mining on literature," Bioinformatics, vol. 29, pp. 638–644, 2013.

[45] Y.-C. Fang, H.-C. Huang, and H.-F. Juan, "MeInfoText: associated gene methylation and cancer information from text mining.," BMC Bioinformatics, vol. 9, p. 22, 2008.

[46] Who, "World Health Statistics 2009," p. 149, 2009.

[47] F. Zhu, P. Patumcharoenpol, C. Zhang, Y. Yang, J. Chan, A. Meechai, W. Vongsangnak, and B. Shen, "Biomedical text mining and its applications in cancer research," J. Biomed. Inform., vol. 46, no. 2, pp. 200–211, 2013.

[48] R. Maier, "Knowledge Management Systems: Information and Communication Technologies for Knowledge Management," Knowl. Manag., vol. 2, p. 720, 2007.

[49] A. Korhonen, I. Silins, L. Sun, and U. Stenius, "The first step in the development of Text Mining technology for Cancer Risk Assessment: identifying and organizing scientific evidence in risk assessment literature.," BMC Bioinformatics, vol. 10, p. 303, 2009.

[50] R. Jelier, M. J. Schuemie, A. Veldhoven, L. C. J. Dorssers, G. Jenster, and J. a Kors, "Anni 2.0: a multipurpose text-mining tool for the life sciences.," Genome Biol., vol. 9, no. 6, p. R96, 2008.

[51] N. Barrett, J. H. Weber-Jahnke, and V. Thai, "Engineering natural language processing solutions for structured information from clinical text: Extracting sentinel events from palliative care consult letters," Stud. Health Technol. Inform., vol. 192, pp. 594–598, 2013.

[52] H. C. Beck, Mass spectrometry in epigenetic research., vol. 593. 2010.

[53] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data To Big Impact," Mis Q., vol. 36, no. 4, pp. 1165–1188, 2012.

[54] D. Chiang, "Hierarchical Phrase-Based Translation ò Ã ¢," no. May 2006, 2007.

[55] W. W. M. Fleuren and W. Alkema, "Application of text mining in the biomedical domain," Methods, vol. 74, pp. 97–106, 2015.

[56] C.-H. Lee and S.-H. Wang, "An information fusion approach to integrate image annotation and text mining methods for geographic knowledge discovery," Expert Syst. Appl., vol. 39, no. 10, pp. 8954–8967, 2012.

[57] D. G. Rajpathak, "An ontology based text mining system for knowledge discovery from the diagnosis data in the automotive domain," Comput. Ind., vol. 64, no. 5, pp. 565–580, 2013.

[58] N. B. Ngwenya and S. Mills, "The use of weblogs within palliative care: A systematic literature review.," Health Informatics J., vol. 20, pp. 13–21, 2014.

[59] W. Der Yu and J. Y. Hsu, "Content-based text mining technique for retrieval of CAD documents," Autom. Constr., vol. 31, pp. 65–74, 2013.

[60] A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Ying Wah, and D. C. L. Ngo, "Text mining for market prediction: A systematic review," Expert Syst. Appl., vol. 41, no. 16, pp. 7653–7670, 2014.

[61] Jimeno Yepes and R. Berlanga, "Knowledge based word-concept model estimation and refinement for biomedical text mining," J. Biomed. Inform., vol. 53, pp. 300–307, 2014.

[62] Hotho, A. Nürnberger, and G. Paaß, "A Brief Survey of Text Mining," LDV Forum - Gld. J. Comput. Linguist. Lang. Technol., vol. 20, pp. 19–62, 2005.

[63] N. Ur-Rahman and J. a. Harding, "Textual data mining for industrial knowledge management and text classification: A business oriented approach," Expert Syst. Appl., vol. 39, no. 5, pp. 4729–4739, 2012.

[64] V. C. Pande and A. S. Khandelwal, "A Survey Of Different Text Mining Techniques," IBMRD's J. Manag. Res., vol. 3, no. 1, pp. 125–133, 2014.

[65] M. M. Mostafa, "More than words: Social networks' text mining for consumer brand sentiments," Expert Syst. Appl., vol. 40, no. 10, pp. 4241–4251, 2013.

[66] N. Ide and K. Suderman, "GrAF: A Graph-based Format for Linguistic Annotations," Proc. Linguist. Annot. Work., no. June, pp. 1–8, 2007.

[67] H. Shatkay, S. Brady, and A. Wong, "Text as data: Using text-based features for proteins representation and for computational prediction of their characteristics," Methods, vol. 74, pp. 54–64, 2015.

[68] D. Munková, M. Munk, and M. Vozár, "Data pre-processing evaluation for text mining: Transaction/sequence model," Procedia Comput. Sci., vol. 18, pp. 1198–1207, 2013.

[69] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," Procedia Comput.

Sci., vol. 17, pp. 26–32, 2013.

[70] U. Erra, S. Senatore, F. Minnella, and G. Caggianese, “Approximate TF-IDF based on topic extraction from massive message stream using the GPU,” Inf. Sci. (Ny)., vol. 292, pp. 143–161, 2015.

[71] Trstenjak, S. Mikac, and D. Donko, “KNN with TF-IDF based framework for text categorization,” Procedia Eng., vol. 69, pp. 1356–1364, 2014.

[72] M. Konopka, “Biomedical ontologies—A review,” Biocybern. Biomed. Eng., vol. 35, no. 2, pp. 75–86, 2015.

[73] T. R. Inniss, J. R. Lee, M. Light, M. a. Grassi, G. Thomas, and A. B. Williams, “Towards applying text mining and natural language processing for biomedical ontology acquisition,” Proc. 1st Int. Work. Text Min. Bioinforma. - TMBIO ’06, p. 7, 2006.

[74] Cedeño and M. Vargas-lombardo, “Framework Based on Ontologies for Palliative Care of Patients with Breast Cancer,” vol. 37, no. 3, pp. 49–57, 2015.