Proceedings of the Federated Conference on Computer Science and Information Systems pp. 439–445
DOI: 10.15439/2016F221
ACSIS, Vol. 8. ISSN 2300-5963

Semantic Knowledge Extraction from Research Documents

Rishabh Upadhyay, Akihiro Fujii
Department of Applied Informatics,
Hosei University, Tokyo, Japan
Email: uhrishabh@gmail.com, fujii@hosei.ac.jp

***Abstract*—In this paper, we designed a knowledge supporting software system in which sentences and keywords are extracted from large scale document database. This system consists of semantic representation scheme for natural language processing of the document database. Documents originally in a form of PDF are broken into triple-store data after pre-processing. The semantic representation is a hyper-graph which consists of collections of binary relations of 'triples'. According to a certain rule based on user's interests, the system identify sentences and words of interests. The relationship of those extracted sentences is visualized in the form of network graph. A user can introduce new rules to extract additional Knowledge from the Database or paper. For practical example, we choose a set of research papers related IoT for the case study purpose. Applying several rules concerning authors' indicated keywords as well as the system's specified discourse words, significant knowledge are extracted from the papers.**

***Index Terms*—Knowledge extraction; Semantics; Ontology; Discourse; Science and technology foresight**

I. Introduction

KNOWLEDGE extraction can be defined as the creation of information from structured or unstructured data. The general purpose of Knowledge discovery is to "extract implicit, previously unknown, and potentially useful information from data" [1]. Due to continuous growth of electronic articles or documents, automated knowledge extraction techniques become more and more important. The extraction of useful information from unstructured or semi-structured articles is attracting attention [3-7]. Knowledge mining can be characterized as concerned with developing and integrating a wide range of data analysis methods that are able to derive directly or incrementally new knowledge from large or small volumes of data from structured or unstructured sources using relevant prior knowledge [2]. Text mining tools in this context have the capability to analyze large or small quantities of natural language text and detect lexical and linguistic usage patterns [8]. The extracted information also should be machine understandable as well as human understandable in terms of open-data perspective.

This paper proposes a new method for Knowledge extraction or mining based on the integration of Semantics Technology (ST), Natural language processing (NLP) and Information extraction (IE). ST and NLP are significant topics in recent years. Knowledge extraction works in iterative manner, starting with a small set of rules which are tested on the available corpora or dataset and extended until the desired recall value is reached. The process of extracting knowledge is guided by certain rules inputting to the system which will define the knowledge according to the interests of a particular user. Semantic technology is based on RDF model. NLP concerns with the correction of the sentences and text which is obtained after IE phase. In this paper, we explore the benefit and application that can be achieved by the integration of these three areas for knowledge mining.

As we assume, one of the application of this system would be foresight scenario building based on the results where experts are working on writing or discussing technology trend of a certain field of research. Those experts are not necessary knowledgeable about technical issue written in a research paper, so that extracting fragments of knowledge and facts should be provided to the user compactly and easily in a limited time period.

In section 2, we give an introduction and background to Text mining as well as Technology forecasting. In section 3, introduce our model that we have used for knowledge extraction or mining. Details of the application of our system are introduced in section 4. Then in section 5 we examine the result acquired by the proposed model. Related works is presented in section 6. Conclusion and future works are introduced in section 7.

II. Background

A. Text Mining

Research in text mining has been carried out since the mid- 80s when the Prof Don Swanson, realized that, by combining information slice or fragments from seemingly unrelated medical articles, it was possible to deduce new hypotheses [13]. "Text mining" is used to describe the

978-83-60810-90-3/$25.00

application of data mining techniques to automated discovery of useful or interesting knowledge from unstructured text such as email, text documents etc. [9]. Several techniques have been proposed for text mining including association mining [10 and 11], decision tree [12], Machine learning, conceptual structure and rule induction methods. In addition, Information extraction and Natural language processing are the fundamental methods for Text Mining [23].

In Information extraction, natural language texts are mapped into predefine, structured representation, or templates, which, when it is filled, represent an extract of key information from the original text [15]. So the IE task is defined by its input and output. The work of [16] and [17] present text mining using IE. The Application of text mining can be extended to various sectors where text documents exist. For example, Crime detection can profit by the identification of similarities between various crimes. Some of the past researches in the field of Text mining or knowledge extractions are as follows:

Rajman and his colleagues [24] presented an approach for knowledge extraction using Text Mining technique. They have presented two example of information that can be extracted from text collections- probabilistic association of keywords and prototypical document instances. They have given the importance of the Natural language processing tools for such knowledge extraction. So his was the base for our method.

Alani and his team [25] have provided an updated for Artequakt System. This system uses Natural Language processing tools to automatically extract knowledge about artists from documents using predefined ontology. Steps for knowledge extraction are as follows: Document Retrieval, Entity Recognition and Extraction procedure. In knowledge extraction procedure, consists of Syntactical analysis, Semantic Analysis and Relation extraction. They have produced acceptable results.

Peter Clark and Phil Harrison [26] worked on knowledge extraction by making database of "tuples" and thus capturing the simple word knowledge. And then using it in improving parsing and the plausibility assessment of paraphrase rules used in textual entailment.

Parikh [27] proposed an approach to learn a semantic parser for extracting nested event structures with or without annotated text. The idea behind this method is to model the annotations as latent variables and incorporate prior that matched with Semantics parses of the events.

B. Technology Forecasting and foresight:

Technology forecasting is used widely by the private sector and by governments for applications ranging from predicting product development or a competitor's technical capabilities to the creation of scenarios for predicting the impact of future technologies [19]. It is "the prediction of the invention, timing, characteristic, performance, technique or process serving some useful purpose". Detailed account of achievements and failures of the technology forecasting over the four decades is given by Cuhls [19]. Johnston [20] proposed five stages in the chronology of foresight, technology forecasting and futurism leading to technology foresight, which can be used for wide understanding of the economic and social processes that shape technology. Foresight can be referred as systematic process of reflection and vision building on the future among a group of stakeholders. Foresight is nowadays referred as an element in a continuous policy learning process that is contributing to a more effective mode of policy making [21]. In a European research group, foresight is described as "...a systematic, participatory, future intelligence-gathering and medium-term vision-building process aimed at present-day decisions and mobilizing joint actions" [22].

III. Related Works

Their have been many research in the field of literature mining or extracting knowledge from research documents. But most of them are related to Biomedical or medicinal field. Our approach was related to the field of Technologies. QasemiZadeh [33] have presented an approach for structure mining from Scientific or research papers. He has only processed using Language processing, but we have combined three main fields to extract knowledge. Cimiano et. al. [34] gave a survey of current methods in ontology constructions and also discussed relation between ontology and Natural language processing.

Mima et. al. [35] gave a method for knowledge management using ontology-based similarity calculation. We have also used ontology for extracting knowledge but apart from ontology we have also given emphasis on Natural language processing and a bit of Information extraction (early stage). Mima have also not presented any information regarding evaluation of the system. Hahm et. al. [36] presented an approach for knowledge extraction from web data. Triple store was produced using web data. But in our method we have given more emphasis on research document and producing triple with it and then ontologies is applied on the produced triple datasets of line and sentences.

IV. Proposed System

Proposed system uses combination of semantics of sentences and natural language processing technique over the sentences. It also provides visualization of the result. We do not expect fully machine processing results from the system. In a sense that after some processing by inference rule and getting sentences which might be significant, user creativity is required to understand what is written in the document. In this sense, our method is hybrid with software processing and expert knowledge in the area. The following Fig.1 describes the proposed model of our system.

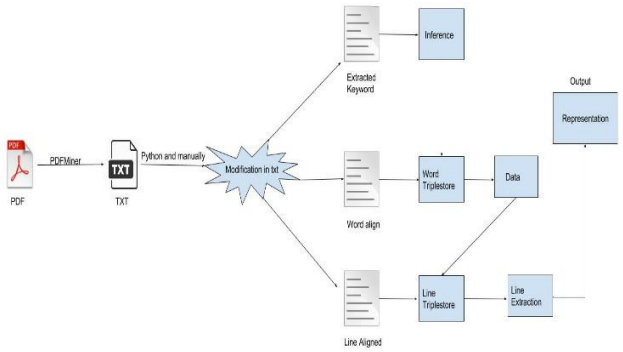

Fig. 1 Pipeline of Proposed System

Our model consists of the following Steps which is given below:

1. Extracting information from PDF file to text file.
2. Pre-processing of data.
3. Extracting the Keywords.
4. Extraction of Discourse words
5. Building a simple Triple-store (Word and Line).
6. Inference Rule.
7. Visualizing the data.

1) Extracting information from PDF file to Text file:

We made a dataset of IOT research papers. We had about 69 research papers dataset. We used some paper out of those papers. The paper were in pdf format. So to work on them we have to extract information from the pdf files. So to extract information we used Python Library "PDFMiner" [29].

2) Pre-processing of data:

After extracting information from *pdf* file into text file pre-processing was done. The extracted documents had problems, as extracted information was on a single line. So to work on them the modification of the extracted information is required. So the modification includes removing the noise (image), make proper alignment of the sentences, checking the extracted data.

So we removed the noise which is caused by extracting the image from the pdf file. As image will contains least information in the form of text so we ignored the text extracted by image. We align each sentence on the separate line and also checked the percent of the extracted data. Most of the file had some errors so we corrected it manually.

3) Extracting the Keywords:

As this system will be extracting the important knowledge from the research papers, rules should be created according to the important Keywords. So, the easiest way was to extract the Keywords from the research papers.

We worked on the research paper with the standard structure such as Title, Abstract, Keywords then Main text. So the extraction of Keywords from the research paper takes places only for specific structure. So we automated the process of extracting the Keyword using python's regular expression. After getting keywords the frequency of the extracted words is obtained. Because those frequent words will be used by the inference rule to create new rules for extracting useful sentences from documents.

4) Extraction using Discourse words:

Discourse words are the words which give "important message" that helps us to understand or comprehend a text. This discourse words ranges from Numbering words to Adding words, linking words to Contrasting words (however) etc. These words are used everywhere from articles to research papers. So we emphasis on these words to represent knowledge or message from the research articles. We went through 5-6 paper to get the list of the discourse words. Those words were used in making new rules for the extraction of the message from the articles. Then after creating the rules for those 5-6 papers we then used those words to create the rules for other research documents to check, if these words can be generalized.

5) Building a simple Triple-store:

Till now we were doing the pre-processing of the data and collecting words for making new rules. In this section we will discuss about the schema to analyze the sentence and word data. We focused on two triple-stores that is, sentence and words separately.

There are many semantics toolkit available. We have referred Python code [28]. We choose this programming language because of its simplicity and flexibility towards various toolkits in the field of semantics and Artificial intelligence.

So from the above toolkit we have produced a dataset of sentences and words. In fact, this dataset is of three type formats that is why single data it is known as "triple". The three type formats are Subject, Predicate and Object.

We have maintained two triple-stores. The format of both the Triple-Store is shown below.

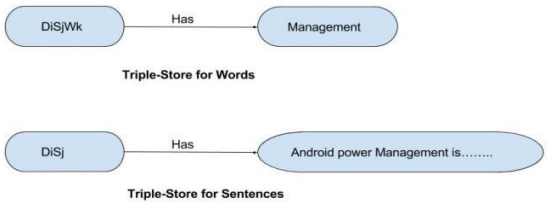

Fig.2 RDF graph example

Above Fig.2 shows a triple syntax called an RDF graph. RDF, an abbreviation for "Resource Description Framework," is a concept adopted in defining knowledge structure. Knowledge fragments are given in a syntax consisting of three elements: the subject, the predicate, and the object.

Let's consider the triple store for words for the above example, the subject is "*DiSjWk*", Predicate is "*Has*" and Object is "Management". So the subject *DiSjWk* stands for *i* th Document, *j* th Sentence and *k* th word. We have chosen the predicate to be "*Has*". The Object will be the extracted word from the paper.

For the sentences triple-Store, We have subject "*DiSj*", Predicate "*Has*" and Object will be the Sentence extracted from the document. Subject *Di* stands for *i*th Document and *Sj* stands for *j*th sentence. This Triple-Store will be base for our Experiments.

So this triple-Store will be used for extracting the knowledge or useful terms from the documents. For extracting the so called "knowledge", we will use Inference rule, which is introduced in next step.

6) Inference Rule:

Inference is the process of deriving new information from the information you already have [28]. So the "information" and which rule to apply to extract the information will vary depending on the context. As we have explained the structure of the Knowledge fragments that are given in a three elements structure. So to describe ontologies, logical expression is configured. So the process to configure the logical expression uses a syntax called predicate. Ontologies are written in OWL. Ontology is defined as the explicit and abstract model representation of already defined finite sets of terms and concepts, involved in knowledge management, knowledge engineering, and intelligent information integration [30]. In simple word, ontology is the collection and classification of terms to recognize their semantic relevance. OWL stands for Web Ontology Writing Language; its standardization has been conducted by W3C. To describe a knowledge structure in a predicate logic, a set of elements that meet a certain condition is constructed, such as "*If A, then B*". After construction, the resulting set is Fundamental to knowledge processing in the semantic data processing. Knowledge processing based on the predicate logic takes the form of generating answers from the collection of the knowledge fragments, such as *"If A is true, then B is satisfied*" and "*If B, then C,*" to queries such as "*Is Z true, if A holds?*". This process is referred to as the reasoning mechanism.

The basic pattern of inference is simply to query the information relevant to the rule and then apply the transformation that turns these bindings into a simple triple data added to the main Triple-Store, with appropriate Subject, predicate and Objects. After getting this new triple data, we use this information to process the knowledge from whole tripe-store, so Fig.3 gives the insight of our Triple data sets.

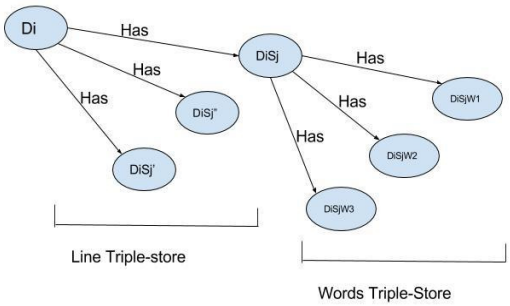

Fig.3 Triple Data Example

We have data sets with line and word Triple store. The knowledge processing or extraction of line data set uses the words triple store. We use the rule on word triple store, added the new triple data using inference rule and then we extract the knowledge from the line data sets. The steps are shown in the Fig.4 below.

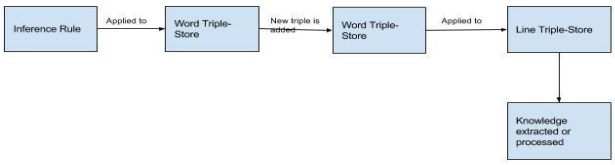

Fig.4 Pipeline of the Inference System

7) Visualization:

So, using the above process we extracted the useful and meaningful line from whole documents. Then we emphasized on visualization of the extracted knowledge. For visualization, we created new triple dataset to make a relationship between words and line. This dataset will show which word is in which line. Subject will be word, Predicate will be "Has" and Object will be sentences. So using this we got a third triple store that will consist of words and sentences together. So we will use this RDF file for visualization of the knowledge.

We used *GraphViz* graph drawing package [31]. This library has support extrinsic data in its graph drawing algorithm. It offers one of the best solution for drawing directed cyclic graphs which is state-of-the-art in graph drawing. *GraphViz* application *neato* [32] is a straightforward method for rapidly visualizing undirected graphs. The graphs are generated by using:

"neato-Teps undirected.gv > undirected.eps",

Where undirected.gv is a file that contains the code of the graph, –Teps specifies an Encapsulated Postscript output format. So using the above given library we visualized our knowledge.

V. APPLICATION

Natural language processing (NLP) is one of main field of artificial intelligence study. Extraction of meaningful information from vast amount of document database is an example of the features that NLP possibly contribute well in performing scenario building practices.

On the other hand, Science and Technology Foresight (STF) is an activity to identify important research areas or directions as well as to search key features and driving forces for those researches. For STF, there are numerous papers and books on methodologies. The scenario development is one of main measures for such future-oriented work. However, even in scenario development alone, there is wide range of possible practices depending on purpose of outcome, participant's knowledge etc.

For the scenario building method in STF, the creativity of those who participating the workshop is the most important. Usually such an activity is held in limited time scale. In order to stimulate the creativity of participant in short period of times at scenario building workshop, for example, quick identification of important sentences of the area of interests from large number of research papers is quite helpful. At the same time results of extractions could be provided with visual imaging with of data processing and graphic visualization software. With the help of such a system, human experts may exercise more efficiently during scenario building workshop.

In this context, we have designed an artificial intelligence support for the scenario building activities in this paper. We applied semantic information about structure of sentences in research papers and support human user quick visualization of extracting of sentences and key words for the purpose.

By using algorithm or method, we have achieved two aspects to a certain extent. Scenario writing is one of the example of such system could support efficiently. In the process, firstly some key driver is necessary. We may use current algorithm to get meaningful sentences that may help understanding scenario drivers. Then our algorithm also could produce different sentences with slightly different angle of interests.

The extracted sentences and words might help the scenario writer to write down case scenario.

VI. EXPERIMENTAL RESULTS.

This section show the example of using Hybrid method. We have a dataset of 69 research papers which we developed using varius sources. This dataset consist of paper which was pubished in IEEE journals and conferences. For results we have given emphasis on two methods, that are:

- Extracting the knowledge using Keywords from the paper.
- Extracting the knowledge using discourse Keywords.

As the dataset consist of the research paper which were in PDF format. So the first step is to extract the information from the PDF document so that it would be easier to work on them. PDFminner was used for this process.Then the prepocessing of the Information takes place. In prerocessing the unwanted extracted information was removed. While using PDFminner, the images in the document was extracted but contains many noise or unwanted characters. So preprocessing emphasis on removing the unwanted characters and also aligning the line using python. Then after the preprocessing we give importance on extarcting the keyword from all the documents and then we can choose the most frequent words from the document. These words were used directly from the word mentioned in the keywords section. As the keyword section consists of the words which are important acording to the context. So all this processing and counting of the word is done automatically using python. The Table1 shown below consists of frequent term in all the 69 documents:

Table 1 Keywords Examples of the 69 papers

Internet of Things	42
Wireless-network sensors	22
RFID	13
Social network	8
System security	8
Energy Efficiency	7
Service Management	7
Enterprise systems	6
Learning technology	5
6lowpan	4

After this the formation of tripe-store takes place. The dataset consists of 69 documents, with about 21K lines and 300K words before filtering. After filtering we got about 200K words and 21K lines. So two triple-store are created one for lines and one for words. Subject is choosen such that it have some relationship between two triple-store. After the making of triple-store and all process now the important step is to extract the knowledge using INFERENCE rule. This consist of the SPARQL type query. So the user makes important rule ,through this rule the knowledge is extracted. The main step in this method is INFERENCE rule. This rule can be created by user according to their preference. So this

rule will help to extract the knowledge from the Huge database rather then reading all the documents. To ease the understanding we also given importance to visualise the knowledge.

The above method was related to our first method that we mentioned in the start of this section.

In next method, we gave attention to the discourse keywords in the documents. Discourse keywords is the keywords which talks and give information about the text. After going through set of papers we came up with some Discourse keywords which are listed in Table2.

Table 2 Examples of Discourse Keywords

consists	become	aimed	Instead
useful	capable	using	Provide
method	propose	enhance	application
future	explore	aspects	Discover
objectives	focused	pedagogy	Crucial
different	various	integration	Import
promote	reflect	classified	Need

So new rule is generated using the discourse term which is mentioned above. We used this words in two form first on training set then on test set. Using training documents we got those words, then we used these words in some other documents to check the effectiveness of the words.

Here, we chose one of the 69 papers as a case study example. Document [37] is a servey paper about IoT applications of RFID(Radio Frequency IDentifier). We extracted some useful sentences which can be considered as significant knowledge descriptions about the application. Some example sentences are given below with the associated discourse term as results from the system.

"aimed" -- "this paper is aimed at drawing a landscape of the current research on rfid sensing from the perspective of iot for personal healthcare

"application" -- "thanks to the recent advantages in biomaterial engineering, a rich variety of tattoo-like thin surface electronic devices have been experimen- ted in the last three years

"based" -- "the body-centric networks investigated so far are mostly based on active devices"

"bring" -- "data processing and human behavior analysis the interactions between people and their habitat bring precious information on behavioral parameters concerning the activity rate during the different phases of the day and of the night"

"useful" -- "air monitoring is also useful in common spaces with the purpose to avoid the inhalation of anesthetic gases"

Althought those are not whole result but only the fragments of whole content of the paper, sentences indicates condenced informations about several issues discussed in the paper. We have evaluated usefulness of extracted sentences over 10 documents out of 69 papers. In each paper, there are around 10 discourse terms, but most of them were same as before and after that the inference rules were introduced based on the network visualization in the process of extraction.

In the Fig.5 we have presented both the methods i.e. Keywords and Discourse words. So first we have extracted Keywords from both the sources and then we have used the semantic analysis on it to extract the knowledge from it. The output is in the form of graph which is easy to understand. One example of the extracted knowledge is given below which is extracted by using keyword "Pedagogy".

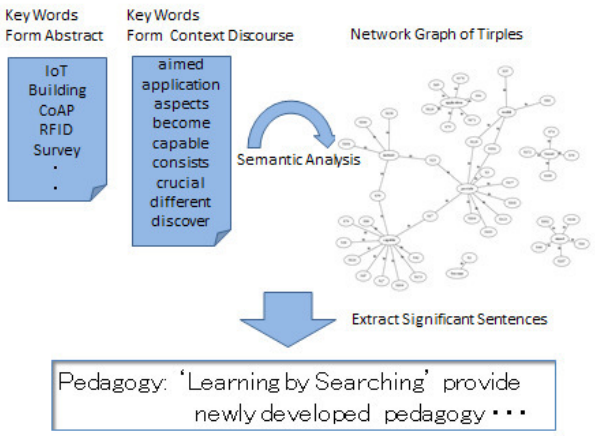

Fig.5 Example Visualization of Konwledge

Apart from this to check the extracted information we calculated the precision of the information. So after using this Hybrid method, which range from Information extraction to Semantics for knowledge extraction, the precision of the model was good.

VII. Conculsion And Future work.

In this article, we have discussed practical sentence extraction procesure and supporting system which we intended to call knowledge extraction system. Since processed data is always stored in a form of triples, resulted dataset is always fully machine readable in every stage of cyclic extraction and cleaning of data. The system assumes human exparts support in selecting so called discourse keywords. Such characteristics is useful and practical in the situations where exparts need to aquire a certain level of knowledge in a research area such as Science and Technology Foresight activities. As we have also shown, that the above introduced or obtainted discourse words can be used in any research documents and on the basis of that words useful information can be obtainted.

There are two directions for enhancing the system. One is to inroduce more sophisticated inference rules over sentences. Purhaps it will come with NLP technique with looking at grammatical structure of sentences. In addition to this, the system can try to extract discourse words from the extracted lines (using Keywords) rather than those words which are mentioned by us. Another direction goes toward practical utilizations. Big data analysis is currently one of the most needed technology in IT related services. The availability of software based data processing is very important aspect of reusing and deepining knowledges obtaind in a stage of processing.

REFERENCES

[1] W. Frawley and G. Piatetsky-Shapiro and C. Matheus, Knowledge Discovery in Databases: An Overview. AI Magazine, 1992,213-228.

[2] Michalski, R.S.: Knowledge Mining: A Proposed New Direction, In: Invited talk at the Sanken Symposium on Data Mining and Semantic Web, Osaka University, Japan, March 10-11, 2003.

[3] Jérôme Darmont, chair. Proceedings of the 15th international conference on extraction and knowledge management, Luxembourg, 2015.

[4] Jerzy Grzymala-Busse, Ingo Schwab, Maria Pia di Buono, editor. Proceedings of the second on Big Data, Small Data, Linked Data and Open Data (ALLDATA 2016) workshop on Knowledge Extraction and Semantic Annotation. Portugal, 2016.

[5] International World Wide Web Conferences (WWW 2015) Second workshop on Knowledge Extraction from Text, Italy, 2015.

[6] XIV Conference of the Spanish Association for Artificial Intelligence (CAEPIA 2011) workshop on Knowledge Extraction and Exploitation from semi-Structured Online Sources, Spain, 2011.

[7] 1st international Workshop on Knowledge Extraction and Consolidation from Social Media collocated with the 11th International Semantic Web Conference (ISWC), USA, 2012.

[8] F. Sebastiani, "Machine learning in Automated Text Categorization," ACM Computing Surveys, vol. 1, no. 34, pp. 1–47, 2002.

[9] J. Han and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco, 2000.

[10] Dion H. Goh and Rebecca P. Ang (2007), "An introduction to association rule mining: An application in counseling and help seeking behavior of adolescents", Journal of Behavior Research Methods39 (2), Singapore, 259-266.

[11] Pak Chung Wong, Paul Whitney and Jim Thomas," Visualizing Association Rules for Text Mining", ", International Conference, Pacific Northwest National Laboratory, USA, 1-5.

[12] C. Apte and F. Damerau and S. M. Weiss and Chid Apte and Fred Damerau and Sholom Weiss," Text Mining with Decision Trees and Decision Rules", In Proceedings of the Conference on Automated Learning and Discorery, Workshop 6: Learning from Text and the Web, 1998.

[13] J. Nightingal, "Digging for data that can change our world," the Guardian, Jan 2006.

[14] Grishman R. (1997), "Information Extraction: Techniques and Challenges", International Summer School, SCIE-97.

[15] Wilks Yorick (1997), "Information Extraction as a Core Language Technology", International Summer School, SCIE-97.

[16] H. Karanikas, C. Tjortjis, and B. Theodoulidis, "An approach to text mining using information extraction," in Proceedings of Workshop of Knowledge Management: Theory and Applications in Principles of Data Mining and Knowledge Discovery 4th European Conference, 2000.

[17] U. Nahm and R. Mooney, "Text mining with information extraction," in Proceedings of the AAAI 2002 Spring Symposium on Mining Answers from Texts and Knowledge Bases, 2002.

[18] Committee on Forecasting Future Disruptive Technologies; Air Force Studies Board; Division on Engineering and Physical Sciences; National Research Council. "Persistent Forecasting of Disruptive Technologies",2009

[19] Cuhls K," From Forecasting to Foresight Processes – New Participative Foresight Activities in Germany", Journal of Forecasting, 23, pp 93–111 European Foresight Monitoring Network, available at http://www.efmn.info/.

[20] Johnston R," The State and Contribution of Foresight: New Challenges". In Proceedings of the Workshop on the Role of Foresight in the Selection of Research Policy Priorities' IPTS, Seville.

[21] Weber, M., 'Foresight and Adaptive Planning as Complementary Elements in Anticipatory Policy-making: A Conceptual and Methodological Approach' In: Jan-Peter Voß, Dierk Bauknecht, René Kemp (eds.) Reflexive Governance For Sustainable Development Edward Elgar, pp. 189-22.

[22] FOREN 2001: A Practical Guide to Regional Foresight. FOREN network, European Commission Research Directorate General, STRATA programme.

[23] S. Jusoh and H. M. Alfawareh, "Techniques Techniques, Applications and Challenging Issue in Text Mining." IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 2, November 2012.

[24] Martin Rajman and Romaric Besancon,"Text mining- Knowledge extraction from unstructured textual data".In:Proceedings of the 6th Conference of the International Federation of Classification Societies, Rome, 1998.

[25] Alani, Harith, Kim, Sanghee, Millard, David E., Weal, Mark J., Lewis, Paul H., Hall, Wendy and Shadbolt, Nigel R," Automatic Extraction of Knowledge from Web Documents", Wendy; Lewis, Paul H. and Shadbolt, Nigel R. In, *2nd International Semantic Web Conference - Workshop on Human Language Technology for the Semantic Web abd Web Services, Sanibel Island, Florida, USA,20 - 23 Oct 2003.*

[26] Peter Clark and Phil Harrison,"Large-Scale Extraction and Use of Knowledge From Text", In: Proceedings of the fifth international conference on Knowledge capture(K-CAP '09),USA, 2009.

[27] Ankur P. Parikh; Hoifung Poon; Kristina Toutanova," Grounded Semantic Parsing for Complex Knowledge Extraction",In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,2015.

[28] Toby Segaran, 'Semantic Web Programming', O'reilly, 2009.

[29] Shinyama, Y. (2010) PDFMiner: Python PDF parser and analyzer. Retrieved on 11 June 2015 from:http://www.unixuser.org/~euske/python/pdfminer/.

[30] D.Fensel, "Ontologies: Silver Bullet for Knowledge Management and e-Commerce", Springer Verlag, Berlin, 2000.

[31] J. Ellson, E. R. Gansner, L. Koutsofios, S. C. North, and G. Woodhull. Graphviz — open source graph drawing tools. In P. Mutzel, M. Junger ¨, and S. Leipert, editors, Proc. 9th Int. Symp. Graph Drawing (GD 2001), number 2265 in Lecture Notes in Computer Science, LNCS, pages 483–484. Springer-Verlag, 2002.

[32] Stephen C. North, "Drawing graphs with NEATO", NEATO User manual, April 26, 2004.

[33] Behrang QasemiZadeh, "Towards Technology Structure Mining from Scientific Literature", 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010.

[34] Cimiano, P., Buitelaar, P., Völker, J.: Ontology construction. In: Indurkhya, N., Damerau, F.J. (eds.) Handbook of Natural Language Processing, 2nd edn., pp. 577–605.

[35] Mima, H., Ananiadou**, S**. & Matsushima, K. (2004) Design and Implementation of a Terminology-based literature mining and knowledge structuring system, in Proceedings of international workshop of Computational Terminology, CompuTerm, Coling, Geneva , Switzerland.

[36] Younggyun Hahm, Hee-Geun Yoon, Se-Young Park, Seong-Bae Park, Jungwon Cha, Dosam Hwang, Key-Sun Choi, Towards Ontology-based Knowledge Extraction from Web Data with Lexicalization of Ontology for Korean QA System, Submitted to NLIWoD, 2014.

[37] S. Amendola, R. Lodato, S. Manzari, C. Occhiuzzi, and G. Marrocco "RFID Technology for IoT-Based Personal Healthcare in Smart Spaces", IEEE INTERNET OF THINGS JOURNAL, VOL. 1, NO. 2, APRIL 2014.