A SYSTEMATIC REVIEW ON PATENT DOCUMENT ANALYSIS, PATENT DOCUMENT SUMMARIZATION TECHNIQUES FOR PATENT RETRIEVAL

[1] Avinash Thakur, [2]V.Hepsiba Mabel,

[1]Research Scholar, [2]Assistant Professor, [1,2,] Department of Computer science and engineering,
[1,2]School of Computing science & Engineering Vellore Institute of technology, Chennai India
I. *Email id: avinash.thakur2017@vitstudent.ac.in, hepsiba.mabel@vit.ac.in*

Received: 14 March 2020 Revised and Accepted: 8 July 2020

Abstract:

The primary objective of patent retrieval is the quest for applicable legal documents for a given application. Patent retrieval processes can differ considerably, reliant on the basis of different recovery tasks. Determining the patentability of the patent application is difficult, i.e. determining whether or not a similar invention has been written. Hence, using the patent document as the application would be beneficial, this could minimize labor cost and time consuming. However, locating the applicable previous art by means of the whole patent document as an enquiry is not a straightforward job; as such a request is composed of thousands of words which cannot reflect a concentrated need for information. In this survey Our approach is based on patent review, patent description to obtain successful research in patent retrieval, & further to patent document summarization

Keywords— Patent Analysis, text mining, comparative summarization, Extractive summarization

Introduction

Patent records contain important findings of the study. However, they are lengthy and dense in technical and legal jargon so interpreting them requires a lot of human effort. Advanced devices are in high demand for assisting patent research. When a consumer conducts a patent mining quest, a large number of items (patents, businesses, and inventors) will return. Digesting the appropriate amount of information is often costly to the consumer. The system will create a succinct and descriptive description for the returned objects automatically, so that the user can easily capture a global picture before clicking-and displaying each object. The following criteria should be met by a high-quality review: (1) cover the critical information in the returned items, (2) is also important for the query and (3) reduce redundancy in the created description. To solve the question of patent description we suggest a maximum method of coverage. The idea is to pick a set of representative sentences as the description for a question of the returned objects. The procedure consists of three concrete steps. First, when the user asks a question, he obtains applicable patents for the topic. Second, it extracts concepts in the patent document from every single paragraph. All the derived ideas form the space of information for the question, and each sentence also reflects the extracted concepts. Finally, it uses linear integer programming to find a series of sentences, whose thoughts cover the space of information to the limit. Extraction of definition in our system of full coverage the definition is the basic unit of information representation. A theory can be a noun, a phrase, an entity named (e.g., person or location), or even a subtree of a sentence parsed with syntax.

All definitions have a score of significance to the application. Concept scoring systems range from pure word frequency to advanced methods of machine learning. Candidate concepts are chosen using bigrams in our method, and are weighted using TF-IDF. In particular, all patents pre-processed by (1) sentence breaking, (2) part-of - speech marking, (3) stemming, and chunking of sentences. Then we compile all the extracted noun phrases from the patents (obtained by phrase chunking) and divide them further as candidate definitions into bigrams. The next move is to rate every single concept of candidate. We segment each patent document especially into five fields: title, description, claim, history, and other. Every field has a weight which represents its importance.

I. RELATED WORK

As there are extensive data available on the literature for text summarization and patent information retrieval, but it is not possible to present all content in front of you, we have selected some of those most related to our research, ie Summary generation/text summarization in the patent as well as other documents

Changjian Fang et al. [1] this paper recommends a novel co-ranking word-sentence concept called Co-Rank for automated extractive text summarization. Co-Rank incorporates the interaction between the word phrase and the graph-based rating pattern. The Co-Rank method can see as mutual support of terms and sentences on the basis that specific words will have prejudice weights.

Ferreira et al. [2] Authors described three separate formulations used to test the strategies. The authors selected the best five results from the various test sets, one of which would achieve as best a combination of four methods: Word Frequency, TF / IDF, Lexical Similarity, and Sentence Length. Also selected was the "Text Rank Value" technique for having reasonable results for 2 of the 3 test data sets. ROUGE's findings for quantitative assessment of the summaries were very similar to those obtained through the qualitative study. The equation TF / IDF is by far the most computationally costly of all tested processes. Methods The length of imprisonment and the amount of parole have the best balance between the execution period and the number of acceptable sentences.

Y. Zhang et al. [3] A text summarization method focused on Convolutionary Neural Networks is proposed in this paper for learning sentence features and jointly performing sentence ranking. It transforms the ranking function into a process of regression. Our proposed structure requires no prior knowledge, and can, therefore, extended with different writing styles to various document review tasks.

P. Sethi et al. [4] Authors were able to auto-summarize news articles & compare their summaries to determine which score parameters would help improve performance. In procedure, modified approaches used to exploit the idea that only news reports are dealing. They were just connecting nouns to lexical strings.

A. T. Al-Taani et al. [5] in this article, the authors analyzed and examined ATS extractive-based strategies built for Arabic texts. Such methods shall include 1. Statistics Approaches. 2. Graphic-based methods. 3. Approaches for Artificial Learning. 4. Methods of clustering 5. Meta – Heuristic Analysis Approaches.

P. P. Tardan et al. [6] it is noted that the writers used computational methodology techniques for this study, including sentence classification and sentence location functions. It describes the weak magnitude of the average outcome between the mathematical method and the qualitative research, plus the positive performance of the calculation of subjectivity. One of unimplemented elements of this research is elimination of sentences concerning trash. Junk sentences are sentences that have little connection with context of text.

J. N. Madhuri et al. [7] Recognizing the required subsections of the text in discussion. Researchers proposed extractive summarization of records using a statistically innovative method focused on sentence recognition phrases. The collected sentences are created as a summary text and translated into audio formats.

M. N. Uddin et al. [8] Study research has performed on document summarization, and the Bangla language summary Extraction has carried out. The critical drawback of the Bangla Interpretation is that it merely removes specific phrases from the text in discussion that is far more distinct from human summarizing. Another disadvantage is that often sentences that arrive early in the document have a higher probability of being in the description.

X. Sun et al. [9] Reinforcement ranking of various representation units within the scientific paper on the Semantic Connection Network will substantially boost the paper's extractive description. It not only sets out a representation method focused on semantic modeling, but also verifies Semantic Link Network's importance in reflecting and defining substance of text. The suggested solution has consistent consistency of both the academic articles and the description of a particular text in the short press.

A. P. Patil et al. [10] In fact, they are aimed at extracting a single English article, not exceeding 300 phrases in length, to a fraction of its original size while retaining cohesion, and then using a lexical database to abstract summary. To obscure analysis produced, software uses external tool WorldNet. WorldNet is a lexical database where terms are clustered across semitone relationships. Python's Natural Language Toolkit (NLTK) is used for accessing database via program. ROUGE used for assessment of summary.

In this paper, H. T. Le et al. [11] introduced a general text summary process consisting of two stages: the removal of sentences and the combination of sentences. The sentence lessening stage is based on debating rules for removing unnecessary words at the beginning of a sentence, and on syntactic constraints for completing the abbreviated sentence ending. The combination stage of the sentence is based on a word graph to show the relationship between words, clauses, and penalties from the input text. Expending a word graph produces novel phrases that chain various sentences of data. Experimental results indicate that our approach to AS problem solving is promising.

E. Naresh et al. [12]. In this survey paper, the writers presented a study of numerous documents using different strategies for categorizing and summarizing text. We can see that each of the methods has its advantages for specific applications. However, technologies like CNNs and RNNs were emerging to be very popular due to their benefits of back propagation and internal memory.

Ramesh et al. [13]. In this study, authors introduce the focus encoder-decoder for the abstractive description function with favorable outcomes, substantially outperforming state-of-the-art findings on 2 separate datasets. In abstract summing up, each of their proposed experimental models tackles a particular issue, resulting in more efficiency improvements. They are also offering a new multi-sentence overview dataset and establishing benchmark numbers. They expect to concentrate on that data as part of future research, and create more reliable models for multi-phase summaries.

Arpita Sahoo et al. [14], authors go through various methodologies and approaches for summarization of text documents. They say there's plenty of information available on the World Wide Web, and it's not possible to go through every article available to know document’s purpose and its utility. A review of these documents will therefore be more useful to the user in determining whether the accessible text is important or not, and it will be more efficient to derive the essence of each article.

Okamoto et al. [15], Authors suggested a method for the analysis of claim structure that uses a technique for extracting information. Applying machine-based computing knowledge retrieval methods instead of utilizing just syntactic sorting is beneficial in reducing the expense of extracting and arranging essential words from patent claims. They also developed a computational method for the simulation of patent ideas and patent assessment and the quest for related patents.

S. C. Gowri et al. [16], the authors of this article have analyzed a maximum inverted index as a reference for the procedure on string vectors for various methods of representation. Reflecting the text is easier than traditional methods by growing paper's pre-processing time and dimensionality, as well as making it simple to monitor whether each report is described in the same category. This research report's analysts concentrated on the encoding documents in string vectors by interpreting terms rather than numerical vectors.

Chintan Shah et al. [17]. Using supervised and unsupervised deep learning methods, they have proposed a hybrid solution for a single-document text description scheme. Authors use Self Organizing Maps (SOM) to know the function and chart it, retrain apps for description on Artificial Neural Network (ANN). Their methodology offers significant change, and extractive summarization methods are effective. Extractive summaries created by this approach sentencing scoring. But this method restricts their application to the single-document path.

Pattabhi R et al. This paper introduced a framework for the mining principle and abstract summarization of patent documents. It uses computational graphic formalism for representation. It obtains an average F-measure of 0.2198 ROUGE score, which is equivalent to the state of the art. The essential purpose of the current research was to determine how often capturing of a sentence structure and, therefore, of the text might help to produce descriptive summaries of the book. Use the conceptual graph helped catch the meaning and semantic and helped provide a theoretical description. [18]

Brügmann S et al. [19], Researchers provided specialized techniques for patent analysis and definition that form part of TOPAS workbench. Each of these approaches meets IP practitioners' needs and represents a commitment to state-of-the-art patent documents handling. Its usage improves the reliability and precision of almost all specific patent applications, including the study of a particular patent or patent collection, the landscaping of a patent collection, and the scan for patents. Individual patent research is particularly beneficial, for example, from TOPAS innovations for more effective navigation in a license or for recognizing personal individuals.

Tseng et al. [20]. Here outlines a sequence of mining techniques that correspond to theoretical method embraced by patent analysts and using them to train. Automating the whole process not only allows us to build final patent maps for the subject analysis, but it also promotes certain patent review activities, since each stage in this phase has its implementation.

Niemann, H. et al. [21]. The authors had presented the patent in this article, which they describe with time as the deployment of patent clusters. To this motive, a standardized method consisting of five steps was created, and six design decisions relevant to those steps were addressing. In conjunction with bicycle technology, the area of carbon fibers was selected to explain the patent structure process, contributing to insights into the development of the technology sector over time.

Brügmann S. et al., [22] introduce a patent documents information representation system and outline the design of strategies that enable the introduction of this framework into the patent production method. There are two styles of approaches questioned. First style strategies allow exposure to the substance of the patent documents presented in a text format – whether by the human user or the computer – by rephrasing and summarizing the documents and translating it into a standard semantics image.

Lopatecki L et al. [23]. “They introduce their research aimed at defining the hierarchy automatically within texts of full patent claims. Starting with a brief introduction to patent claims and common use cases for searching in claims, the findings of a preliminary background review of English claims from the European Patents Full text (EPFULL) database will be discussing. Also, point out some possibilities with which claim dependency is indicating in text”.[23]

Parapatics P, Dittenbach M et al. “To boost the passing efficiency for more automated processing, they propose an approach to break patent claims into several sections. In this study, they explore how the structure of the statements can be used to split them into multiple components and rearrange them to enhance the performance of natural language processing systems, such as parsers for dependence, and increase readability”.[24]

K. H. Law et al. [25] this paper outlines a technical architecture focused on experience to enable the collection of patents and relevant material through several disparate and uncoordinated sources of information in the US patent network.

Sheremetyeva S. et al. [26] propose an NLP approach to evaluate patent claims, mixing abstract grammar formalisms with data-intensive methods, thus improving the robustness of the study. Their analyzer's performance is a superficial inter-lingual representation capturing the form and meaning of an argument document. The technique may be used in any application relating to patents, such as machine translation, enhanced readability of patent claims, collection of knowledge, compilation, description, creation, etc.

Shinmori A, Okumura M (2004), et al. [27] proposed a framework for representing the patent claims structure and a method for the automatic analysis. A tool for clarifying terminology in patent claims and identifying informative parts of the detailed description section of patent requirements is examining. Authors wholeheartedly believe with both approaches that they will find patent statements easy to grasp.

Shinmori A et al. [28]. They are researching NLP innovations to allow patent statements accessible faster. For this to happen, the definition of the comments must be defined descriptively. Authors notice that some popular phrases are used in claim definitions, and they can be used as clues to analyze the semantic framework of patent claims. By utilizing these cue words, they suggest a framework for evaluating the conceptual structure of patent claims and disclosing the assessment outcome.

Verberne S et al. [29]. Within this article, they intend to validate and measure the problems found in the literature about the handling of patent claims. Authors concentrate on the following three issues, which are crucial to patent claim review in the context of the number of references in the articles on patent research and patent retrieval: (1) The duration of sentences is much longer than with the normal usage of language; (2) several modern words are used in patent claims which are complicated to comprehend; (3) the syntactic form of patent claims is complex.

Wanner, L et al. [30]. They present ongoing work on an advanced patent-processing service PAT Expert in this article. PAT Expert's fundamental premise is that to satisfy the demands of consumers of patent management systems, the quality of Patent information must be recourse. They implement a material description scheme for the design of patent documentation and sketch techniques, which facilitates the incorporation of this system into the patent processing cycle.

Yang SY et al. [31], "This paper describes strategies for extracting conceptual diagrams from a patent claim utilizing syntactic knowledge (POS, and tree of dependence) and semantic details (ontology of the background). Because of the proliferation of specialized domain words and the lengthy sentences in patent applications, it is complicated to explicitly submit an NLP Parser to decipher the simple texts in the patent argument. This paper uses strategies such as finite state machines, part-of-speech marks, definition maps, ontology of domain, and the tree of dependence to transform a patent assertion into an explicitly specified logical model".

Avinash et. al [33] This paper gives a comprehensive review of all domains of Patent documents analysis.

II. **METHODOLOGY**

In this paper, We discussed three incorporated aspects of patent mining methods, comprising enquiry generation to enhance patent retrieval efficiency, patent description to understand both the commonality and the distinction between patent pairs, and key patent mining from a large volume of patent documents. [3]

Patent Retrieval: Patent recovery is a subdomain of knowledge recovery, in which patents are the basic elements to be sought. Patent scanning is radically different from searching general web records because of the features of patents and special conditions of patent retrieval. [4]

A. Automatic extraction of keywords: Grounded on the exploration of patent official papers, our outline is capable of routinely extracting significant but unique keywords from a specific patent document which integrates special typescripts of patent documents. [3]

B. Expansion of specific keywords: Based on the familiar base and tenure thesaurus, Our framework may expand a list of keywords relevant to a given query sentence. The expansion is achieved by integrating material proximity with topic pertinence.

C. Filtering of results with topic: "Our system is able to retrieve appropriate patent documents based on the extended search query. The results are obtained by identifying all possible relevant patent documents and then filtering them within the relevant topics."[3, 6]

Single Patent Document Analysis: "Patents are among the main invention paper carriers. In-depth review of patent records allows essential technological details and relationships to be uncovered, which can provide useful knowledge for developing R&D strategies."[7,8]

Multiple Patent Document Analysis: Reviewing large volumes of patent documents would enable us to understand technological progress effectively, to understand modern developments and to capture the arrival of new technology. However, reviewing several patent documents is a non-trivial task, since there can be a lot of underlying relationships between various documents, which requires a lot of human efforts.

III. SUMMURY OF RESEARCH

As per above discussion, all methods mention in existing systems are work for text classification and text categorization. Our problem statement is to generate a summary from multiple patents documents from different domains at same time. By considering need for research, we will mainly be focused on patent documents and generate comparative summary from multiple patent reports; it may broadly divide in component-wise summary of each story as well as whole document summary. This summary generation performance evaluated by using ROUGH matrix in existing system, and gives satisfied result [2],[10],[18], so will also carry forward same method for performance measure. As per some Author's discussion [15], [23], [27], [28][29]. Claim is one of most crucial parts of patent, and theirs is lack of research so that we will be focusing on Claim. Below are some of the well known techniques

Sr. No.	Method	Explanation	Formula
1	Part of Speech (POS) Tagging	Part of Speech Tagging is how to group or describe texts by speech form, such as nouns, verbs, adverbs, adjectives, etc. Numerous algorithms such as the Secret Markov model and the Viterb form of tagging POS will apply	
2	Stop Word Filtering	Stop words are those removed prior to or post to text (corpus) has been processed; it is entirely non-objective depending on case. A, an, in, for instance may be viewed as stop terms and extracted from basic text.	
3	Stemming	Stemming is a mechanism by which inflected (or often derived) verbs are reduced to their word stalk, base or origin. For instance, removing verbs from -editing, using singular rather than a plural noun etc. Numerous NLP instruments use stemmer algorithms.	

4	Feature Calculation	This attributes are usually used in sentence ranking to measure score of Sentence S, showing degree, whether it belongs to description or not.	
5	Title Similarity	A sentence gets a strong notch if it has full amount of identical terms in text. Number of words will calculate subject similarities in subject conviction and overall number of words in description	f1= $\frac{S \cap T}{t}$ In formula; S = word set in sentence T = word set in title S = Related terms in sentence and document title
6	Sentence Position	In that feature its location in the text decides a sentence. If it is first five sentences of an article, the position of the sentence in a book gives the phrases meaning. First they take into consideration five essay sentences.	F2 = 5/5 for 1st, 4/5 for 2nd, 3/5 for 3rd,2/5 for 4th, 1/5 for 5th,0/5 for other sentences. On the other hand, when determining the meaning of the sentence location a clear and straight calculation is used. F2 = 1, if the phrase begins in the text, f2 = 0, If the phrase starts in the middle of the paragraphs in the book, f2=1, if the phrase ends in the text.

7	Term Weight (Term frequency)	"The cumulative weight for word is determined for paper by calculating tf and idf. Here idf corresponds to inverse frequency of documents, which literally says whether word is normal or uncommon among all documents".[12]	wi = tfi.idfi = tfi .log $\frac{N}{ni}$ Tfi = phrase frequency I in text N = cumulative number of sentences ni = number of sentences I appears in. You can measure the function as follows. f3 = $\sum Wi(S)\ k\ i$=1 Max($\sum Wi(SiN)\ k\ i$=1) k=numberof words sentence
8	Sentence Length	"This feature is ideal for removing unnecessarily short sentences such as datelines or author names. Not needed to add brief sentences to the list".[13]	f4 = $Number\ of\ Words\ in\ Sentence$ $Number\ of\ words\ occuring\ in\ longest\ sentence$
9	Thematic Word	It function refers to domain-specific words, which are supposed to appear as related subjects frequently in a document. The score for this feature is designed in phrase to complete definition of thematic terms in speech as an average of the sum of thematic terms.	f5 = $Number\ of\ th\,ematic\ word\ in\ S$ $Maximum\ (Number\ of\ Th\,ematic\ Word)$
10	Proper Nouns	"Generally speaking, a sentence that contains proper nouns is important, and most likely includes them in the document description. The score for	f6 =

		this attribute is to measure significant sentence, which requires a maximum number of correct nouns, as the percentage of correct number nouns that appear in a paragraph over the length of a sentence".[14]	$Number\ of\ Proper\ nouns\ in\ S$ $Sentence\ Length$
11	Sentence to Sentence Similarity	The formula for cosine similarity provides relation between statement S and each other. As vectors are defined word weight wi and wj of term t to n term in sentence Si and Sj.	Similarity(Si, Sj) = $\sum WitWjt\ N\ t$=1 √$\sum Wit$ 2 $n\ t$=1 √$\sum Wjt\ 2n\ t$=1 f7=$\sum Similarity(Si,Sj)$ $Maximum(\sum Similarity(Si,Sj))$
12	Numerical Data	"Generally speaking, a sentence that contains numerical details is important and most usually incorporates it in a document description. The score is calculated as a ratio of the sum of numerical data in sentence to sentence duration".[7,8]	f8 = $Number\ of\ Numerical\ Data$ in S $Sentence\ Length$
13	Bayesian Classifier	"Instead of a training collection of documents with hand-selected document extracts, construct a classification function that estimates the probability of including a given sentence in an extract. Then new extracts can be generated by ranking penalties according to this probability and selecting the top scoring ones listed by a number of users."[4,5]	For. sentence we quantify the likelihood it will be included in a description of features S provided k from f j; j = 1…k, which can expressed using Bayes' rule as follows: P (s $\epsilon\ S \mid f1, f2, \ldots, fk$) = $P\ (f1, f2, fk \mid s\ \epsilon\ S)\ P\ (s\epsilon S)\ \ (f1,2,\ldots,fk)$ Assuming statistical independence of features: P(s $\epsilon\ S \mid f1$, 2,.. fk) = $\prod\ P(fj \mid s\ \epsilon\ S)\ P(s\ \epsilon\ S)\ k$ j=1 $\prod$ /$P(fj)$

IV.CONCLUSION

Before we perform this experiment our results support some of our speculations. Firstly, text description techniques aid in the study and arrangement of patents, either automatically or manually. It's because patents

appear to be lengthy, terminology-rich, and full of specifics, which can confuse human and computer subject research alike. Second, machine-derived summaries can be as useful as or even better than those manually collected. This can be seen by re-
Inspection of individual words and higher coefficients. Third, in patent research, patent claim is proving very significant.

Our results come to a conclusion, where we can use these ANNs to identify patents and see if they are as unfair as expected by the correlation coefficient. To find parallels between the two sentences apply the intersection matrix. Also, in the actual function of classification, the section excerpts should be contrasted with the full texts and other segments to demonstrate their advantages. To check those results, our future research will perform further experiments. Using multiple paper approaches to extractive summary generation for patents belonging to the same or different categories to enforce a patent summary process.

REFERENCES

[1] Changjian Fang, Dejun Mu, Zhenghong Deng, Zhiang Wu, Word-Sentence Co-Ranking for Automatic Extractive Text Summarization, Expert Systems With Applications (2016),

[2] Ferreira, Rafael & Cabral, Luciano & Lins, Rafael & Silva, Gabriel & Freitas, Fred &
Cavalcanti, George & Lima, Rinaldo & Simske, Steven & Favaro, Luciano. (2013). Assessing sentence scoring techniques for extractive text summarization. Expert Systems with Applications. 40. 5755–5764. 10.1016/j.eswa.2013.04.023.

[3] Y. Zhang, M. J. Er and M. Pratama, "Extractive document summarization based on convolutional neural networks," IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, 2016, pp. 918-922.

[4] P. Sethi, S. Sonawane, S. Khanwalker and R. B. Keskar, "Automatic text summarization of news articles," 2017 International Conference on Big Data, IoT and Data Science (BID), Pune, 2017, pp. 23-29.

[5] A. T. Al-Taani, "Automatic text summarization approaches," 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), Dubai, 2017, pp. 93-94.

[6] P. P. Tardan, A. Erwin, K. I. Eng and W. Muliady, "Automatic text summarization based on semantic analysis approach for documents in Indonesian language," 2013 International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, 2013, pp. 47-52.

[7] J. N. Madhuri and R. Ganesh Kumar, "Extractive Text Summarization Using Sentence Ranking," 2019 International Conference on Data Science and Communication (IconDSC), Bangalore, India, 2019, pp. 1-3.

[8] M. N. Uddin and S. A. Khan, "A study on text summarization techniques and implement few of them for Bangla language," 2007 10th international conference on computer and information technology, Dhaka, 2007, pp. 1-4.

[9] X. Sun and H. Zhuge, "Summarization of Scientific Paper Through Reinforcement Ranking on Semantic Link Network," in IEEE Access, vol. 6, pp. 40611-40625, 2018.
[10] A. P. Patil, S. Dalmia, S. A. A. Ansari, T. Aul and V. Bhatnagar, "Automatic text summarizer," 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), New Delhi, 2014, pp. 1530-1534.

[11] H. T. Le and T. M. Le, "An approach to abstractive text summarization," 2013 International Conference on Soft Computing and Pattern Recognition (SoCPaR), Hanoi, 2013, pp. 371-376.

[12] E. Naresh and Kumar B P, Vijaya and V S, Pruthvi and K, Anusha and V, Akshatha, Survey on Classification and Summarization of Documents (May 17, 2019). International Journal of Research in Advent Technology, Vol.7, No.6S, June 2019

[13] Ramesh Nallapati, Zhou, Bowen, dos Santos, Cicero Xiang, Bing, "Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond", Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL), pages 280–290, Berlin, Germany, August 7-12, 2016.

[14] Arpita Sahoo, Dr.Ajit Kumar Nayak, "Review Paper on Extractive Text Summarization", International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 5, Issue 4, April 2018

[15] Okamoto, Masayuki & Shan, Zifei & Orihara, Ryohei. (2017). "Applying Information Extraction for Patent Structure Analysis". SIGIR'17, August 7-11, 2017, Shinjuku, Tokyo, Japan10.1145/3077136.3080698.

[16] S. C. Gowri et al, "A Study on Information Retrieval and Extraction for Text Data Words using Data Mining Classifier" International Journal of Computer Science and Mobile Computing, Vol.4 Issue.10, October-2015, pg. 121-126

[17] C. Shah and A. Jivani, "A Hybrid Approach of Text Summarization Using Latent Semantic Analysis and Deep Learning," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 2039-2044.

[18] Pattabhi R K Rao and Sobha Lalitha Devi, "Patent document summarization using conceptual graphs" International Journal on Natural Language Computing (IJNLC) Vol. 6, No.3, June 2017DOI: 10.5121/ijnlc.2017.6302 15

[19] Brügmann S, et. al., "Towards content-oriented patent document processing: Intelligent patent analysis and summarization", World Patent Information, ELSEVIER, (2014), http://dx.doi.org/10.1016/j.wpi.2014.10.003

[20] Tseng, Yuen-Hsien & Lin, Chi-Jen & Lin, Yu-I. (2007). Text mining techniques for patent analysis. Information Processing & Management, ELSEVIER 43. 1216-1247. 10.1016/j.ipm.2006.11.011.

[21] Niemann, H., et al., Use of a new patent text-mining and visualization method for identifying patenting patterns over time: Concept, method and test ap..., journal: Technological Forecasting in Social Change (2016), Elsevier, http://dx.doi.org/10.1016/j.techfore.2016.10.004

[22] Brügmann S, Bouayad-Agha N, Burga A, et al (2015) Towards content-oriented patent document processing: Intelligent patent analysis and summarization. Patent Claim Structure Recognition 15 World Patent Information 40:30–42.

[23] Lopatecki L (2008) Comparative analysis of German and American patent claim text structures as a basis for the evaluation of patent scope. Master's thesis, University of Hildesheim, Germany

[24] Parapatics P, Dittenbach M (2011) Patent claim decomposition for improved information extraction. In: Lupu M, Mayer K, Tait J, Trippe AJ (eds) Current Challenges in Patent Information Retrieval, The Information Retrieval Series, vol 29, Springer Berlin Heidelberg, pp 197–216

[25] S. Taduri, G. T. Lau, K. H. Law, and J. P. Kesan. A patent system ontology for facilitating retrieval of patent-related information. In Proceedings of the 6th International Conference on Theory and Practice of ElectronicGovernance (pp. 146-157). ACM, 2012

[26] Sheremetyeva S (2003) Natural language analysis of patent claims. In: Proceedings of the ACL-2003 workshop on patent corpus processing, Association for Computational Linguistics, Morristown, NJ, USA, pp 66–73

[27] Shinmori A, Okumura M, Marukawa Y, Iwayama M (2003a) Patent claim processing for readability: Structure analysis and term explanation. In: Proceedings of the ACL-2003 workshop on patent corpus processing, Association for Computational Linguistics, Morristown, NJ, USA, pp 56–65

[28] Shinmori A, Okumura M, Marukawa Y, Iwayama M (2003b) Rhetorical structure analysis of Japanese patent claims using cue phrases. In: Proceedings of the third NTCIR workshop on research in information retrieval, automatic text summarization, and question answering, National Institute of Informatics, Japan

[30] Wanner L, Baeza-Yates R, Brügmann S, Codina J, Diallo B, Escorsa E, Giereth M, Kompatsiaris Y, Papadopoulos S, Pianta E, et al. (2008) Towards contentoriented patent document processing. World Patent Information 30(1):21–33

[31] Yang SY, Soo VW (2012) Extract conceptual graphs from plain texts in patent claims. Engineering Applications of Artificial Intelligence 25(4):874–887

[32] Okamoto, Masayuki & Shan, Zifei & Orihara, Ryohei. (2017). Applying Information Extraction for Patent Structure Analysis. 10.1145/3077136.3080698.

[33] Avinash & Hepsiba A Critical Literature Review on Patent document mining, analysis and evaluation.https://www.jardcs.org/backissues/archives-special.php?year=2018&issue=13-26