An Effective Study on Research-Based Information Extraction with Text Mining Methodologies for Health Risk Assessment

[1]T. Charan Singh, [2]M.Kiran Kumar, [3]T.Santosh, [4]S.Siva Kumar

[1, 2]Research Scholar, Sri satya sai University of technology and medical sciences, Sehore

[3,4]Assistant Professor, Department of Information Technology, Bhoj Reddy Engineering College for women, Hyderabad, Telangana

ABSTRACT

This paper displays a Research-Based information extraction engineering utilizing text mining methods to survey the human health risk of electromagnetic fields (EMFs) produced by remote sensor gadgets in the Internet of Things (IoT). The proposed design utilizes three text mining strategies to separate three sorts of information—reason articulation, research class, and wellspring of EMF exposure—from the Research-Based information to assist researchers with evaluating the human health risk of EMFs. For the reason proclamation, an agent sentence communicating the writers' aims and intentions were separated from the unique text of the articles through procedures of up-and-comer sentence choice, point dictionary creation, and weighting. For the research classification, the articles were characterized into three examination types—epidemiological, creature exploratory, and cell test—utilizing a weighting procedure based on the predefined include a dictionary of every classification. At last, all words speaking to recurrence groups remembered for the conceptual text of the articles were separated to recognize the wellspring of EMF exposure. The previously mentioned text mining methods were utilized to extricate the information from 320 logical articles and the exhibition of this design was demonstrated through master confirmation. The exploratory outcomes show that the proposed engineering can remove the ideal information to evaluate the human health risk of EMFs from the Research-Based information with high precision.

Keywords: *EMF exposure, information extraction, text mining, Research-Based information.*

I. INTRODUCTION

New research in biomedicine relies upon utilizing existing logical information – an assignment which bio-researchers are finding progressively troublesome. Given the twofold exponential development pace of biomedical writing over ongoing years [1], there is presently a squeezing need to create innovation that can make information in distributed writing increasingly open and helpful for researchers. Such innovation can be based on text mining. Drawing on strategies from regular language handling, information recovery, and information mining, text mining can consequently recover, extricate and find novel information even in colossal assortments of composed text. In spite of the fact that it can't yet supplant people in complex assignments, it can empower people to recognize and check the necessary Information in writing more proficiently and reveal important information darkened by the volume of accessible information.

As of late, biomedical text mining has expanded in fame. Methods have been created to help, for instance, the extraction of reports, databases, word references, ontologies, outlines, and explicit information (for example

associations among proteins and qualities, novel research theories) from pertinent writing [2–4]. The assessment of such methods has uncovered promising outcomes. Notwithstanding, a significant part of the assessment has been immediate in nature and has utilized pre-decided best quality levels. There is no broad acknowledgment of the need to move biomedical text mining research nearer to rehearse: to incorporate innovation to help genuine logical errands (for example the procedure of logical revelation) and to assess its convenience in the context of such undertakings [3,5].

Right now, a novel RESEARCH-BASED information extraction design identified with the human health risk assessment of EMF exposure is proposed. The proposed design uses text mining strategies to expand the productivity of information extraction and the objectivity of separated information to take care of the issues of existing database-based techniques that require unnecessary time and exertion and rely upon the decisions of explicit specialists. The proposed design separates three kinds of information—reason articulation, research class, and wellspring of EMF exposure—from the theoretical text of logical articles utilizing three text mining procedures. For the reason articulation, an agent sentence communicating the writers' goals and intentions are extricated from the dynamic text of the articles through procedures of competitor sentence determination, subject vocabulary creation, and weighting. For the research classification, the articles are ordered into three examination types—epidemiological, creature test, and cell test—utilizing a weighting procedure based on the predefined highlight vocabulary of every class. At last, all words speaking to recurrence groups (e.g., 'MHz' and 'GHz') remembered for the conceptual text of the articles are separated to recognize the wellspring of EMF exposure. The previously mentioned text mining strategies were utilized to remove the information from 100 logical articles and the presentation of this engineering was demonstrated through master confirmation. Our exploratory outcomes show that the proposed engineering can remove the ideal information to survey the human health risk of EMFs from the RESEARCH-BASED information with high precision.

II. LITERATURE REVIEW

Many research works added to the field of IE using different systems. The essential focal point of this researches was to decide how extraordinary text mining techniques can be used as the organized informational collections exist in the text record group. This part starts with characterizing the subject of the research, assessing past researches, and afterward significant systems are applied utilizing information extraction and text mining. So as to decide the subject of each research region and to build up a developmental and various leveled association between these themes, utilize the strategy for text mining. Subjects are exhibited through representation apparatuses. In addition, these apparatuses are utilized so as to show the association between these points and to offer intelligent capacities with the goal that clients can successfully discover cross-space themes and know the patterns of cross-area research.

Moloshnikov et al. built up a calculation for discovering reports on a specific subject

contingent upon a chose reference assortment of archives. Furthermore, the context-semantic chart for representation topics in list items was additionally evolved. The calculation relies upon the joining of a gathering of entropic, probabilistic and semantic engineers for mining of weighted catchphrases and set of words that clarify the predetermined point. Results showed that the normal exactness is 99% and the review is 84%. A one of a kind procedure was additionally made for causing charts based on the calculation, to can expel key expressions with loads. It offers the chance to show a plan of sub-themes in enormous arrangements of archives in minimal diagram structure.

So as to offer a reference for extra researches of different researchers, we examined the research status of text mining innovation when it was utilized in the biomedical field that covers 10 years. Biomedical text mining writing joined in SCI from 2004 to 2013 was recouped, separated and afterward inspected from the perspective of research establishments, yearly changes, research territories, nearby conveyance, diary sources, and catchphrases. A conspicuous increment in the measure of overall biomedical text mining writing is watched. Among this worldwide writing, a colossal rate is taken up by writing identified with named element acknowledgment, element connection extraction, text arrangement, text grouping, shortenings extraction, and co-event investigation. Concentrates did in the USA and UK are viewed as present in the essential position.

So as to remove between language groups through multilingual archives relying upon the Closed Concepts Mining and vector model, another factual methodology was proposed by. Formal Concept Analysis strategies are utilized for mining Closed Concepts from comparable corpora and later these Closed Concepts and vector models are used in the bunching and course of action of multilingual reports. A test assessment is done over a lot of French-English bilingual reports of CLEF's 2003. With a striking similarity score and so as to evacuate the bilingual classes of archives, results uncovered that the association between vector model and Formal Concept Analysis is extremely valuable.

Santosh recommended the diagram mining-based report content (for example text fields) abuse. That is, the inquiry created the diagram contingent upon the clients' necessities. This is a simple and compelling diagram mining strategy to extricate comparative examples through the reports and changed the inquiry chart into model charts which are used when the clients are absent. An astute answer for record information misuse has been made. This is portrayed by effortlessness, convenience, precision, simplicity of improvement, and adaptability. So as to comprehend chart models, it needn't bother with an immense assortment of record pictures. In addition, since model learning expends under 10 s for an information design for each class, all things considered, changes, corrections, and substitutions should be possible in the info designs. Information abuse normal execution is appeared to have 86.64% as Precision, and 90.80% as a Recall. In any case, the proposed system neglected to offer comprehensive and exact answers for the examples that have an immense assortment of fields in a crisscrosses course of action because of the question diagram multifaceted design.

Sirsat et al. proposed two systems for mining text through online sources. The primary procedure managed the information that is required to be demonstrated legitimately in the reports that should be mined. Text mining and IE are considered as the main compelling devices for playing out that method. The subsequent one worried about the records that hold genuine information in an unstructured configuration rather than nonfigurative information. IE can assist with changing the unstructured information displayed in the report corpus into an organized one. So as to find nonfigurative examples in the extricated information, information mining calculations and systems can be utilized.

Tune and Kim exhibited the main endeavor to apply text mining ways to deal with an enormous assortment of full-text articles for finding the information structure of the region. Rather than relying upon the reference information exhibited in Web of Science, PubMed Central full-text articles have been utilized for bibliometric assessment. Most importantly, this helped the production of text mining schedules so as to build up a handcrafted reference database following the full-text mining. Discoveries indicated that the vast majority of the records that were distributed in the bioinformatics region were not referred to by others. Furthermore, a consistent and direct ascent has been seen in the number of distributions across production years. Results uncovered that most of the recovered examinations were motivated by USA-based foundations followed by European establishments. Results announced that the significant essential focal point of the significant themes was on natural components. Be that as it may, as indicated by PageRank, the best 10 articles were profoundly worried about computational elements.

So as to encourage the exact extraction of text from PDF records of research articles that can be used in text mining applications, a "Format Aware PDF Text Extraction" (LA-PDF Text) framework was introduced by. Text squares are mined from PDF-organized fill-text research articles under this framework and afterward, the framework classifies them into coherent units relying upon decides that encapsulate specific areas. Just the textual substance of the research articles is centered around the LA-PDF Text framework. This framework fills in as a reason for new tests into more devel-oped extraction strategies managing multi-modular substances like pictures and diagrams. The framework experiences three stages: (1) Identifying touching text hinders with the assistance of spatial design handling so as to find squares of adjoining text,

(2) Categorization of text hinders into allegorical classes with the assistance of a standard-based strategy, and (3) Joining ordered text squares together by orchestrating them precisely which brings about the extraction of text from area astute gathered squares. An assessment of the exactness of the square revelation calculation utilized in stage 2 was performed. It was additionally demonstrated that the framework can distinguish and group them

into allegorical classes with Recall = 0.89%, Precision = 0.96%, and F = 0.91%. In addition, the precision of the text mined with the assistance of LA-PDF Text is contrasted with the text from an Open

Access subset of PubMed Central. This precision is then contrasted and the text that was mined utilizing the PDF2Text framework. These are the two habitually utilized strategies to extricate text from PDF.

Mooney and Bunescu portrayed two methods for utilizing normal language information extraction for text mining. To start with, general information can be mined legitimately from the text. A venture where an information base of 6580 human protein associations was removed by mining around 750,000 Medline abstracts in which reevaluated for instance of this strategy. Second, organized information can be mined through text reports or website pages. So as to discover designs in the mined information, conventional KDD techniques can be applied. The performed take a shot at the DiscoTEX framework and its application to Amazon book depictions, software engineering work postings, and resumes were considered for instance of this system. So as to find units and relations in text, research in IE continues making progressively proficient calculations. Important and noteworthy information can be mined successfully from the continually creating assortment of electronic archives and website pages by utilizing present-day approaches in human language innovation and computational semantics and connecting them with the cutting edge systems utilized in AI and traditional information mining methods. IE manages to determine a specific arrangement of applicable things through common language records.

So as to find themes that repeat in articles of text corpus, another technique TopCat (Topic Categories) was proposed by. IE was utilized by this system so as to find named elements in singular articles and to portray them as an assortment of things of an article. Along these lines, through acknowledgment of successive itemsets that normally happened with named elements, the issues in information mining or database context were examined. Affiliation rule information mining procedure is utilized by TopCat to find these incessant itemsets. By utilizing a hypergraph parting system, TopCat further groups the named substances which find an assortment of incessant itemsets with noteworthy overlie. So as to find archives with respect to the subject, the IR method was utilized.

III. RELATED WORK

Text Mining

The improvement in the fields of web, computerized libraries, specialized documentation, restorative information has made it simpler to get to a bigger measure of textual archives, which meet up to create helpful information assets [7]. Along these lines, it makes text mining (TM) or the information revelation from textual databases a difficult assignment inferable from fulfilling the guidelines of the profundity of normal language which is utilized by the vast majority of the accessible archives. The accessible textual information as databases and online sources [7–9] brings up an issue about who is answerable for keeping a beware of the information and breaking down it? Keeping in see the relating condition, it is unimaginable to expect to break down and viably remove the valuable information physically. There is a need to utilize software arrangements which may utilize automatic apparatuses for breaking down a lot of textual material, extricate applicable

information, dissect important information, and sort out significant information. Attributable to the expanding requests to acquire information from an enormous number of textual reports available on the web, text mining is increasing a noteworthy significance in research [10, 11]. For the one might say that information mining includes organized information, while text manages certain highlights and is relatively unstructured and normally require preprocessing. Besides, text mining is an interrelated field with Natural Language Processing (NLP). NLP is one of the hotly debated issues that are worried about the interrelation among the gigantic measure of unstructured accessible text [14], other than the examination and translation of person dialects [15, 16].

Information Extraction

An inception point for PCs to assess unstructured original copies is to utilize Information Extraction (IE). IE software

mining and looking through instruments [19]. IE manages to find and separating significant information from normal language texts [18]. It comprises of isolating fitting text parts, removing the offered information in such parts, and changing the information into the utilitarian structure. Fragmentary extraction from area-specific texts is presently conceivable; however, complete IE from the arbitrary text is as yet a proceeding with study target [20].

Separating Knowledge from Text

Under the vast majority of the conditions, just explicit information is acquired from the information separated from unstructured text rather than unique

most part, text mining and information mining are viewed as like each other, with a recognition that equivalent methods might be utilized in the two ideas to mine text [4, 12, 13], and [3]. Notwithstanding, both are diverse

perceives key expressions and connections remembered for the original copy. This is performed through finding the predefined plans in a text; this method is called design coordinating. Standard language text archives comprise of information that can't be used for mining. IE concurs with the documentation, picking fitting articles, and the relationship among them to make them increasingly accessible for included direction [17, 18]. As opposed to Information Retrieval, which manages how to perceive significant reports from an archive assortment, IE yields organized information arranged for post-handling, which is basic to different utilizations of Web

information. In such a case, it is required to utilize a text mining task alongside extra methods to mine information from the information close by [12, 22]. DiscoTEX (Discovery from Text EXtraction) is one of the significant methodologies utilized for text mining. It includes utilizing IE first to assemble organized information from unstructured text, trailed by utilizing conventional Knowledge Discovery from Database (KDD) apparatuses to dis-spread information from this information. This structure for text mining was introduced by [21]. Right now, the adapted IE framework is utilized to change over unstructured text into progressively organized information. This information is then exposed to mining to create

significant connections. For a situation that the information extricated from a corpus of records is as dynamic information rather than solid information, IE will in general fill in as the "finding information" from the text. Disclosure of information by separating information, for example, key-expressions or catchphrases extraction from the text might be utilized for other text mining undertakings, for example, arrangement, grouping, synopsis, and theme discovery.

Text Mining Methods and Techniques

Text mining is typically utilized to get brisk outcomes it has been oppressed research under various application regions. Based on individual territories of use, text mining can be sorted as text order, text grouping, affiliation rule extraction, and text perception. They are examined in the following sub-segments.

Text Clustering

Text bunching is based on the Cluster speculation which suggests that applicable reports must have a bigger number of likenesses with each other than the non-significant ones. The Clustering strategy is a trust-commendable system that is commonly utilized for breaking down bigger measures of information like information mining. It has been demonstrated that text bunching is one of the best apparatuses utilized for text topic investigation. In addition, it encourages the strategy for point investigation where named substances having a simultaneous event is gathered, trailed by exposing them to the grouping procedure so that visit thing is put in sets by applying the hyper diagram-based technique. Each arrangement of named substances is represented by a group that is identified with one of the continuous themes in the corpus. The procedure of point following inside unique text information has picked up the enthusiasm from the researchers who are chipping away at the subject of text grouping in the computerized field. Different strategies and calculations based on the unaided archive the executives are remembered for the procedure of record bunching. In the bunching procedure, the numbers, properties, and relationship of the assembled sets are at first obscure. The gathering of reports is performed by classifying them into a specific classification, for example, therapeutic, money related, and additionally lawful.

Affiliation Rule Extraction

An investigation by contended that the strategy for affiliation rule mining (ARM) is utilized to distinguish connections inside a bigger gathering of factors in a dataset. The ARM distinguishes the variable-esteem mixes which will, in general, happen frequently. The technique for ARM in information mining otherwise called information revelation in databases; that is like the connection examination that discovers the connections between two factors. Wong et al. [provided that the Association Rules for Text Mining are significantly worried to investigate the connections between different themes or authentic thoughts utilized for describing a corpus. They plan to find key affiliation rules comparative with a corpus so that the event of specific subjects in an article

may relate to the event of another point also.

K-Means Algorithms

The k-mean methodology partitions the informational collection into k groups, where each bunch is exposed to be spoken to by the mean of focuses; called the centroid. A two-advance dreary procedure is utilized for the utilization of the calculation: (1) Assigning each point to the closest centroid. (2) Evaluating the centroids for an as of late created gathering. The procedure is finished when the bunch centroid goes to a steady worth. The k-mean calculation has a broad application attributable to its immediate parallelization. Besides, the request for particular information doesn't influence the k-mean calculation which credits the numerical qualities to it. It is required to make reference to the greatest estimation of k toward the start of the procedure. The portrayal of the bunch is made by the k-medoid calculation that picks the article connecting the focal point of the group. However, the determination of the k objects is done haphazardly in the calculation. The chose objects to help to decide the separation. A bunch is framed based on the closest item to k, though different articles procure the situation of k recursively till the necessary nature of the group is accomplished.

Information Visualization

Information perception places extraordinary textual bases in a visual chain of importance or plan and offers perusing capacities just as general looking. This procedure offers improved and speedier exhaustive information, which helps us to mine enormous collection records. The administrators can recognize the hues, associations, and holes. The arrangement of reports can be shown as an organized format using an ordering or vector space model.

Word Cloud

Jayashankar and Sridaran characterized word mists or label mists as the visual portrayal of words for a specific composed substance organized according to its recurrence. Word cloud is among the most much of the time utilized technique to display text information in a graphical way; making it accommodating for breaking down different types of text information, for example, expositions and short answers or composed assessments to a review or poll. Word cloud will in general fill in as a fundamental stage for a top to bottom examination of certain text material. For instance, a word cloud helps with finding the pertinence between a given text and the necessary information. Regardless, the technique has certain downsides also. One of the significant downsides that is it doesn't think about the etymological information regarding the words and their particular connection to the given subject while giving an absolutely factual synopsis to the isolated words. Subsequently, in many frameworks, the word mists are often utilized in a factual way for condensing text, giving almost no or no methods for connecting the information. It is seen this could be one of the most impacting standards of perception for a large portion of the

investigation conditions. Therefore, right now, I have utilized the utilization of word mists as the focal strategy to text examination.

IV. PROPOSAL METHODOLOGY

Figure 1 outlines the proposed practical engineering for RESEARCH-BASED information extraction. The proposed design incorporates three distinct strategies for separating three sorts of information from the dynamic text of articles as information: (1) reason articulation extraction (PSE), (2) research class extraction (RCE), and (3) EMF exposure source extraction (EESE). Every strategy applies distinctive preprocessing capacities for conceptual text investigation as portrayed in detail in Sect. 3.

The PSE technique comprises of four procedures: (1) preprocessing, (2) competitor sentence choice, (3) subject vocabulary creation, and (4) the weighting of the reason proclamation. Subsequent to preprocessing, the applicant sentence determination and theme dictionary creation forms are executed autonomously. For the competitor sentence determination process, two kinds of predefined dictionaries are utilized: an intentional dictionary (PL) and an upsetting vocabulary (DL). The PL remembers the words for the title of each article and the words that are utilized for depicting the reasonable explanation. Then again, the DL incorporates the words that are utilized for depicting the test results. The aftereffects of the up-and-comer sentence determination process incorporate numerous up-and-comer sentences removed from the conceptual text. For the point dictionary creation process, there are 10 sets of words and their numeric loads are made through the inactive semantic investigation (LSA) subject displaying procedure. At long last, in the weighting procedure, the words and loads from the subject vocabulary are utilized to relegate a load to every applicant sentence, and the sentence with the biggest weight is distinguished as the reason articulation.

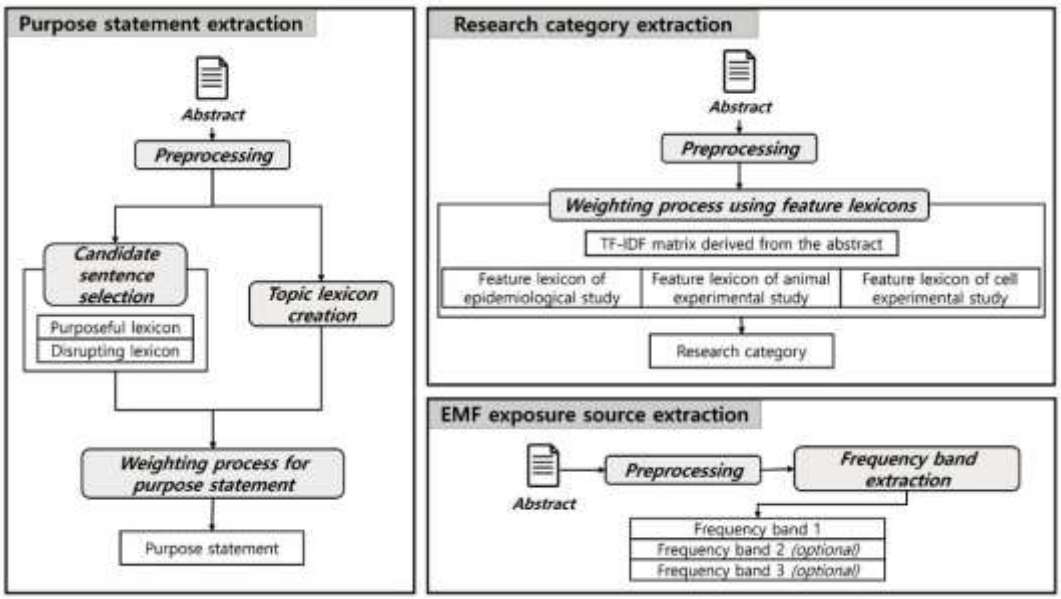

Fig. 1. Functional architecture of RESEARCH-BASED information extraction.

The RCE technique comprises of two procedures: (1) preprocessing and (2) weighting utilizing highlight vocabularies. A component vocabulary comprises of words identified with every one of the three research classes (i.e., epidemiological investigation, creature exploratory examination, and cell test study), which are characterized before the weighting procedure. In the initial step of the weighting procedure, the term recurrence opposite report recurrence (tf-idf) lattice is gotten from the dynamic text of an article. At that point, words that co-happen in theory text and every one of the three component dictionaries is separated. At last, the heaviness of every classification is determined by adding the tf-idf estimations of the words co-happening in theory text and each element dictionary. From the weighting procedure, the classification with the biggest weight is separated as the research classification of the article.

The EESE strategy comprises of two procedures: (1) preprocessing and (2) recurrence band extraction. Subsequent to preprocessing, all words speaking to recurrence groups (e.g., 'MHz' and 'GHz') referenced in theory text of articles are extricated, and at most three diverse recurrence groups are considered as the EMF exposure source.

CONCLUSION

This paper introduced a RESEARCH-BASED information extraction engineering for the human health risk assessment of EMFs created by remote sensor gadgets in the IoT condition utilizing text mining procedures. Diverse text mining systems were applied to separate three kinds of information: reason proclamation, research class, and wellspring of EMF exposure. Initially, the reason articulation was extricated utilizing three procedures: applicant sentence determination, subject dictionary creation, and weighting. Second, for the research class, a weighting procedure was applied to utilize highlight dictionaries. At long last, all words speaking to recurrence groups were separated from the unique text of articles to recognize the wellspring of EMF exposure. To assess the proposed text mining methods, a presentation assessment was directed by specialists utilizing 100 logical articles downloaded from the EMF-Portal database. The trial results demonstrated that the proposed engineering effectively extricated the reason articulation from 84 articles, the research classification from 69 articles, and the EMF exposure source from 92 articles.

REFERENCE

1.D. Poljak, Electromagnetic fields: Environmental exposure. In: Nriagu JO (ed.) Encyclopedia of Environmental Health, volume 2, pp. 259–268 Burlington: Elsevier, 2011.

2. D. Poljak, D. Cavka, H. Dodig, C. Peratta, A. Peratta, On the Use of Boundary Element Analysis in Bioelectromagnetics, Eng. Anal. Bound. Elem. (Special issue on Bioelectromagnetics) Vol. 49 (2014) pp. 2–14.

3. Navathe, S.B., Ramez, E.: Data warehousing and data mining. Fundam. Database Syst., 841– 872 (2000)

*4. Gupta, V., Lehal, G.S.: A survey of text mining techniques and applications. J. Emerg. Technol. Web Intell. **1**(1), 60–76 (2009)*

5. Gupta, S., Kaiser, G.E., Grimm, P., Chiang, M.F., Starren, J.: Automating content extraction of

*html documents. World Wide Web **8**(2), 179–224 (2005)*

*6. Hassani, H., Huang, X., Silva, E.S., Ghodsi, M.: A review of data mining applications in crime. Statistical Anal. Data Min.: ASA Data Sci. J. **9**(3), 139–154 (2016)*

*7. Feldman, R., Dagan, I.: Knowledge discovery in textual databases (KDT). KDD **95**, 112–117 (1995)*

8. Tan, A.H.: Text mining: The state of the art and the challenges. In: Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases, vol. 8, 65–70 (1999).

9. Hearst, M.A.: Untangling text data mining. In: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics, pp. 3–10. Association for Computational Linguistics (1999)

10. Hunter L, Cohen KB (2006) Biomedical language processing: What's beyond PubMed? Mol Cell 21: 589–594.

11. Ananiadou S, McNaught J (2006) Text Mining for Biology And Biomedicine. Norwood, MA: Artech House, Inc.

12. M. Kiran Kumar, S. Jessica Saritha (2015) An efficient approach to query reformulation in WEB SEARCH, Volume: 04 Issue: 06,IJRET.

13. Cohen KB, Yu H, Bourne PE, Hirschman L (2008) Translating biology: Text mining tools that work. In: Proceedings of the Pacific Symposium on Biocomputing;PSB-08.

14. Altman RB, Bergman CM, Blake J, Blaschke C, Cohen A, et al. (2008) Text mining for biology–the way forward: opinions from leading scientists. Genome Biology 9 Suppl 2.

15. Alex B, Grover C, Haddow B, Kabadjov M, Klein E, et al. (2008) Assisted curation: does text mining really help. In: Proceedings of the Pacific Symposium on Biocomputing;PSB-08.

16. Karamanis N, Seal R, Lewin I, McQuilton P, Vlachos A, et al. (2008) Natural language processing in aid of ybase curators. BMC Bioinformatics 9.

17. Guo Y, Korhonen A, Liakata M, Silins I, Ho¨gberg J, et al. (2011) A comparison and user-based evaluation of models of textual information structure in the context of cancer risk assessment. BMC Bioinformatics 12.

18. Jelier R, Schuemie M, Veldhoven A, Dorssers L, Jenster G, et al. (2008) Anni 2.0: a multipurpose text-mining tool for the life sciences. Genome Biology 9.

19. Frijters R, van Vugt M, Smeets R, van Schaik R, de Vlieg J, et al. (2010) Literature mining for the discovery of hidden connections between drugs, genes and diseases. PLoS Comput Biol 6: e1000943.

*20. C. H. Wei, H. Y. Kao, and Z. Lu: Nucleic Acids Res. **41** (2013) 518. https://doi.org/10.1093/nar/gkt441*

*21. R. Y. N. Lee, K. L. Howe, T. W. Harris, V. Arnaboldi, S. Cain, J. Chan, W. J. Chen, P. Davis, S. Gao, C. Grove, R.Kishore, H.-M. Muller, C. Nakamura, P. Nuin, M. Paulini, D. Raciti, F. Rodgers, M. Russell, G. Schindelman, M. A. Tuli, K. V. Auken, Q. Wang, G. Williams, A. Wright, K. Yook, M. Berriman, P. Kersey, T. Schedl, L. Stein, and P. W. Sternberg: Nucleic Acids Res. **46** (2017) 869. https://doi.org/10.1093/nar/gkx998.*

***22.** C. Simon, K. Davidsen, C. Hansen, E. Seymour, M. B. Barnkob, and L. R. Olsen: BMC Bioinf. **19** (2019) 57. https://doi.org/10.1186/s12859-019-2607-x*