Optimization Techniques for Web Content Mining- A Survey

1M. LAVANYA AND 2M. USHA RANI

1 Department of Master of Computer Applications
Sree Vidyaniikethan Engineering College (Autonomous), A.Rangampet, Tirupati, Andhra Pradesh, INDIA.
2 Department of Computer Science, Sri Padmavati Mahila Visvavidyalayam, (SPMVV Woman’s’ University), Tirupati, Andhra Pradesh, INDIA.
lavanya4 79@rediffmail.com, musharohan@yahoo.com

ABSTRACT

Web Mining is the method used to creep through various web resources to automatically determine and dig out information from Web documents and services. Web mining can assist marketing prototypes and adapt market to bring good products and services to right customers. We have studied various Sources to access data from Deep Web Pages, Web pages, etc. In this paper we discuss many optimization techniques in promotion of research in web content mining for accomplishment of required data from huge web resources. We have a taxonomy of web content mining how the data can be retrieved using various techniques from different web resources available. As per the user required data which is available in various chunks of deep web page or web pages have been retrieved by using different automatic wrappers irrespective of locations in web pages or deep web pages. Many other types of information are also making use of to improve the performances of the wrappers. CTVS deals with the method of combining tags. We have studied a novel and more useful process to extract chunks from web pages that contain a set of horizontal and layered data records. We have discussed an optimized technique about required relevant data extraction from irrelevant data regions like navigational bars, images, etc.

Keywords: Web resources; Web content Mining, Wrappers, Deep web Pages, Dynamic content.

1. INTRODUCTION

There are various causes for the come out of web mining [31]. The volume of the web is very huge on the orders of Zeta bytes and it still fabricating rapidly. Many Enterprises, users or welfare organizations offer their free information through web. Also, the content of the web pages are much more complex than any other text documents. Due to insufficient standard structure, they contain more typical style than standardized format [27]. Web Mining is the method used to creep through various web resources to automatically determine and dig out information from Web documents and services [8]. Web mining can assist marketing prototypes and adapt market to bring good products and services to right customers. It can help in making resolutions in client relationship management and also improve quality of mining. Nonetheless, a huge number of documents in this web are not organized in a meticulous order. Another cause is that the web is a highly dynamic information source [27]. As a Consequence there is rapid increase in World Wide Web and information in it also updated frequently. Educational Societies, shops and bazaars, organizational advertisements and Web service centres update their pages regularly. Web users may have different backgrounds, interests and usage purposes. It is also stated that only a small portion of information in the web is relevant or useful. Any web user can be interested in only small portion of the web.

The main issues lead to a research for successful detection and utilize of resources in World Wide Web, which also leads to web mining. The complete retrieval of information from web results leads to new research area called Web Mining. There are two different approaches were proposed for defining Web mining [5]. The first approach is a ‘process-centric view’, which defines Web mining as a sequence of ordered tasks. Second one is a ‘data-centric view’, which defines web mining with respect to the types of web data that was used in the mining process [14]. Web involves three types of data [19]; the actual data on the WWW, the web log data obtained from the users who browsed the web pages and the web structure data. Thus, the web mining should focus on three important dimensions; web structure mining, web content mining and web usage mining [17].

Web as a Database Web as a source containing some semantic information on top of the semi structured Web lets users query the Web as they would a database [22]. Researchers can use content and hyperlink mining approaches in which XML represents the semantics to build such a multilayered Web. As huge amount of information which is available in dynamic source is complicated and rapidly increasing all over the world. Web is rich source for data mining principles, of which a vital role is Web content mining. The application of web content mining techniques to extract relevant data from Web resources. The process of creating a data extraction program for e.g. generating XML based on relevant data taken from HTML is usually referred to as wrapper generation [26]. Different from data extraction, web navigation is a more process-oriented step-by-step task and needs different treatment. Web pages often contain advertisements, image-maps, logos, search boxes, navigational links, related links, footers and headers, and copyright information along with the primary content [30]. These blocks are not relevant to the main content of the page. These items are required for web site owners.
but they will hamper the web data mining and decrease performance of the search engines [7]. Recently, many research works are going on for the extraction of information from deep web in better way and many solutions have been proposed by the researchers. Advently researches have proposed a type called as Wrapper generation [5] where in this method, locate rules that are intended to extract the information from web pages. In this many optimization techniques for retrieval of data is considered as wrapping of data from different forms to one form to make available the user with more ocular and understandable appearance of the query results. Data are of varied nature and contain various formats, layouts, and presentation. To browse these data, specialized search engines (known as meta search engine) such as Google Scholar are developed which can locate, understand, and process this particular data before presenting them for user viewing [4]. Before a meta search engine can search for data, it needs to locate the data (also known as data records) from other search engines, extract them, filter out the irrelevant data, and rank them accordingly. The tools to extract this data from various search engines are termed wrapper Jer L. H. (2010). However, extraction of data from search engines is difficult due to the following reasons 1) HTML language is confusing and not regularly presented 2) Data in search engines are semi structured, they contain various forms and layouts 3) Search engine result pages may also contain other irrelevant data (known as data regions) in addition to the relevant data. Manual wrappers are also fault prone, and not easy to update because developer needs to make required updating to the wrapper as required by the web data base for a particular company [4]. Hence the developer has to alternative to the development of supervised and semi supervised wrappers which can extract data records automatically. Before these wrappers came into existence, previously they use human labeling of the data so that the wrappers can extract the relevant region of the data, which is of necessary to the user. Semi supervised wrappers can guess the extraction rule once data labeling and data extraction are carried out. However, these wrappers also have issues, like they require human labeling and intervention before they begin their operation. To overcome this issue, fully automatic wrappers were developed to extract data records without any human involvement Arvind A. and Hector G-M (2003), Valter C., Giansalvatore M., and Paolo M. (2001). The main issue in extraction of data from web is to remove the irrelevant and the redundant data available from deep web pages which is not directly extracted through search engines [1]. The extraneous data may consist of navigation panel, popup windows and advertisement. Web programming languages which are programming independent language are used to extract data from the deep web pages. Various optimization techniques provide a eminence in examining web data extraction. With advent growth in usage of web data, the assets of data regarding many applications and enterprises has become online [6]. With many techniques are available in searching of data from deep web and have many limitations. Many explorations for searching keywords is found than searching through browsers which returns large amount of data than what user is required where as browsing is not adoptable for finding particular keywords in deep web pages. Wrapper is program that executes the mapping. A main aim of wrapper generation tool is to produce wrappers that are extremely precise and strong while acquiring less work from wrapper programmer. During studies it was observed that significant fraction of web content lies outside the PIW [5]. A large amount of the web and data is hidden behind search forms (large number of data is extracted only through HTML forms). These types of web are called the hidden web or deep web [6].

2. SOURCES TO ACCESS DATA FROM DIFFERENT WEB RESOURCES

Definition (Deep Web, Hidden Web, Invisible Web) the content of the web which are directly not accessible through general search engines, hyperlinks etc. The main sources of deep web pages are completeplanet.com, BrightPlanet.com, Thousands of deep web databases [30]. Deep web is the web that is dynamically generated from data sources such as databases or file systems. Web pages often contain advertisements, image-maps, logos, search boxes, navigational links, related links, footers and headers, and copyright information along with the primary content [31]. These blocks are not relevant to the main content of the page. These items are required for web site owners but they will hamper the web data mining and decrease performance of the search engines. The amount of data in deep web exceeds by far that of the surface web. This calls for deep web crawlers to excavate the data so that they can be reused, indexed, and searched upon in an integrated environment [24]. Crawling deep web [24, 25, 26, 27, 28] is the process of collecting hidden data by issuing queries through various search interfaces including HTML forms, web services and programmable web APIs. Crawling deep web data sources is important for several reasons, such as indexing deep web data sources, or backing up data. In this area, two research challenges have been substantially addressed. One is learning and understanding the interface and the returning result so that query submission [29] and data extraction can be automated.

Categories of Web Resource Data Types

In this sub section we have studied various data types available in deep web from which the content is to be extracted. The data in the web resources is mainly deep web pages that are obtainable for the use of users. Content data from deep consists of free text, semi-structured data like HTML pages and more structured data like automatically generated HTML pages, XML files or data in tables related to web content[14]. Textual,
image, audio and video data types falls into this category. The most common deep web content data is HTML pages in the web. HTML (Hypertext Markup Language) is designed to determine the logical organizations of documents with hypertext extensions. The HTML elements can be examined in two categories: those that define how the body of the document is to be displayed by the browser, and those that define the information about the document, such as the title or relationships to other documents.

![Deep Web Page](image)

**Fig.1 Taxonomy of various data types in Deep Web Page**

3. **TAXONOMY FOR WEB CONTENT MINING**

Taxonomy for Web content mining can be broadly divided into following categories depending on kinds of data to be mined on Web resources:

- **Web Content Mining**
  - Structured
    - Web crawler
    - Wrapper Generation
      - DOM tree Wrappers
      - Visual Wrappers
  - Page Content Mining
  - Semi Structured
- **Web Structure Mining**
- **Web Usage Mining**

Web content mining is the activity of extracting knowledge from the content of documents on World Wide Web like mining the content of html files. Web document text mining, resource discovery based on concepts indexing or agent-based technology fall in this category. Web structure mining is the process of extracting knowledge from the link structure of the World Wide Web. Finally, web usage mining, also known as Web Log Mining, is the process of discovering interesting patterns from web access logs on servers.

3.1 **Web Content Mining**

Web Content Mining is the process of extracting useful information from the contents of Web documents. Content data is the collection of information designed to be conveyed to the users. It may consist of text, images, audio, video, or structured records such as lists and tables. Text mining and its application to Web content has been the most widely studied forms of web content mining. Some of the research issues including the text mining are; topic discovery, extracting association patterns, clustering of web documents and classification of Web Pages. Research activities in these fields also involve using techniques from other disciplines such as Information Retrieval (IR) and Natural Language Processing (NLP). There is also significant body of work exist for discovering knowledge from images in the fields of image processing and computer vision. The application of these techniques to Web content mining has not been very effective yet. The research on unstructured documents like pure text files also falls into web content mining. Unstructured documents are free texts in www such as newspaper pages. Most of the researches in this area uses bag of words in order to represent unstructured documents [20]. Web content mining, which mainly concentrates on the information of single document, web structure mining tries to discover the link structures of the hyperlinks between documents. A lot of the knowledge in the Web is inside documents, i.e., in their content/data/documents. The discovery process of this useful information from that content is called Web Content Mining. Web Content Mining is useful to recognize the data presented in deep Web pages documents, Classify Web Documents, stumble on...
3.1 Document Classification

Classification’s is the main source for machine learning, pattern recognition, and text analysis. The main theme behind how to categorize pages depends on supervised and unsupervised methods [22]. Previously works in document classification applied text-mining techniques to Web data directly. With new trend in research showed that exploiting the Web graph structure and semi structured content in the form of HTML tags improved results.

3.2 Web Structure Mining

A web graph consists of web pages as nodes and hyperlinks as edges, which represents the connection between two web pages. Web structure mining can be defined as a activity of determining structured information from the web. The goal of web structure mining is to extract structural information about the web site and its web pages. Making use of topology information of hyperlinks, web structure mining can categorize the Web pages and access results such as the similarity and relationship between different Web sites.

Web Structure Mining can be divided into two categories based on the type of structure data used. The structural data for Web structure mining is the link information and document structure. Given a collection of web pages and topology, interesting facts related to page connectivity can be discovered.

3.3 Web Usage Mining

Web Usage Mining is the process of applying data mining techniques to discover interesting patterns from Web usage data. Web usage mining provides better understanding for serving the needs of Web-based applications [20]. Web Usage data keeps information about the identity or origin of Web users with their browsing behaviour in a web domain. Web usage mining itself can be divided into subcategories based on the type of web usage data used. As shown in the figure2 we have an taxonomy of how the data can be retrieved using various techniques.

There are two types of automatic wrappers available currently 1) Document Object Model (DOM) Tree Based Wrappers 2) Visual Based Wrappers. Wrapper Incorporating Set of Heuristic Techniques (WISH) Jer L. H., Eugene S., Simon E. (2010)[2] uses frequency measures to match the tree structures of data records. WISH works in a time complexity of O(n) and is able to match tree structures of data records containing iterative and disjunctive data. However, tree matching algorithm of WISH is not able to match data records with dissimilar tree structures.

![Fig.2 Taxonomy For Web Content Mining](image)

4. METHODS FOR EXTRACTING DATA IN STRUCTURED WEB

There are three basic methods for extracting the data from structured web pages as follows:

i. web crawler
ii. Wrapper Generation
iii. Page content mining

4.1 Web Crawler

Web Crawler is a program that searches throughout web pages in a sophisticated manner. This program execution is a process called as crawling. This type of web crawling is used in search engines. There are two
categories in Web Crawler specified as an External and Internal Web Crawler [6]. External crawler orders
unknown websites and initiates an internal crawl on the first page of the website only. Internal crawler will only
crawl through internal pages of the websites returned by the external crawler. The websites spawned by the
internal crawlers are more trustworthy. Web Crawlers are a core for search engines, and specifications on their
algorithms and architecture are reserved as company undisclosed. Web crawlers are improvable to position
recognizing information in the user negotiator field. It is significant for Web crawlers to recognize themselves
when ever required administrators of Web databases can consult owner if required. In some extraneous cases,
crawlers may be capturing in a crawler trap and sometimes they may have flood of requests to web servers and
the owners have to stop crawling. Recognition is handful benefit for administrators who are really interested in
accessing the web pages to be indexed by a particular ontology search engine. Requesting the data in the basis
requires producing a relevant URL for the document that contains the solutions. Building the URL is quite
different balanced to translation from a set of standard operators to another set of operators native to a database-
like source? The solutions are optimized typed objects in some standard representation. Rather, the solutions to
be extracted from the semi-structured data stored in HTML documents, and as per the query these documents
may contain data that is irrelevant. Due to instability in structure of documents which may affect the rigorous
process.

4.2 Wrapper Generations
We search for the information on web, wrapper generation process on web pages that are retrieved through
search engines. Once the relevant URL for the document is provided then searching the data contains
solution.W4F divides wrapper development process in three phases: first, the user describes how to access the
document, second he describes what pieces of data to extract and third he declares what target structure to use
for storing the data extracted. Once users have complete knowledge of what type of data is required then we can
follow basic primitives for wrapper generation can be differentiated like
(1) Manual wrapper generation where the system requests user in programming a specific wrapper which is not
generalized.
(2) Wrapper induction, where the user provides templates and samples for extraction patterns, and the system
provide a suitable wrapper using machine learning techniques.
(3) semi-automatic interactive wrapper generation, where the wrapper designer for both sampled data and
wrapper programming for system interactive method involving generalized things to be included along with
visual programming techniques.

4.3 Visual Wrapper Generation
As deep Web pages from the same Web database share the same visual template, once the data records and data
items on a deep Web page have been extracted, we can use these extracted data records and data items to
generate the extraction wrapper for the Web database so that new deep Web pages from the same Web database
can be processed using the wrappers quickly without reapplying the entire extraction process. Our wrappers
include data record wrapper and data item wrapper. They are the programs that do data record extraction and
data item extraction with a set of parameter obtained from sample pages. For each Web database, we use a
normal deep Web page containing the maximum number of data records to generate the wrappers. The wrappers
of previous works mainly depend on the structures or the locations of the data records and data items in the tag
tree, such as tag path. In contrast, we mainly use the visual information to generate our wrappers. Note that
some other kinds of information are also utilized to enhance the performances of the wrappers, such as the data
types of the data items and the frequent symbols appearing in the data items.

5. OPTIMIZATION TECHNIQUES FOR RETRIEVAL OF DATA FROM WEB
5.1 Handling user queries:
To compare resemblance between data records, Weifeng Sue et al. (2012) proposed an approach called
Combining Tag and Value Similarity for Data Extraction and Alignment (CTVS). This method deals with the
process of combining tags [1]. The user had provided manually queries. The output of CTVS is based on the
manual user query which is given as an input for the construction of tag tree. The outcome of tag tree emerges
into the data region identification. This data region is divided into sub- data regions that penetrates through the
record chunks in which specific records are recognize before combining data. Then it pierces into the query
result segment identifications. Finally, the query result records are recognized.

5.2 Mining of data from horizontal and layered data records of web pages which is a combination of
VSAP (Visual Structure based Analysis of web Pages) and VCED(Visual Clue based Extraction of web Data)
We have studied a novel and more useful method to extract chunks from web pages that contain a set of
horizontal or layered data records. This technique is implemented in two phases [11]:
i) Recognizing and accessing chunks based on location of data region/ data records/ data items on the screen at
which the tags are rendered, data from web pages is accessed using an algorithm here after called as VSAP
(Visual Structure based Analysis of web Pages[12]).
ii) Retrieved worthy chunk is recognized and accessed from flat and nested data records based on location of data regions using an algorithm called as VCED [13].

<table>
<thead>
<tr>
<th>Features of VSAP</th>
<th>Features of VCED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortitude of bounding data regions (rectangle).</td>
<td>i. Data record extractor.</td>
</tr>
<tr>
<td>Recognizing data regions</td>
<td>ii. Data record identifier of flat or nested.</td>
</tr>
<tr>
<td>Identifying largest rectangle region.</td>
<td></td>
</tr>
<tr>
<td>Container identifier.</td>
<td></td>
</tr>
<tr>
<td>Data region identifier (Filter).</td>
<td></td>
</tr>
</tbody>
</table>

The VSAP technique is based on three observations:
(a) A group of data records, that contains descriptions of a set of similar objects, is typically presented in a contiguous region of a page. (b) The area covered by a rectangle that bounds the data region is more than the area covered by rectangles bounding other regions, e.g., advertisements and links. (c) The height of a irrelevant data record within a collection of data records is less than the average height of relevant data records.

In this method a more effective method to identify and extract horizontal and layered data records from a given web page automatically is have been proposed. The method is called VCED. By Definition horizontal and layered data records are defined as below:
A horizontal data record is defined as a collection of data items that together represents a single meaningful entity. e.g., the product having single size, look, price etc.,
A layered data record is defined as one that provides multiple description of the same entity. E.g., the same type of products but different sizes, looks, prices etc.,

5.3 Optimization method for Deep web pages using Query Techniques

The data from the Deep Web cannot be accessed by Search engine and web crawlers directly. The only way to access the hidden database is through query interfaces and filling up number of HTML forms for a specific domain. A technique called HEET(Hidden Web QuEry Technique) for modeling and query the hidden web is presented. It technique for modeling of consecutive forms and introduce the concept of widespread form. Pages in the hidden web are dynamically extracted after submitting the query through different search forms. It is very much challenging for user to retrieve and analyze relevant as well as important information from the deep web automatically.

While browsing different types of Health web site, it was observed that the web sites are designed in same pattern and users need to fill up number of consecutive forms to retrieve the in formations. User may give any type of unexpected query in search engine text box. The process can be efficiently completed by using an automatic form querying system, but it is not an easy task to design this type of automated query processing method due to several challenges.
The user manually fills input values to web forms and extract data from the returned web pages.
- Automatic filling of forms
- Extraction of results
- Navigational complexity

Most consecutive forms have dependencies between the main and descendant form. While comparing the existing model and HEET, it is going select four options in child form rather than one option in child form and submits the query which takes less submissions and retrieval time. Modelling the consecutive forms into a single form for more results in a single submission of query form which saves the query submission time, execution time, result extraction time.

5.4 Vision-Based Deep Web Data Extraction For Web Document Clustering (VDEC)

Recently, web data extraction has become more demanding due to the difficulty and the assortment of web structures and representation. This is an predictable occurrence since the web resources has been so admired and there are now many types of web contents, including text, videos, images, speeches, multimedia, dynamic content, or flashes. Until now, a huge number of techniques have been proposed to address this problem, but all of them have inherent limitations because they are Web-page-programming-language dependent. In this paper, we have discussed an approach for detection and removal of noisy data to extract main content information and deep web clustering that is both fast and accurate [32].

Steps to be followed for VDEC
1. Provide input as the deep web pages, \( W_p \)
2. Fragment deep web page using \(<\text{div}>\) tag
3. Calculate Noise chunk removal value for each leaf nodes of deep web pages
   3.1 Compute surplus noise removal for all leaf nodes using Hyperlink Percentage and noise score.
      3.1.1 Compute Hyperlink word Percentage,
      3.1.2 Compute Noise score,
3.2 Compute duplicate noise removal for all leaf nodes using cosine similarity
4. Calculate sub-chunk weightage value $SC_w$ for each Noiseless chunk of deep Web pages
   4.1 Compute Title word relevancy for noiseless chunk
   4.2 Compute Keyword frequency based chunk importance
   4.3 Compute Position features based chunk importance
   4.4 Compute sub-chunk weightage
5. Compute main-chunk weightage value $M_i$, i.e. set of keywords
6. After computing the extraction of keywords for all deep web pages, set of
   Keywords were clustered using Fuzzy c-means clustering.

5.5 Optimized Technique for data region extraction
We studied that this wrapper uses two phases of extraction, global and local extraction. In the first phase, wrapper will identify the list of regions based on their visual boundary. A region is considered as potential region if its visual boundary is of acceptable size (e.g. more than 500 pixels). Potential region is defined as region which may contain the relevant data, search results. Once the list of potential regions is identified, our wrapper chooses the region which has the largest size (calculated based on its visual boundary), since search results usually have the largest visual boundary. When the largest region is identified, wrapper will use local extraction stage to remove the remaining irrelevant data within the region such as search identifiers. As search identifiers usually have different visual boundary compared to data records, they have used visual boundary of data records to remove these irrelevant data.[2]

5.5.1 To extract data from Global Region
A new proposed approach they have traversed the DOM Tree using BFS algorithm where nodes are considered either as HTML Tags or HTML Texts. Therefore, in this case, root is the $<HTML>$ tag in the HTML page input source. To identify probable data regions, they implemented searchRegion function which locates the list of regions based on the visual boundary of HTML Tags.

5.5.2 To extract data from Local Region
In Local Extraction, they checks for repetitive nodes in a particular level of a tree for a given system. These repetitive patterns are defined as group of data records. Groups of data records can be defined as a set of data records having similar parent HTML tag, containing repetitive sequence of HTML tags and are located in the same level of the DOM tree. This wrapper uses the Adaptive Search extraction technique to determine and label potential tree nodes that represent data records. Subtrees which store data records may be contained in potential tree nodes. The nodes in the same level of a tree are checked to determine their similarity (whether they have the same contents). If none of the nodes can satisfy this criterion, the search will go one level lower and perform the search again on all the lower level nodes.

CONCLUSION
In this paper we have discussed about various web resources available for web content mining. We have focused on the taxonomy of web content mining with various technologies included for retrieval of data from web. We have studied domain dependent approach have been described for retrieving the data behind a given template. The study exposes about several techniques involved in data extraction from the deep web pages. In VDEC, technique mainly focuses on performances of memory management considering the amount of time consumed has been significantly reduced. Here, automation takes part in a vital role in improving the performance. After analyzing, if a query result page has more than one data region that contains result records and the records in the different data regions are not similar to each other, then CTVS will select only one of the data regions and discard the others. A novel and effective method based on a combination of VSAP and VCED is proposed to mine the data region and data items from the horizontal and layered data records in a web page automatically. In HEET approach have been proposed for modelling the consecutive forms into a single form for more results in a single submission of query form which saves the query submission time, execution time, result extraction time. After analyzing all wrappers a new approach has been discussed where they have considered both relevant and irrelevant region of web page for web content mining.

REFERENCES
[1] Pui Leng Goh, Jer Lang Hong, Ee Xion Tan and Wei Wei Goh, “Region Based Data Extraction” School of Computing and IT, Taylor’s University, Malaysia
[22] Pranam Kolari And Anupam Joshi"Webmining:Research And Practice W E B E N G I N E E R I N G”, University Of Maryland, Baltimore County 1521-9615/04/$20.00 © 2004 Iee, Copublished By The Iee Cs And The Aip 50 Computing In Science & Engineering
[28] Ping Wu, Ji-Rong Wen, Huan Liu, Wei-Ying Ma,Query Selection Techniques for Efficient Crawling of Structured Web Sources, ICDE, 2006. pp 47-56.


AUTHOR BIOGRAPHY

Ms. M. Lavanya obtained Bachelor’s degree in Sciences (Computer Science) from S.V.University, Tirupathi. Then she obtained her Master’s degree in Computer Applications from S.V.University. She is working as Assistant Professor at the Department of Master of Computer Applications at Sree Vidyanikethan Engineering College, A.Rangampet, Tirupati. She is pursuing her Ph.D. in Computer Science in the area of Data Warehousing and Data Mining. She is in teaching since 2003. She presented many papers at National and International Conferences and published articles in National & International journals.

Dr. M. Usha Rani is an Associate Professor in the Department of Computer Science and HOD for MCA, Sri Padmavati Mahila Viswavidyalayam (SPMVV Woman’s University), Tirupati. She did her Ph.D. in Computer Science in the area of Artificial Intelligence and Expert Systems. She is in teaching since 1992. She presented many papers at National and International Conferences and published articles in national & international journals. She also has written 4 books like Data Mining - Applications: Opportunities and Challenges, Superficial Overview of Data Mining Tools, Data Warehousing & Data Mining and Intelligent Systems & Communications. She is guiding M.Phil. and Ph.D. in the areas like Artificial Intelligence, Data Warehousing and Data Mining, Computer Networks and Network Security etc.