ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

Survey on Information Extraction and Summarization of Multiple News Articles

V.Swathi Lakshmi[1], Dr. S. Lakshmana Pandian[2],

Dept. of Computer Science and Engineering, Pondicherry Engineering College, Puducherry, India[1]

Associate Professor, Dept. of CSE, Pondicherry Engineering College, Puducherry, India[2]

ABSTRACT: Now a days the web contains huge and growing amount of information. The information explosion in web consumes more time to read and grasp information from different sources. News Articles are published by large groups of News Providers and they generate tons of news contents each and every minute of the day. The user needs a lot of analysis to understand the main idea described in the documents. News summarization is a process, which creates a shorter version of multiple news articles by preserving the information and overall meaning. It helps humans to digest the main contents of related documents. Hence we proposed an IE-supported summarization system by combining Ontology based Information Extraction and Abstraction-based Multi-document summarization.

KEYWORDS: Information extraction, ontology, Multi-Document Summarization, Natural Language Processing (NLP).

I. INTRODUCTION

The amount of information available today is tremendous and the problem of finding the relevant pieces and making sense of these is becoming more and more essential. Nowadays, a great deal of information comes from the Internet in a textual form. The challenge of finding relevant documents on the web is mainly handled by information retrieval techniques utilized in search engines such as Google, Bing, Yahoo, etc. Search engines usually return thousands of pages for a single query, and even the use of sophisticated ranking algorithms can't provide us the exact information we are looking for. A typical user goes through the top-ranked pages and tries to find the relevant pieces of information he or she is interested in, manually. Obviously, a short summary of the retrieved pages would be very helpful in such situations. In general, construction of summaries is an ideal way to cope with the information overload.

Information Extraction aims to retrieve certain types of information from natural language text by processing them automatically. Ontology-based information extraction has recently emerged as a subfield of information extraction. Here, ontology's which provide formal and explicit specifications of conceptualizations, play a crucial role in the information extraction process. Because of the use of anthologies, this field is related to knowledge representation and has the potential to assist the development of the Semantic Web. Information extractions take an unrestricted text and "summarize" that text with respect to a prespecified topic or domain of interest. Find useful information about the domain from the summarized text. An example IE system where news articles are taken as input and information is extracted on certain events related to a specific company [1]. For example, once user will give the query in the search engine. The search engine collects all domain information from web document to the information extractor extracts all the information from those documents and submits it to the ontology. From the ontology all domain related documents are clustered separately with the related domain names. The information extractor again extracts all needed information from the ontology clusters and passes it to the summarizer.

A summary is a shortened version of a text that contains the main points of the original content. Automatic summarization is the creation of a summary by a computer program. Although automatic summarization is a hot topic of research nowadays, only very few software tools are available to the end users and none of them are particularly popular. The reason for this should be the low quality of automatically produced summaries. In general, creation of a good summary requires a lot of intelligence. A clear evidence of this is the use of summaries in educational process. Like many other natural language processing (NLP) tasks, a high quality automatic summarization will require understanding of a natural language, at least to a certain degree. This is known to be an AI-complete task, that is, it

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

requires a computer with human-level intelligence that strong AI claims. Despite philosophical discussions about possibilities for strong AI, even with existing methods, scientists could achieve decent results in NLP tasks that are quite challenging, such as machine translation, speech recognition, domain specific question answering, etc. Although none of these problems are near to be solved yet, the results are promising to be useful. Improving the quality of automatic summarization to this level of usefulness is the motivation behind the increasing amount of research in the field, including mine.

II. ONTOLOGY-BASED INFORMATION EXTRACTION

Information Extraction is a key NLP technology to introduce complementary information and knowledge into a document. The general goal of information extraction is to discover structured information from unstructured or semi-structured text .The term "Ontology-based Information Extraction" has been conceived only a few years ago and has recently emerged as a subfield of IE. Originally, ontology is a philosophical term which is defined as "the description of the objective existence of the world, namely, the existence". Ontology is formalized clearly to the shared conceptual model. Ontology gives the basic terms of the vocabulary and the relationships, to capture the relevant domain knowledge, and propose a common understanding of the field to identify the common vocabulary, and give a clear formal definition. OBIE is different from traditional IE because it finds type of extracted entity by linking it to its semantic description in the formal ontology. The task of OBIE has received a specific attention in the last few years with many publications that describe systems. For (Wimalasuriya and Dou, 2010), the key characteristics of OBIE systems are:

• **Process natural language text documents:** The inputs of OBIE systems are limited to unstructured or semi-structured documents. Systems that use images, diagrams or videos as input cannot thus be categorized as OBIE systems. Categorizing systems that extract information from PDF documents is more problematic. For instance, Oro and Ruffolo have developed a system that processes PDF documents and presents the output in the form of an ontology [2]. On the first glance, this looks like a typical OBIE system. However, this system makes extensive use of the spatial relationships and images of the PDF documents. Oro and Ruffolo also recognize that their system is significantly different from "NLP-Oriented" systems. They state that systems that extract information from unstructured text can be categorized into two groups as NLP-Oriented and PDF-oriented [2]. In our view, only the systems of the former category can be classified as OBIE systems . While systems belonging to the latter category are important and interesting because of the widespread use of the PDF format, they should be seen as a separate type of systems.

• **Present the output using anthologies:** Y. Li and K. Bontcheva [3] identify the use of a formal ontology as one of the system inputs and as the target output as an important characteristic that distinguishes OBIE systems from IE systems. While this statement holds true for most OBIE systems, there are some OBIE systems that construct the ontology to be used through the information extraction process itself instead of treating it as an input. Since constructing ontology in this manner should not disqualify a system from being an OBIE system, we believe that it is prudent to remove the requirement to have ontology as an input for the system. However, the requirement to represent the output using ontologies appears to be reasonable.

2.1 Information Extraction process guided by ontology*:* We believe that "guide" is a suitable word to describe the interaction between the ontology and the information extraction process in an OBIE system: in all OBIE systems, the information extraction process is guided by the ontology to extract things such as classes, properties and instances. This Means that no new information extraction method is invented but an existing method is oriented to identify the components of ontology. An important question here is whether the information extractors, which are the components of an information extraction system that extract different ontological concepts, should be considered a part of the ontology or not [4, 5].To the best of our knowledge, this is an open question. When we see say that the information extraction process of an OBIE system is guided by ontology, we consciously accommodate both possibilities: the information extractors may be either part of ontology or may lie outside the ontology.

Although ontology-based information extraction is a relatively new field of study, it is generally agreed that it has a lot of potential. The following points highlight this potential.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

1. **Automatically processing the information contained in natural language text:** A large fraction of the information contained in the World Wide Web takes the form of natural language text. Further, according to some estimates around 80% of the data of a typical corporation are in natural language [6]. Ontology-based information extraction systems as well as general information extraction systems are necessary to process this information automatically. This is essential because manually processing them is becoming increasingly difficult due to their increasing volumes.

2. **Creating semantic contents for the Semantic Web:** Although the success of the Semantic Web relies heavily on the existence of semantic contents that can be processed by software agents, the creation of such contents has been quite slow. Popov et al. state that it is hard to imagine that such contents would be manually created given the size of the Web and assert that the automatic metadata generation would be the snowball to unleash an avalanche of metadata through the Web, making the Semantic Web come true [7]. Ontology-based information extraction provides such an automatic mechanism to generate semantic contents (metadata) by converting the information contained in existing web pages into ontologies. This process is also known as the semantic annotation of web pages.

3. **Improving the quality of ontologies:** As pointed out by Kietz et al. [8] and Maynard et al. [5] among others, ontology-based information extraction can also be used to evaluate the quality of ontology. If a given domain ontology can be successfully used by an OBIE system to extract the semantic contents from a set of documents related to that domain, it can be deduced that the ontology is a good representation of the domain. Further, the weaknesses of the ontology can be identified by analyzing the types of semantic content it has failed to extract.

III. AUTOMATIC TEXT SUMMARIZATION

Automatic text summarization [20] is the process of reducing a text document with a computer program in order to create a summary that retains the most important points of the original document. As the problem of information overload has grown, and as the quantity of data has increased, so has interest in automatic summarization. Technologies that can make a coherent summary take into account variables such as length, writing style and syntax. An example of the use of summarization technology is search engines such as Google.

3.1 Characteristic of Text Summarization

- **Frequency-based:** frequency of words or key terms, proximity, and location within the text.
- **Knowledge-based:** generally depend on rich knowledge sources to interpret the conceptual structure of the text. Knowledge-based approaches are usually very knowledge-intensive and domain specific.
- **Discourse-based:** theories of text cohesion and coherence of sentence in how much they push the limits of text understanding and the complexity as well as automation of that processing.

3.2 Level of Processing

Summarization can be characterized as level of processing at the surface, entity, or discourse level.

i. **Surface level approaches:** represent information in terms of shallow feature to yield a salience function to extract the most important information. These features include: *Term frequency* (statistics provide a salient terms, the important sentences are the ones that contain words that occur frequently), *Location* (position in text, position in paragraph such as the lead method, the title base method), *Cue words and phrases* (e.g., cue: “in summary”,” in conclusion”, “the paper describes”, “our investigation”; or emphasizes: “significantly”, “important”, “hardly”, “impossible”)

ii. **Entity level approaches:** try to model text entities (simple word, compound nouns, name entities, etc) for creating an internal representation of text. These approaches include: *Similarity* (similarity between the sentence and the rest of the document, similarity between the sentence and the title of the document),*Proximity* (distance between text units), *Cohesion* (terms of connectivity such as word co-occurrence, local salience, co reference),*Logical relations* (such as agreement, contradiction, entailment, and consistency) and *meaning representation-based relations* (establish relation between entities in text)

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

iii. **Discourse level approaches:** model the global structure of the text and its relation to communicative goals that include: *Format of the document* (e.g., hypertext markup, document outlines),*Threads of topics* (topic signature, can be displayed in text) and *Rhetorical structure of the text* (argumentation or narrative structure)

3.3 Classification of text summarization

Depending on the number of documents a summarizer takes as input, there are two types of summarizer: Single Document Text Summarizer and Multi Document Text Summarizer

1) Single document summarizer: Summary is generated from a single document. Generated summary is Coherent i.e. no redundant information is available. Text Teaser [9] is an application that extracts the most important sentences of an article. The purpose of the API is to provide a preview of what the article is all about. The API accepts the text of the article, and it will return the summary by extracting most important Sentences in it.
2) Multi Document Text Summarizer: Summary is generated from multiple text documents. For example, given a collection of news articles, a multi document text summarizer is able to create a concise overview of the important events. One big issue of this type of summarizer is to reduce the redundant information. In the case of multi document summarization of articles about the same event, source articles can contain both repetitions and contradictions. Extracting all the similar sentences would produce a verbose and repetitive summary, while extracting only some of the similar sentences would produce a summary biased towards some sources. MultiGen [10] uses a comparison of extracted similar sentences to select the appropriate phrases to include in the summary and reformulates them as new text.

3.4. Techniques of Text Summarization

Text Summarization techniques can be broadly classified as Extractive and Abstractive.

1) **Extractive Summarization:** This technique aims to pick out most relevant sentences from different documents and maintain low redundancy of information in summary. In extractive-based summarization methods, picking out most important document regions like phrases, sentences, paragraphs etc. In this technique first, assigns a score to each sentence and then gives ranks to the sentences according to their scores. The sentence with highest score will get the top rank. The score for a sentence is calculated by using statistical features including sentence position, cue words, term frequency, document frequency, topic signature, etc. The highest ranked sentences from different documents can then be grouped into different cluster in terms of their Similarity with other ranked sentences. Finally select representative sentences from each cluster to generate a Simple and non redundant summary

2) Abstractive Summarization: This method involves natural language understanding tool to generate Summary from document(s). An abstract summary ([11], [12]) may contain words or phrases which mayn't be exist in the original document(s). The process of abstractive summarization is complex because it cannot be formulated mathematically or logically. The quality of abstractive summary depends on deep linguistic skills.

IV. RELATED WORK

There are several works that have been proposed and implemented in the past by various researchers about information extraction and text summarization. Some of the most important works are cited below.

4.1 Text information extraction based on OWL ontologies Hongsheng wang, Lu Yuan, Hong Shao presented in IEEE 2008 an OWL ontology-based text information extraction system and has made a clear description of every module contained in the system and the cooperative relationships among them. This use two algorithm for implementation. One is semantic information extraction and other is semantic information re-recognizing. Experiment results show that, the two algorithms are effective and accurate, especially when the domain ontologies are well-defined. System has good portability.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

4.2 Web information extraction based on news domain ontology theory Junfang Shi, LiLiu proposed in IEEE 2010 a web information extraction method based on news ontology. Accurate and Interested information was identified by using news domain ontology. With the help of page preprocessing, XPath and page conversion technology. Testing from news site shows that the approach proposed doesn't rely on the page structure and it increased the recall and precision of information extraction.

4.3 Ontology-based information extraction system in E-commerce websites Yang Xiudan, IEEE 2011 used the concept of ontology to analyze the structure and content of the website, in order to extract the information based on ontology from the e-commerce website for the users with the help of ontology model.. The paper makes an experiment test of the test tool GATE to extract from website and evaluate the results objectively. This Increase the performance of the system.

4.4 Ontology-based information extraction from twitter Kamel Nebhi[13] proposed an approach for ontology-based information extraction from Twitter. This system provides an integrated disambiguation module based on popularity score and syntax-based similarity. Result shows that, the system performed significantly better using disambiguation process.

4.5 Abstraction based domain ontology extraction for Idea Creation Delin Jing, Hongji Yang, Yingchun Tian Proposed in IEEE 2013, an abstraction method to support one of the essential parts in creative idea creation- domain ontology extraction. Abstraction techniques are classified, selected and integrated while elements of domain ontology are defined including concepts and relations

4.6 Multi-Document Summarization

Recently year most researchers for automatic text summarization have transferred their efforts from single documents to multiple documents. There are many techniques that people use for multi document summarization from the part to present.

Kathleen R. McKeown. et al.(2001) [14] presented MultiGen and DEMS for Columbia multi-document summarization system built on the observation that depending on the intended purpose of the summary and on the types of document summarized. This technique focused on the summarization of sets of documents that all describe the same event or news. They used an enhanced version of MultiGen to summarize the document. They used alternative system DEMS (Dissimilarity Engine for Multi document Summarization) for biographical documents. While processing stage, the
Input articles are transformed into a uniform XML format. After that, the router components of the system determined the type of each input document set and direct the input texts to the summarizers.

Jun ichi Fukumoto (2004) [15] proposed a summarization system which automatically classified type of document set and summarized a document set with its appropriate summarization mechanism. This system classified a document set into three types, a series of events, a set of the same events and related events, by using information of high frequency nouns and named entity. The unnecessary parts are deleted after summarized each document and generated multi-document summary. They used single document summarization mechanism for each document of a document set and removed similar parts between summarized documents for generation of a target summary. They applied a TF/IDF based sentence extraction for single document summarization and used of single document summarization for multi-document summarization. Their mechanism of document set classification does not work well in the evaluation because their current implementation has some system bugs in Classification mechanism.

Junlin Zhang., Le Sun. and Quan Zhou (2005) [16] proposed a new approach for multi-document summarization under the Hub/Authority framework. This approach combines the text content with some cues such as cue phrase, sentence length and first sentence and explores the sub-topics in the multi-documents by bringing the features of these sub-topics into graph-based sentence ranking algorithms. They improved two main different points from the old graph-based method: (1) combines the text content with some cues such as cue phrase, sentence length and first sentence. (2) Explores the sub-topics in the multi-documents and brings the features of these sub-topics into graph-based sentence

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

ranking algorithms. Also, they used the Markov Models to order the subtopics that the final summarization should contain and output the text summarization according to the sentence ranking score of all sentences within one sub-topics as user' requirement. The result of evaluation this method on DUC 2004 data proved that the idea of combining the surface and content features under the Hub/Authority framework is an effective graph-ranking schema in multi-document generic text summarization.

Yan-Min, Xiao-Long and Bing-Quan. (2005) [17] used lexical chain for multi document summarization in Chinese document based on How net knowledge database. In their algorithm start from pre-process the text: remove redundant similarities and remain differences in information content among multiple documents, then constructs lexical chains and identifies strong chains. Then significant Sentences are extracted from each document and ordered sentences, then recognized and removed redundant information. Finally, the summary is generated in chronological order, and the anaphora resolution technology is applied to improve the fluency of the summary.

De-Xi Liu et al. (2006) [18] proposed a cluster-based method for Chinese multi-document summarization which two steps: sentence clustering and sentence selection. Sentence clustering has two strategies to determine the number of clusters automatically: one strategy makes full use of the summary length fixed by the user while the other one is stability based, which can infer the optimal cluster number automatically. From each Sentence selection, select one sentence to represent the topic by weight the sentence based on the terms included in the sentence. Judith D.

Schlesinger et al. (2008) [19] proposed CLASSY system. This text summarization is an automatic multi-lingual document summarization. It used CLASSY system architecture to summarize document. CLASSY (Clustering, Linguistics, And Statistics for Summarization Yield) is an automatic, extract generating, summarization system that uses linguistic trimming and statistical methods to generate generic or topic (/query)-driven summaries for single documents or clusters of documents. CLASSY used trimming rules to shorten sentences, identify sentences, select sentences and organize the selected sentences for the final summary. The main approach of this research is to generate a multi-lingual summarization document based on summaries document from Machine Translate (MT) document. CLASSY architecture consists of five steps: document preparation, sentence trimming, sentence scoring, redundancy reduction, and sentence ordering. The CLASSY will generate very good summaries when using signature term that Computed from English document and machine translated version of Arabic document. The quality of machine translation affected directly to the quality of this summary architecture.

V. CONCLUSION

The large amount of news published every day exceeds the human reading capacity and motivated pioneer to the Informaion Extraction and multi-document summarization efforts. Despite the large body of work on summarization in the news domain, most works do not explore the fact that news documents describe events. In this paper, we use Ontology Based Information Extraction method guides the system that how to pull out efficient and relevant information using the Information Extraction methods. In future there are several directions with OBIE like improving the efficiency of IE system to improve the precision and recall. It has been observed that in the context of multi-document summarization of news articles, extraction may be inappropriate because it may produce summaries which are overly verbose or biased towards some sources, so in this paper we used Abstraction-based summarization to generate efficient multi-document summarization from multiple news articles. Our multi-document summarization framework improves the summarization of documents containing detectable events, achieving state-of-the-art results in multi-document summarization.

REFERENCES

[1] Daya C. Wimalasuriya, Dejing Dou "Ontology-Based Information Extraction: *An Introduction and Current Approaches" By the Journal of Information Science, XX (X)* , pp. 1-20,2009.

[2] E. Oro and M. Ruffolo, Towards a system for ontology-based information extraction from PDF documents. In: R. Meersman and Z. Tari (eds.), *Proceedings of the OTM 2008 Confederated International Conferences, CoopIS, DOA, GADA, IS, and ODBASE 2008. Part II on On the Move to Meaningful Internet Systems*, (Springer-Verlag, Berlin, 2008).

[3] Y. Li and K. Bontcheva, Hierarchical, perceptron-like learning for ontology-based information extraction. In: *Proceedings of the 16th International Conference on World Wide Web*, (ACM, New York, 2007).

[4] F. Wu and D. S. Weld, Autonomously semantifying wikipedia. In: *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management*, (ACM, New York, 2007).

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

[5] D. Maynard, W. Peters, and Y. Li, Metrics for evaluation of ontology-based information extraction. In: *Proceedings of the WWW 2006 Workshop on Evaluation of Ontologies for the Web*, (ACM, New York, 2006).
[6] J. J. Ritsko and D. I. Seidman, Preface, *IBM Systems Journal* 43(3) (2004) 449-450.
[7] B. Popov, A. Kiryakov, D. Ognyanoff, D. Manov and A. Kirilov, KIM - a semantic platform for information extraction and retrieval, *Journal of Natural Language Engineering* 10(3-4) (2004) 375-392.
[8] J. Kietz, A. Maedche, and R. Volz, A method for semi-automatic ontology acquisition from a corporate intranet. In: *Proceedings of the EKAW'00 Workshop on Ontologies and Text*, (Springer, Berlin, 2000).
[9] Purabi Choudhury A Survey of News Article Clustering and Summarization (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 2849-2851
[10] Kathleen R. McKeown, Regina Barzilay, David Evans, Vasileios Hatzivassiloglou, Simone Teufel, M. Yen Kan, and Barry Schiffman. 2001. Columbia multi-document summarization:Approach and evaluation. In Donna Harman and Daniel Marcu, editors, Proceedings of the First Document Under-standing Conference (DUC '01), NewOrleans, Louisiana, USA. 1-21.
[11] Inderjeet Mani, (1999) "Advances in Automated Text Summarization", MIT Press.
[12] Sun J, Shen D, (2005) "Web-page Summarization using Clickthrough data", In proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development inInformation Retrieval.
[13] Kamel Nebhi ,"Ontology-Based Information Extraction from Twitter "Proceedings of the Workshop on Information Extraction and Entity Analytics on Social Media Data, pages 17–22, COLING 2012, Mumbai, December 2012.
[14] Kathleen R. McKeown, Regina Barzilay, David Evans,Vasileios Hatzivassiloglou,Simone Teufel, M. Yen Kan, and Barry Schiffman. "*Columbia multi-document summarization: Approach and evaluation". In Donna Harman and Daniel Marcu, editors, Proceedings of the FirstDocument Under-standing Conference (DUC '01), New Orleans, Louisiana, USA. PP 1-21,2001.*
[15] Jun ichi Fukumoto "Multi-Summarization using document set type classification". *In Proceedings of NTCIR- 4, Tokyo, 2004.*
[16] Junlin Zhang., Le Sun. and Quan Zhou. "A cues-based hub-authority approach for multi-document summarization". *In Proceeding of NLP-KE'05*,IEEE, 642-645,2005.
[17] Yan-Min C., Xiao-Long W. and Bing-Quan "Multi-Document Summarization Based on Lexical Chains". *In Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou*. IEEE, pp 1937-1942,2005.
[18] De-Xi Liu, Yan-Xiang Hi., Dong-Hong Ji. and Hua Yang. "A Novel Chinese Multi-Document Summarization Using Clustering Based Sentence Extraction". *In Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian*, pp 2592-2596,2006.
[19] Judith D. Schlesinger, Dianne P. O'Leary, and John M. Conroy."Arabic/English Multi-document Summarization with CLASSY—the Past and the Future. In Computational Linguistics and Intelligent Text Processing", *9th International Conference*, Haifa, Israel, Proceedings. Springer. PP 568–581.February 2008.
[20] Automatic summarization From Wikipedia, the free encyclopedia

BIOGRAPHY

V. Swathi Lakshmi received her Bachelor of Technology in Information technology from Manakula Vinayagar Institute of Technology College, Pondicherry in 2014 and currently pursing Master of Technology in Computer Science Engineering (CSE) from Pondicherry Engineering College, Pondicherry, India. Her area of interest includes Natural language processing and database management system.

Dr. S. Lakshmana Pandian received his Bachelor of Engineering in Electrical and Electronics Engineering, from Government College of Engineering, Tirunelveli in 1993, and Master of Engineering in 1998 in Computer Science and Engineering, from Government College of Engineering, Tirunelveli and Ph.D. degree in the Department of CSE from Anna University, Chennai in 2011. He is working as Associate Professors in the department of CSE in Pondicherry Engineering College, India. He has more than 7 publications in International Journals. He has presented more than 8 papers in International conferences. He has guided both UG and PG candidate's projects. His areas of interest include Language Technology (Natural language processing, Compilers, Automata theory and computations) and Embedded Systems.