IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657
Volume 7, Issue 3, 2019

Comparison of Text Mining Tools, Techniques and Issues

Zaheeruddin Ahmed , Dr.B.L.Raju, [2], Dr.Jatin Sharma, [3]

Abstract: Now-a-days, online reviews in the e-commerce website are increasingly written by the consumers of the product. More than 80 percent of the data present in them is unstructured. These reviews have become an important source of information for the new customers to research about these products online. The curious customer research often leads to decision making towards purchasing the product based on online reviews. In contrast to structured data, unstructured data such as texts, speech, videos and pictures do not come with a data model that enables a computer to use them directly. Nowadays, computers can interpret the knowledge encoded in unstructured data using methods from text analytics, image recognition and speech recognition. Therefore, unstructured data are used increasingly in decision-making processes. But although decisions are commonly based on unstructured data, data quality assessment methods for unstructured data are lacking. While databases store only structured data, most of the data is unstructured like text documents, web pages, emails etc. Text mining is what is required if useful information needs to be extracted from tons of text. But where to begin, what are the popular tools, which techniques are used, what are the features. Beginning is always the toughest, so in this work tries to explore the tools available for text mining to help new researchers and practitioners in the field of text mining.

Keywords: Text Mining, Text Analytics, Text Mining Tools, Techniques for text mining, Data Analysis, quality of unstructured data

I. INTRODUCTION

There is tons of textual information, but text is unstructured data and a means is required to extract useful information from unstructured data. Text Mining is used for the purpose of extracting useful hidden information and patterns in text [1]. It is very clear that we have text everywhere and not all data can be stored in databases. Sometimes we need to analyze data that is not stored in corporate databases, for instance we might need to extract useful information from a company's website, emails or reports. Obviously we cannot use the same techniques on text as we used on structured data stored in databases. Since unstructured data is complex, more powerful techniques are required to extract information from it. So we need text mining tools that can analyze unstructured data i.e.

Professor & Principal
Charan Singh University

ISSN 2347–3657
Volume 7, Issue 3, 2019

Text. In other words simple data mining tools are not adequate to handle text, we require specialized tools with more powerful algorithms that can process and analyze text.

Text mining is used in every field be it for business intelligence, social media analysis, sentiment analysis, biomedical analysis, software process analysis and even for security analysis. For instance Dekhtyar et al. [2], describe the use of text mining to derive information from software engineering text and the benefits of augmenting software repositories with natural language text. Sateli et al. [3] describe how text mining can be used for improving the quality of software specification without requiring any prior knowledge of Natural Language Processing (NLP). Also Malhotra et al. [4] apply text mining and Support Vector Machine (SVM) for predicting the severity of software bug reports. Sharma et al. [5] also use text mining in conjunction with Naive Baiyes Multinomia (NBM) and K-Nearest Neighbour (KNN) classifiers to predict severity of bugs.

Jurek et al. [6] discuss the use of lexicon-based sentiment analysis for social media analytics. In [7] Eom and Zhang, develop a tool (PubMiner) that helps in extracting biomedical
information from literature abstracts by employing machine learning and data mining techniques. Thus text mining is used in a variety of fields and one needs to knowledge about the tools that are available to assist in text mining so that gathering information becomes easier and faster.
In this paper we try to explore some of the popular and easily available tools for text mining, what techniques employed and what features are provided by each tool. Section 2 discusses
the different types of text mining tools. Section 3 identifies the popular text mining techniques used in these tools. Section 4 identifies the features of popular text mining tools. Section 5
provides a conclusion.

II. TYPES OF TEXT MINING TOOLS

We used the following search string to determine popular text mining tools [(Text) AND (Mining OR Analytics) AND (Tool)]. From the search results we identified 55 popular text mining tools and studied their features. Table 1. lists these tools along with their features and techniques employed by them. In the following sections we analyse the popular techniques and features of text mining tools. Text Mining Tools can be classified into three categories as shown in Figure1.

1. Proprietary Text Mining Tools: These tools are commercial text mining tools owned by a company. To use these tools purchase is required. Although demo/trial versions are available free of cost but have limited functionality. 39 out of these 55 tools are proprietary tools.

2. Open Source Text Mining Tools: These tools are available free of cost and also there source code and one can even contribute in their development.13 out of these 55 text mining tools are open source. Online Text Mining Tools: These tools can be run from the website itself. Only a web browser is required. These tools are generally simple and provide limited functionality.

ISSN 2347–3657
Volume 7, Issue 3, 2019

Three out of these 55 text mining tools are online web based tools.

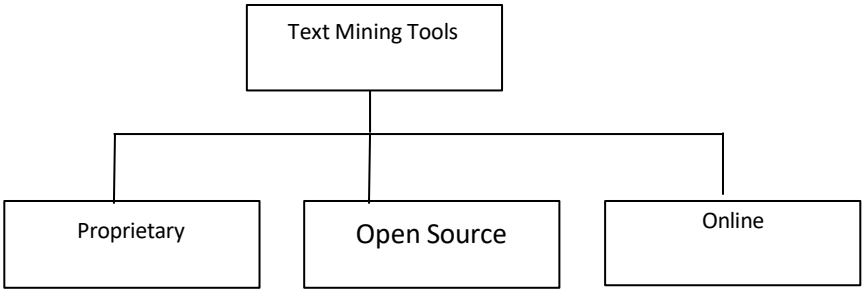

Fig. 1. Types of Text Mining Tools

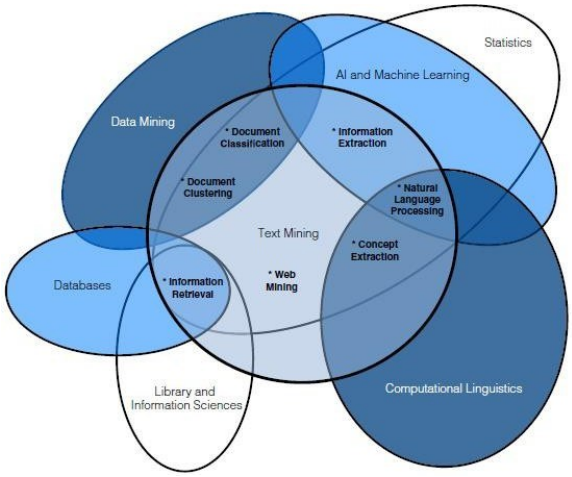

Fig. 1. Venn diagram of text mining interaction with other fields [10]

TABLE 1: TEXT MINING TOOLS

Tool	Type	Techniques supported	Features/Uses	Website	Additional Remarks
Ranks.nl	Online	Keyword analysis	Page Analysis, Article Analysis, Multi page analysis	Http://www.ranks.nl/	Website has been put together using Perl, Mysql, Javascript and HTML. Input Supported: Text/URL
Text Sentiment Visualizer	Online	Deep neural networks and D3.js.	Sentiment analysis	Http://sentiment.lucasestevam.com/	Input Supported: Text/URL
Textalyser	Online	Text Analysis, Keyword Analysis	Text analysis	Http://textalyser.net/	Input Supported: Text/URL
Alceste	Proprietary	Hierarchical descending classification, ascending classification, thematic classification	Textual data analysis, Multilingual analysis, temporal analysis	Http://www.image-zafar.com/Logicieluk.html	OS required-Win XP, VISTA, 7, 8 et Mac OS-X
Anderson Analytics odintext	Proprietary	Advanced statistics and other machine learning techniques	Text analytics	Http://odintext.com/#	

Ascribe	Proprietary	Hybrid technology approach, natural language processing, machine learning and semi-automated coding tools	Text analytics	Http://goascribe.com/	
Basis Technology Rosette	Proprietary	Linguistic analysis, statistical modeling, and machine learning	Text Analytics, multilingual text analytics	Http://www.rosette.com/	Integrated with curl, Python, PHP, JAVA, C#, nodejs, Ruby
Buzzlogix text analysis api	Proprietary	Semantic Text Analysis using natural language processing	Text analysis, sentiment analysis, classification, keyword analysis	Https://www.buzzlogix.com/text-analysis/	
Clarabridge	Proprietary	Linguistic and statistical algorithms, Natural Language Processing (NLP).	Text analytics	Http://www.clarabridge.com/text-analytics/	
Clustify	Proprietary	Classification	Categorization of documents	Http://www.cluster-text.com/	
Dataladder productmatch	Proprietary	Machine learning	Data cleansing, classification	Http://dataladder.com/products/productmatch/	
Discovertext	Proprietary	Cloud-based text analytics, Active Learning machine classification engine	Text analytics	Http://discovertext.com/	
Tool	Type	Techniques supported	Features/Uses	Website	Additional Remarks
Dtsearch	Proprietary	Advanced data classification	Text search	Http://www.dtsearch. com/	
Eaagle text mining software	Proprietary	Knowledge discovery algorithms	Text analysis	Http://wp.eaagle.com /?Page_id=16	
Expert System cogito tool	Proprietary	Artificial intelligence algorithms, semantic analysis, natural language processing	Knowledge management, semantic comprehension, decision making	Http://www.expertsys tem.com/	
IBM infosphere Warehouse Enterprise Edition	Proprietary	OLAP, data mining	Advanced analytics, data mining and text analytics.	Http://www-03.ibm.com/software/ products/en/infospher e-warehouse-pack- customer-insight	OS Required: Windows XP, or Windows 7 (32-bit) machine
IBM SPSS Predictive Analytics	Proprietary	Predictive analytic/modelling, Artificial intelligence algorithms	Data mining and text mining,,statistical analysis	Http://www.ibm.com /analytics/us/en/techn ology/spss/	

Intellexer	Proprietary	Natural language processing	Text analysis and information management, comparison and categorization of documents	Http://www.intellexer .com/knowledge_ma nagement.html	
Ureveal	Proprietary	Patented text anlytics methods including unbiased learning, olap	Data analysis, text analytics	Http://www.ureveal.c om/	
Kbsportal	Proprietary	Natural Language Processing as a SAAS Web Service	Text analytics, document categorization,	Http://kbsportal.com/	
KNIME	Open source	Data Blending & Transformation Math & Statistical Functions Advanced Predictive Algorithms, including Weka support	Text analysis, data I/O, preprocessing and cleansing, modeling, analysis and data mining	Http://www.knime.or g/knime	Integrated capabilities for Python, R, SQL, Java, JSON, and XML Prerequisites:Eclipse platform
Langsoft	Proprietary	Artificial intelligence and natural language processing	Question answering, logical inference, content recognition and text attribution	Http://www.langsoft. ch/refer.htm	
Lexalytics	Proprietary	Machine-learning techniques and expert- tuned industry rules, natural language processing, advanced algorithms	Sentiment Analysis, Categorization & Named Entity Extraction	Https://www.lexalyti cs.com/	
Lextek Profiling Engine	Proprietary	Information retrieval and natural language processing	Document and Knowledge Management	Http://www.lextek.co m/	
Linguamatics I2E	Proprietary	Natural Language Processing	Text mining	Http://www.linguamatics.com/products - services/about-i2e	I2E WSAPI is available to use from any programming language (Java, javascript, C++, C#, Python, etc.)
Loop AI Labs	Proprietary	Machine learning, artificial intelligence	Text processing and analysis	Http://www.loop.ai/	
Meaningcloud	Proprietary	Feature-level sentiment analysis, social media	Text analytics, semantic anaysis,	Http://www.meaning cloud.com/	
		algorithms	extraction and manipulation workbench, text processing	ic.com/textpipe	

Textquest	Proprietary	Content analysis, readability analysis.	Text analysis	Http://www.textquest.de/pages/en/general-information.php?Lang=EN	OS Required:MS-Windows and Apple Mac OS-X
Treparel KMX Text Analytics	Proprietary	Machine Learning, SVM powered Classification, fuzzy or probabilistic matching	Text analysis	Http://treparel.com/	
Visualtext	Open source	Natural language processing systems	Information extraction systems and text analyzers.	Http://www.textanalysis.com/	OS Required: Windows 7, XP, Linux.
VP Student Edition	Proprietary	Knowledge network, trend analysis	Text mining and visualisation, text analysis, text processing	Http://vpinstitute.org/wordpress/vp-marketplace/	OS Required:Windows XP (SP2), Vista, and Windows 7 platforms. Some Visual Basic scripts may require Microsoft Excel (2003 or later) and Java.
Aika	Open source	Machine learning, artificial neuronal networks, frequent pattern mining and grammar induction	Syllabification	Http://www.aika-software.org/	Aika is implemented as a Java library.
Data Science Toolkit	Open source	Advanced algorithms	Sentiment Analysis, Language Detection, Topic Classification	Http://www.datasciencetoolkit.org/	
Datumbox	Open source	Machine learning, keyword extraction	Text analysis, search engine optimization, social media monitoring, sentiment analysis	Http://www.datumbox.com/	
GATE	Open source	Natural language processing	Text processing	Https://gate.ac.uk/	
Lingpipe	Open source	Computational linguistics	Text processing, text classification	Http://alias-i.com/lingpipe/	
Microsoft Distributed Machine Learning Toolkit DMTK	Open source	Data parallelization, lightlda, topic model algorithm, and Distributed (Multisense) Word Embedding algorithm	Text and big data analytics	Http://www.dmtk.io/index.html	
Open Calais	Open source	Enhanced tagging engine	Text processing	Http://www.opencalais.com/	
Rapidminer Text Mining	Open source	Machine learning	Data mining and text analytics, text processing	Http://docs.rapidminer.com/	
Reverb: Open Information Extraction Software,	Open source	NLP, WEKA algorithms	Information extraction	Http://reverb.cs.washington.edu/	

ISSN 2347–3657
Volume 7, Issue 3, 2019

S-EM (Spy-EM)	Open source	Naive Bayes and EM algorithm	Text learning or classification system	Https://www.cs.uic.edu/~liub/S-EM/S-EM-download.html	OS Required:Windows
TXM - Unicode, XML, TEI text/corpus analysis platform	Open source	Statistical functions based on R packages	Text analysis	Https://sourceforge.net/projects/txm/	OS Required: Windows, Linux and Mac OS X.
R Programming	Open source	Statistical and graphical techniques	Data transformation and text analysis	https://www.r-project.org/	

III. TEXT MINING TECHNIQUES

As can be observed from Table 1. Some of the popular techniques for text mining are shown in Figure 2:

- ***Natural Language Processing and Machine Learning***: Most of the tools employ Natural Language Processing (21%) and/or Machine Learning techniques (21%) for mining text.
- ***Statistical Methods***: as used for data mining are also applied for text mining. In fact most of the tools use statistical methods (11%) in conjunction with othermethods.
- ***Artificial Intelligence*** **(nine percent)**: techniques such as neural networks are also employed in many text mining tools.
- ***Classification techniques*** **(eight percent)**: are also used to categorize text and documents. These classification techniques must be able to handle unstructured data.
- *Linguistic Learning* (five percent), *Semantic Analysis* (five percent), *and Predictive Modeling* (seven percent) techniques are also employed for mining text.

IV. FEATURES OF TEXT MINING

We analysed the features of tools mentioned in Table1 and the various uses of Text Mining tools are as shown in Figure 3. The major uses of a text mining tool are for:

- ***Text Analytics*****:** involves extracting useful information and patterns from text. Most tools provide this feature.
- ***Text Processing***: involves transforming and manipulating unstructured text so that analysis methods can be appliedto it.
- ***Classification/Categorization***: Many tools are used for classification and categorization of text/documents.
- ***Sentiment Analysis***: is used to identify subjective information from text. Many tools provide for sentiment analysis also called as Opinion Mining.
- ***Knowledge Discovery***: deals with identification of useful information from huge amount of text. Most tools provide for Knowledge discovery and information retrieval features.

ISSN 2347–3657
Volume 7, Issue 3, 2019

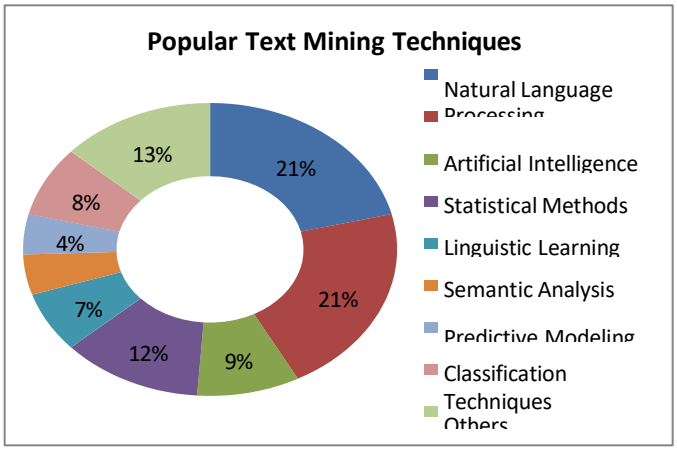

Fig. 2. Popular Text Mining Techniques

Semantic Analysis: involves checking the syntactic structures with the meaning of the text as a whole. Many tools are available that not only provide syntactic analysis but also semantic analysis of the text.

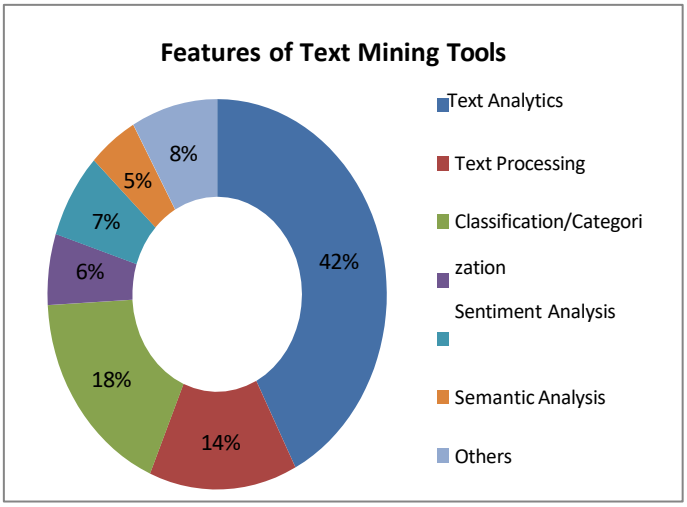

Fig. 3. Features of Text Mining Tools

IV. THE REFLECTIVE PROCESS

Different text mining techniques are available that are applied for analyzing the text patterns and their mining process [16]. Figure 3 shows the Venn diagram for the interrelationship among text mining techniques and their core functionality. Document classification (text classification, document standardization), information retrieval (keyword search / querying and indexing), document clustering (phrase clustering), natural language processing (spelling correction, lemmatization, grammatical parsing, and word sense disambiguation), information extraction (relationship extraction / link analysis), and web mining (web link analysis) [6].

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 7, Issue 3, 2019

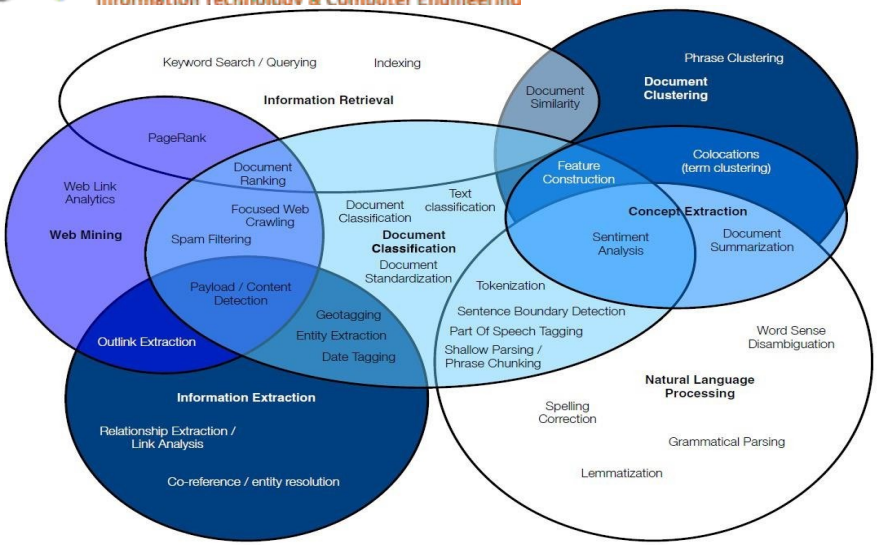

Fig. 4. Inter-relationship among different text mining techniques and their core functionalities [11]

A. Information Extraction

Information Extraction (IE) is a technique that extract meaningful information from large amount of text. Domain experts specify the attributes and relation according to the domain [12]. IE systems are used to extract specific attributes and entities from the document and establish their relationship [13]. The extracted corpus is stored into database for further processing. Precision and recall process is used to check and evaluate the relevance of results on the extracted data. In-depth and complete information about the relevant field is required to perform information extraction process to attain more relevant results [14].

B. Information Retrieval

Information Retrieval (IR) is a process of extracting relevant and associated patterns according to a given set of words or phrases. There is a close relationship in text mining and information retrieval for textual data. In IR systems, different algorithms are used to track the user's behavior and search relevant data accordingly [14]. Google and Yahoo search engines are using information retrieval system more frequently to extract relevant documents according to a phrase on Web. These search engines use query based algorithms to track the trends and attain more significant results. These search engines provide user more relevant and appropriate information that satisfy them according to their needs [20].

C. Natural Language Processing

Natural language processing (NLP) concerns to the automatic processing and analysis of unstructured textual information. It perform different types of analysis such as Named Entity Recognition (NER) for abbreviation and their synonyms extraction to find the relationships among them [21]. NER identify all the instances of specified object from a group of documents. These entities and their instances allow the identification of relationship and other information to attain their key concept. However, this technique lacks complete dictionary list for all named entities used for identification [22], [21]. Complex query based algorithms need to be used to attain acceptable results. In real world, a single entity has numerous terms like TV and

ISSN 2347–3657
Volume 7, Issue 3, 2019

Television. Sometimes, a group of successive words have a multi-word names to identify the boundaries and resolve overlapping issues by using classification technique. Approaches to deal with NER usually fall into four categories: lexicon, rule, statistical based or mixture of these approached. NER systems have achieved the relevance level from 75 to 85 percent [23].

To extract synonym and abbreviation from textual data, co-referencing technique is frequently in use for NLP. Natural Languages (NL) have lot of complexities as a text extracted from different sources don't have identical words or abbreviation. There is a need to detect such issues and make rules for their uniform identification [24]. For example, NER and co-referencing approaches establish a logical relationship to extract and identify the role of person in an organization (use the name of a person at once and then use pronoun instead of name again and again) [25].

D. Clustering

Clustering is an unsupervised process to classify the text documents in groups by applying different clustering algorithms. In a cluster, similar terms or patterns are grouped extracted from various documents. Clustering is performed in top-down and bottom up manner. In NLP, various types of mining tools and techniques are applied for the analysis on unstructured text. Different techniques of clustering are hierarchical, distribution, density, centroid, and k-mean [25].

E. Text Summarization

Text summarization is a process of collecting and producing concise representation of original text documents [26]. Pre-processing and processing operations are performed on the raw text for summarization. Tokenization, stop word removal, and stemming methods are applied for pre-processing. Lexicon lists are generated at processing stage of text summarization.

In past, automatic text summarization was performed on the basis of occurrence a certain word or phrase in document. Later on, additional methods of text mining were introduced with standard text mining process to improve the relevance and accuracy of results [27]. To summarize the text documents, weighted heuristics method extract features by following specific rules. Sentence length, fixed phrase, paragraph, thematic word, and upper case word identification features can be implemented and analyzed for text summerization. Text summarization techniques can be applied on multiple documents at the same time. Quality and
type of classifiers depend on nature and theme of the text documents [28].

V. APPLICATION OF TEXT MINING

- A. Digital Libraries
- B. Academic and Research Field
- C. Life Science
- D. Social Media
- E. Business Intelligence

VI. ISSUES IN TEXT MINING FIELD

Many issues occur during the text mining process and effect the efficiency and effectiveness of decision making. Complexities can arise at the intermediate stage of text mining. In preprocessing stage various rules and regulations are defined to standardize the text that make text mining process efficient. Before applying pattern analysis on the document there is a need to

ISSN 2347–3657
Volume 7, Issue 3, 2019

convert unstructured data into intermediate form but at this stage mining process has its own complications. Sometime real theme or data mislay its importance due to the modification in the text sequence [29]. Another major issue is a multilingual text refinement dependency that create problems. Only few tools are available that support multiple languages [30]. Various algorithms and techniques are used independently to support multilingual text. Because numerous important documents persist outside the text mining process because various tools dont support them. These issues create a lots of problems in knowledge discovery and decision making process. Infect real benefit is difficult to attain by using the existing text mining techniques and tools because its rarely support multilingual documents [31]. Integration of domain knowledge is an important area as it performs specific operations on specified corpus and attain desired outcomes. In this situations domain knowledge from which document corpus to be extracted need to integrate with the computing abilities from which information have to be attained. According to the requirements of the field, experts are needed to work collaboratively from diverse domains to extract more effective, precise and accurate results [32], [33]. The use of synonyms, polysems and antonyms in the documents create problems (abstruseness) for the text mining tools that take both in the same context. It is difficult to categorize the documents when collection of document is large and generated from diverse fields having the same domain.

Abbreviations gives changed meaning in different situation is also a big issue [34]. Varying concepts of granularity change the context of text according to the condition and domain knowledge. There is need to describe rules according to the field that will be used as a standard in the area and can be embedded in text mining tools as a plug-in. It entails lots of effort and time to develop and deploy plug-ins in all fields separately. To develop plug-ins in depth and proper knowledge about the specific domain will be required [35], [36]. Natural languages have lots of complications in itself that create problem in text refinement methods and the identification of entity relationship. Words having same spelling but give diverse meaning, for example, fly and fly. Text mining tools considered both as similar while one is verb and other is noun. Grammatical rules according to the nature and context is still an open issue in the field of text mining [36].

CONCLUSION

The availability of huge volume of text based data need to be examined to extract valuable information. Text mining techniques are used to analyze the interesting and relevant information effectively and efficiently from large amount of unstructured data. This paper presents a brief overview of text mining techniques that help to improve the text mining process.

Data Mining is a very established research field and there are many tools available for data mining. But most of them are not able to handle unstructured data. Since most data is in the form of text i.e. unstructured, there was a need for text mining. Almost everybody works with text, text mining tools are required. There are many open source, proprietary and online text mining tools. This paper discussed some of the popular text mining tools and their features and techniques providing an

overview to researchers and analyst who are new in the field of text mining.

REFERENCES

[1] Feldman, Ronen, and James Sanger. The text mining handbook: advanced approaches in analyzing unstructured data. Cambridge University Press, 2007.

[2] Dekhtyar, Alexander, Jane Huffman Hayes, and Tim Menzies. "Text is software too." MSR 2004: International Workshop on Mining Software Repositories at ICSE'04: Edinburgh, Scotland. 2004.

[3] Sateli, B., Angius, E., Rajivelu, S. S., & Witte, R. (2012). Can text mining assistants help to improve requirements specifications. Mining Unstructured Data (MUD 2012), Canada.

[4] Malhotra, Ruchika, et al. "Severity Assessment of Software Defect Reports using Text Classification." International Journal of Computer Applications83.11 (2013).

[5] Sharma, Gitika, Sumit Sharma, and Shruti Gujral. "A Novel Way of Assessing Software Bug Severity Using Dictionary of Critical Terms."Procedia Computer Science 70 (2015): 632-639.

[6] Jurek, Anna, Maurice D. Mulvenna, and Yaxin Bi. "Improved lexicon-based sentiment analysis for social media analytics." Security Informatics 4.1 (2015): 1-13.

[7] Eom, Jae-Hong, and Byoung-Tak Zhang. "Pubminer: Machine learning-based text mining system for biomedical information mining." Artificial Intelligence: Methodology, Systems, and Applications. Springer Berlin Heidelberg, 2004. 216-225.

[8] Gamon, M., Aue, A., Corston-Oliver, S., & Ringger, E. (2005). Pulse: Mining customer opinions from free text. In Advances in Intelligent Data Analysis VI(pp. 121-132). Springer Berlin Heidelberg.

[9] Bragge, Johanna, and Jan Storgårds. "Profiling academic research on digital games using text mining tools." Proceedings of DiGRA 2007 Conference. 200

[10] S. M. Weiss, N. Indurkhya, T. Zhang, and F. Damerau, Text mining: predictive methods for analyzing unstructured information. Springer Science and Business Media, 2010.

[11] W. He, "Examining students online interaction in a live video streaming environment using data mining and text mining," Computers in Human Behavior, vol. 29, no. 1, pp. 90–102, 2013.

[12] R. Agrawal and M. Batra, "A detailed study on text mining techniques," International Journal of Soft Computing and Engineering (IJSCE) ISSN, pp. 2231–2307, 2013.

[13] D. S. Dang and P. H. Ahmad, "A review of text mining techniques associated with various application areas," International Journal of Science and Research (IJSR), vol. 4, no. 2, pp. 2461–2466, 2015.

[14] R. Steinberger, "A survey of methods to ease the development of highly multilingual text mining applications," Language Resources and Evaluation, vol. 46, no. 2, pp. 155–176, 2012.

[20] N. Zhong, Y. Li, and S.-T. Wu, "Effective pattern discovery for text mining," IEEE transactions on knowledge and data engineering, vol. 24, no. 1, pp. 30–44, 2012.

[21] B. Laxman and D. Sujatha, "Improved method for pattern discovery in text mining," International Journal of Research in Engineering and Technology, vol. 2, no. 1, pp. 2321–2328, 2013.

[22] A. Henriksson, H. Moen, M. Skeppstedt, V. Daudaravi˘cius, and

ISSN 2347–3657
Volume 7, Issue 3, 2019

M. Duneld, "Synonym extraction and abbreviation expansion with ensembles of semantic spaces," Journal of biomedical semantics, vol. 5, no. 1, p. 1, 2014.

[23] A. M. Cohen and W. R. Hersh, "A survey of current work in biomedical text mining," Briefings in bioinformatics, vol. 6, no. 1, pp. 57–71, 2005.

[24] E. A. Calvillo, A. Padilla, J. Mu˜noz, J. Ponce, and J. T. Fernandez,"Searching research papers using clustering and text mining," in Electronics, Communications and Computing (CONIELECOMP), 2013 International Conference on. IEEE, 2013, pp. 78–81.

[25] B. L. Narayana and S. P. Kumar, "A new clustering technique on text in sentence for text mining," IJSEAT, vol. 3, no. 3, pp. 69–71, 2015.

[26] B. A. Mukhedkar, D. Sakhare, and R. Kumar, "Pragmatic analysis based document summarization," International Journal of Computer Science and Information Security, vol. 14, no. 4, p. 145, 2016.

[27] C. P. Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," Information Sciences,vol. 275, pp. 314–347, 2014.

[28] R. Al-Hashemi, "Text summarization extraction system (tses) using extracted keywords." Int. Arab J. e-Technol., vol. 1, no. 4, pp. 164–168, 2010.

[29] A. Henriksson, J. Zhao, H. Dalianis, and H. Bostr¨om, "Ensembles of randomized trees using diverse distributed representations of clinical events," BMC Medical Informatics and Decision Making, vol. 16, no. 2,p. 69, 2016.

[30] H. Solanki, "Comparative study of data mining tools and analysis with unified data mining theory," International Journal of Computer Applications, vol. 75, no. 16, 2013.

[31] A. Kumaran, R. Makin, V. Pattisapu, and S. E. Sharif, "Automatic extraction of synonymy information:-extended abstract," OTT06, vol. 1, p. 55, 2007.

[32] B. L. Narayana and S. P. Kumar, "A new clustering technique on text in sentence for text mining," IJSEAT, vol. 3, no. 3, pp. 69–71, 2015.

[33] A. Henriksson, J. Zhao, H. Dalianis, and H. Bostr¨om, "Ensembles of randomized trees using diverse distributed representations of clinical events," BMC Medical Informatics and Decision Making, vol. 16, no. 2, p. 69, 2016.

[34] A. Kaklauskas, M. Seniut, D. Amaratunga, I. Lill, A. Safonov, N. Vatin,J. Cerkauskas, I. Jackute, A. Kuzminske, and L. Peciure, "Text analytics for android project," Procedia Economics and Finance, vol. 18, pp. 610–617, 2014.

[35] A. Kumaran, R. Makin, V. Pattisapu, and S. E. Sharif, "Automatic extraction of synonymy information:-extended abstract," OTT06, vol. 1, p. 55, 2007.

[36] N. Samsudin, M. Puteh, A. R. Hamdan, and M. Z. A. Nazri, "Immune based feature selection for opinion mining," in Proceedings of the World Congress on Engineering, vol. 3, 2013, pp. 3–5.