International Journal of Engineering Research and General Science Volume 3, Issue 5, September-October, 2015
ISSN 2091-2730

A Survey to Automatic Summarization Techniques

Sherry
Computer Science Department
Thapar University
Patiala (Punjab), India

sherry.0989@gmail.com

Dr.Parteek Bhatia
Computer Science Department
Thapar University
Patiala (Punjab), India
parteek.bhatia@thapar.edu

Abstract- This paper presents a review for automatic text summarization techniques. Text Summarization is extraction of required information from the huge documents by ignoring the irrelevant details with the help of distinct approaches and provides the summary in compact way which reduces the effort and time of user. Due to the huge availability of data in all the fields the management and study of the information become difficult. Due to this the interest in the research of new automatic summarization techniques has been increased. Different approaches are used to determine summary. Plain text and multilingual text summarization play very important role in summary generation. This paper provides the review of all the existing text summarization techniques.

Keywords-Summary;UNL,NLization, UNLization;IAN; EUGENE;Multilingual.

INTRODUCTION

Automatic Text summarization has become very important part of our life due to huge volume of data that required to be compressed for people so, that it becomes possible to go through essential contents in short period of time. The research on the automatic text summarization has been started from 1950's.After a gap of decade's progress in the field of language processing was done due to the growing volume of online text. So, the large amount of electronic documents which are available in Internet has stimulated the development of excellent information retrieval systems. For example, Google always shows some part of the text corresponding to the query of the user. The user has to decide whether the document is interested or not only on the basis of extracted text. The user has to browse all the documents until a right document come. The solution is a use of automatic text summarizer which displays only the important and essential information about the document. Hence, it becomes very easy for the user to choose the right document.
The demand of the text summarization has observed in various domains like education, medical, government offices, research, business *etc*. So, the development of automatic summarization systems has importance due to research point of view. There are different approaches used for text summarization on the basis of single document, multi document summarization

APPLICATIONS OF TEXT SUMMRIES

Text Summarization is used is in medical field, in multimedia news summarization, in producing intelligent reports, in text for hand held devices, in text-to-Speech for blind people, in education and in summarizing meetings [1]. Many other scenarios use text summarization. For example, an information retrieval system uses automatic summarization to produce the list of retrievals. Now a day's summary of the email messages and news articles is sent to mobile devices as Short Message Service (SMS). Search engines also use summary mechanisms. The summary of the web pages is shown on the screen as a result of particular search.

SUMMARIZATION TECHNIQUES

a) Single Document Summarization

In single document summarization only one document is provided for summary generation. It is a simple and earliest approach for summarization. Extractive and abstractive both summaries methods can be applied on single document summarization.

b) Multi Document Summarization

Multi document summarization is also very important part of summarization. More than one information sources are provided for summary generation. Many web based clustering systems like news were inspired from multi document summary. But task of multi document technique is more difficult and complex than single document techniques. The real aim is not only to remove redundancy

and indentify correct text for summary but also to provide novelty and ensuring that final summary should be coherent and complete in itself. So it was a challenge for them to consider all the documents and relate the summary

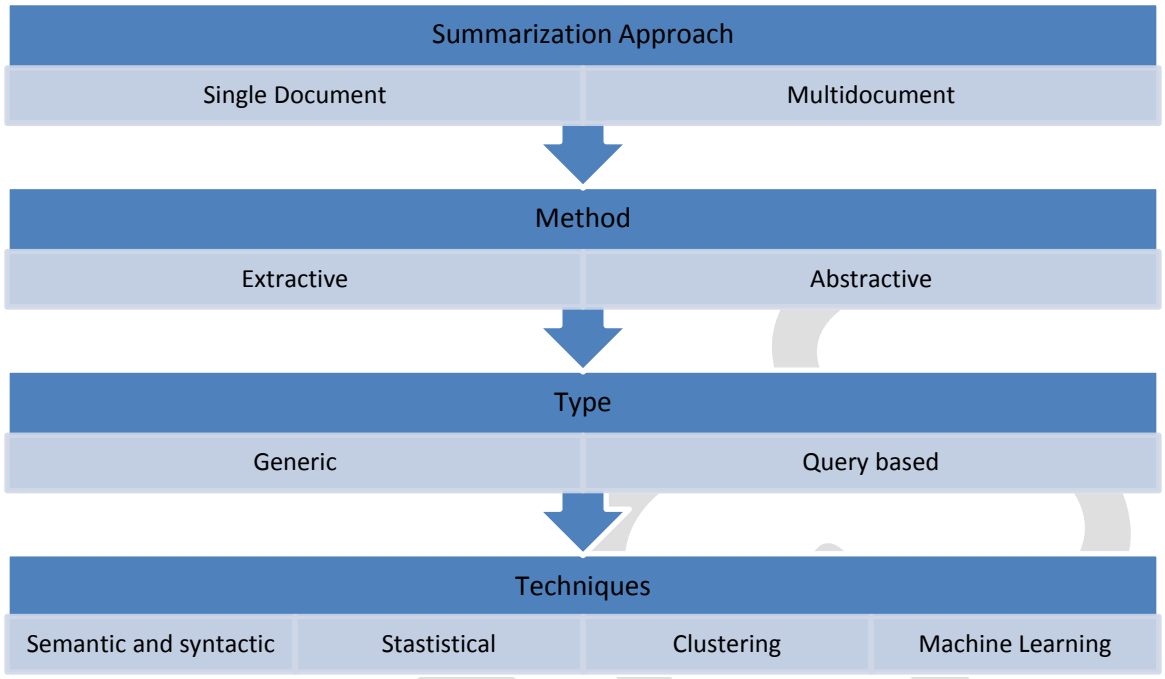

Figure 1.1: Summarization approaches

The work was initiated by the NLP group in the University of Columbia where summary system is called SUMMONS. At start procedures and challenges were different but later on people from different communities added their own perspective to the problem. Some approaches use clustering to indentify common themes and later on each cluster is represented by a sentence others produce more than one sentences from a cluster, while some uses maximal marginal relevance to work dynamically to include a passage if it is novel with the previous passages [2].

a) Extractive Summaries

Extractive summaries are simplest form of summaries. These approaches select the important information in the form of sentences, paragraphs *etc.* from the source document; combine them to generate short sentences. The selection of the sentences is made on the basis of different features like linguistic, statistical features *etc*. No synonyms of the words are used in extractive summaries for simplification [3]. Extractive summaries are not suitable for multi documents because sometimes these are biased towards some information sources.

b) Abstractive Summaries

Abstractive summaries are complex as compare to extractive summaries. These summaries consider the proper understanding of the source document and redefining it into new simple words [4].Actually it generates internal semantic representation; then use the natural language techniques for the summary creation which is close to human mind summary generation. Abstractive summaries add synonyms for the more simplification of the summaries.

c) Query based Summaries

Query based summaries are applicable in case of both single document and multi document summaries. Query based summaries are produced on the basis of is to retrieve sentences which satisfy user query. The score of the sentences are calculated on the basis of frequency of words in a document. A sentence with the query phrase provided a high score than others. The sentences with the high value of the score are extracted for summaries. Partial sentences may also be extracted and further union of them is carried out [5].

d) Generic Summaries

Generic summaries are not query based. Query based summaries are biased because they do not provide the overall review of the source document. They deal with the user queries only hence not suitable for content overview. To define the category of the

document and to describe the main key points of the document generic summaries are required. A best generic summary considers the main topics of the documents and tries to minimize the redundancy as less as possible [6].

e) Summary Techniques

There are four summary techniques which will be described as follows.

- Semantic and Syntactic (Rule-based)
- Statistical Technique
- Clustering Technique
- Machine Learning Technique

Semantic and Syntactic (Rule-based)
There are many semantic analysis techniques which are applied on text summarization to find the relation between different sentences. Following are the three semantic and syntactic summary techniques.

- Graph Representation
- Lexical Chains
- Natural language processing

In the graph representation lexical graphs, Graph matching, weighted graphs and unweighted graphs are used for summarization. In lexical chains word net, co-reference chains and lexical semantics *etc.* are used for summarization. Natural language processing used information extraction, part of speech tagger for the summarization. Summarization techniques under NLP are divided into two categories as follows.

- Plain Text Summarization
- Multilingual Summarization

In Plain text summaries resultant summary is in the same natural language but in multilingual text summarization resultant summary is in different natural language. Initial work in plain text summarization was started in 1950's. Most initial work on the text summarization was targeted on technical documents. Luhn (1958) proposed the first algorithm for text summarization at IBM. The author proposed text summarizer which was based upon frequency of a particular word in a document [7].The main motivation was to short the news information, biographical information. According to the Luhn summary has different categories, some of the summaries are difficult to generate than other. Different categories are Extractive, Abstractive, Indicative, Informative and Critical. Extractive summaries are simplest. These summaries contain the sentences which have already presented in text. Abstractive summaries contain some new text also. Indicative summaries represent the scope of the whole document without including whole content. Informative summaries represent the important factual content of the text document. Critical summaries represent reviews on scientific papers about their work and results. Baxendale (1958) also did work related to extractive summaries at IBM. The author more focused on the "sentence position". Approximately 200 documents are analyzed for research [8]. Edmundson (1969) proposed the system for document extraction. The proposed algorithm was the first algorithm for extractive summaries. Two previous features sentence frequency and position of the sentence were used along with the new features like cue words and skeleton. Cue words are words like hardly, significant etc. Skeleton define the heading of the document. After evaluation 44% results matched with the manual results [9].

Multilingual text summarization is come into existence in 2005. This technique is still in early stage but this different framework has many advantages in the newswire field in which information is combined from different foreign news agencies. Evans (2005) described the scenario in which there is always a preferred language in which summary is required, different multiple source documents are in demand and in different languages are available. They preferred English as a source language and documents are from the news articles in English language and Arabic. The logic was to generate the summary of English articles without discarding the details contains in Arabic. IBM's machine is used to do a transformation of Arabic language to English. The system checks the transformed document in Arabic corresponding to a document of English for each sentence. If match is found then sentence is found relevant for summary. Hence more grammatical summary is found this way, since machine translation is still not perfect of that. To find out the similarities between sentences Simfinder tool was used. This is a clustering based tool based upon similarity over different semantic and lexical features which is using long linear regression model. Universal Networking Language is mostly used in multilingual summarization.

Martins and Rino proposed algorithm for the text summarization using UNL. They presented UNLSumm model to prune the UNL text by means of heuristics that totally focus upon unnecessary binary relations. The system used decoder to produce corresponding

summary in Brazilian Portuguese. Their pruning heuristics are based upon the relations of UNL. Although each relation is not candidate for pruning because some relations like "agt" or "obj" convey important information [10]. Only some of the relations are candidates for pruning. According to this algorithm initially there was 84 heuristics were divided into two groups A and B shown by Figure 2.2 and Figure 2.3. Group A considers 39 heuristics. It also called as single pruning and removes the independent binary relations one by one. Group B heuristics are complex than the Group A heuristics. Group B heuristics are called chained pruning, *i.e.*, once the binary relation is excluded the interconnected binary relation is also excluded.

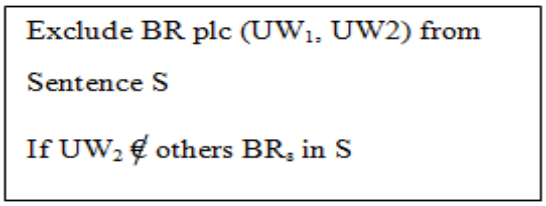

Figure 2.2: Group A heuristics

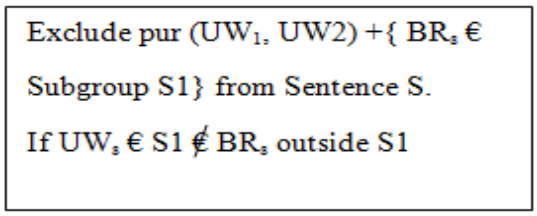

Figure 2.3: Group B heuristics

According to Figure 2.2 Group A heuristic delete the place relation from the UNL document, provided frequency of UW_2 is one means UW_2 should not be a part of any other relation in the same UNL sentence. While applying Group B heuristics frequency of UW_2 should be 2. These heuristics are more complicated because deleting a desired relation containing UW_1, UW_2 leaves blank [ ] in any other relation where UW_2 is placed. Hence to avoid this situation the relation containing [ ] is also removed from the UNL document. For example, if purpose relation as shown in Figure 2.3 is deleted containing UW_2 then any other relation in same UNL sentence containing UW_2 no more will the part of UNL document.
The serious problem regarding these heuristics is to decide the heuristics application order when considering both type of pruning. By default Group A heuristics are always applied and in case of interdependency when dangling of binary relations occurs, Group B are applied. However, Group A and B work on the same binary relations but sometimes after applying Group B heuristic results into more than one dangling relations. Hence, to give a priority to Group A or Group B heuristics is one of the major issues. The precision of the Heuristics is calculated represented by (2.1).

$$Precision(H) = Sat\_Num/Total\_Num \quad \ldots (2.1)$$

Sat_Num = No of applications of H leading to satisfactory results

Total_Num = Total No including satisfactory and unsatisfactory results

There are some limitations of approach which are as follows.

- Sometimes it covers non-relevant information.
- There is an upper bound to the number of heuristics applied for each entry.
- Application order is relevant and providing satisfactory results or not.

Managaikarasi and Gunasundari (2012) proposed an idea of text Summarization. The most important work they have done is improved methodology, which scans the document and transform into UNL graph. The system introduced UNL as a language for knowledge representation and information representation that can be describe in natural language conversation. They proposed method to find the summary of UNL document. The documents are collected from websites based upon the education domain. These documents contain images and unwanted information also. In the First Step stop words are removed. The sentence splitter is used to split document into sentences. The delimiter used is blank space here. In the next step sentences are again splitted into words. Then Morphological analyzer analyzes these words to find out the root word. These root words provide to UNL dictionary. Tenses and heuristic relations of the root words are indentified. The graph is constructed from given information. During graph construction counter field is also updated. Counter field is provided to find the important concepts, based upon threshold. Highest concepts sentences are finally picked for the resultant summary of the document. The system is tested on education domain document for summary. There was manually tested on the summary with experts. During summary preparation the data is collected from the news service providers. Each document includes the irrelevant information like images, tables *etc*. So, there is a need of creation of ideal summary for evaluation of results. For the ideal results the documents are distributed to three judges and rank is given to the sentences

according to their importance in text document. The future work for project is develop a well managed tool for evaluations and updating of UNL dictionaries with the help of root words provided by Morphological analyzer. It also identifies more and more UNL relations with the help of heuristic rules [11].

Pandian and Kalpana (2013) also proposed text Summarization mechanism using UNL. They focused on the tourism domain document which is UNL based. The Bengali UNL system is developed by them. UNL representation used by the system was for simple sentences not for complicated sentences. The main focus was mostly on DeConversion part which converts the universal networking language to Tamil language. The source document is scanned and it is converted into intermediate language. It further undergoes generation process for final output. For the summary process the source document undergoes a process of EnConversion which includes the steps like parts of speech tag, parts of speech parsing, identification of entity, and identification of relation, creation of dictionary and generation rules. Source document is converted into UNL document of UNL expressions. In the first phase parts of speech tagging and parts of speech parsing is carried out with the help of Stanford Parser. The outcome of the parser is used to find the entities and relationship between entities. Further rules are constructed and knowledge base is obtained for the generation of UNL expressions. UNL document containing UNL expressions are passed to the DeConvertor for the generation of the final summary and final output for three levels of users (level1 user, level2 user and level 3 users). The DeConversion module is constructed in such a way that it will perform the function of both summarizer and DeConvertor. To obtained the summary DeConversion module scan the word dictionary and find the relation between the different universal words, attributes of the universal words are collected and relation between universal words are taken. Further the unnecessary information like determiners, prepositions are reduced to obtained the final summary document. Final summary document is produced for the different levels of user's base upon the classification of ages. The distribution level of the summary document based upon the IQ level. DeConversion module produce summary in three steps which are as follows.

- Analysis and preparations of the dictionary information,
- Preparing DeConversion rules and
- DeConversion to produce the output summary document.

The experimental analysis is carried out using NetBeans IDE. The analysis obtained different levels of users based upon their level of IQ to access intelligence. The overall population is considered for experimental analysis. The performance of the overall system is analyzed and considered in the form of Decisiveness for all levels. It is defined by number of words compressed at different levels [12]. It is calculated for all the levels by using the given formula which is shown in 2.4.

$$\text{Decisiveness for the level User(DLU)} = \left\{100 - \frac{\text{No of words in level summarized document}}{\text{Total No of words in original document}} * 100\right\}$$

Decisiveness is find out and graph is plotted against decisiveness ratio and document. After plotting a graph it is observed that the compression for original document for level 1 user is more than rest level users. Same is true for rest two level users.
Sornlertlamvanich *et al.* proposed an approach for Summarization using Universal Networking Language. While producing summary this approach considers surface and semantic information of the UNL. The multilingualism can also be realized using DeConvertors from the summarized UNL document to the resultant target natural language document under the framework of UNL. Algorithm consists of four steps. In the first step the score of each UNL sentence is calculated. Score of the sentence is calculated by using weight of each universal word. Weight of each universal word is calculated by using the factor of frequency and inverse document frequency. After the score calculation some top most sentences based upon score are chosen for the future summary. By using the semantic information of the UNL the redundant words are removed from the summary in third step. This is mathematically calculated by using contribution function. The values obtained through contribution function are compared with the threshold value 1.5. To make summary more natural and real different sentences are merged based upon the head of the sentence and no of words in the sentence in fourth step. This algorithm is applicable for multiple document summarizations. Their experiment proved that use of the UNL improves the summary quality as compare to the plain text summarization. The semantic information of the UNL can also be applied to improve the naturalness in sentence level of summary [13].

Sherry and Parteek Bhatia (2015) also proposed an algorithm for multilingual text summarization technique. The algorithm was based upon the hybrid approach *i.e.* it extracts the best features of the previous algorithms and adds new features also. The system used UNL for language transformation. It was a six step algorithm. The overall complexity of an algorithm decreased due to removal of unnecessary relations in first step. Second step calculates the score for each UNL sentence. In the third step the best sentences are chosen for summary. Fourth step refine summary by calculating contribution functions on modifier relation. At the end sentences are merged and again UNL summary is processed for removing of unwanted words [14].

Statistical Technique

To extract relevant sentences for summary some summarization systems use statistical techniques. Binomial distribution, sentence compression and relevant scores these are statistical method used for summarization. Hidden Markov model also uses this approach.

Conroy and O'Leary (2001) applied hidden Markov model approach for plain text summarization. They used sequential model for the local independence. The system had three features: position of a particular sentence in document, number of different terns in a particular sentence, likeliness of particular sentence terms.

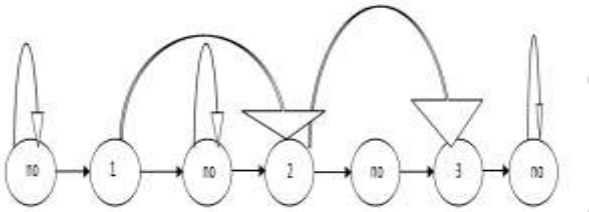

Figure 2.4: Markov model to extract the three summary sentences [23]

In Figure 2.4 Markov model is shown for $2s + 1$ states. Summary states and non summary states are alternated. In this "no" represents the non summary states and numerical numbers represents the summary states. There is a jump to next state in case of summary state. The Figure 2.4 represents the model with 7 nodes corresponds to $s = 3$.

2.7.3 Clustering Technique

Clustering is a process in which different objects are grouped based upon their properties, *i.e.*, objects with the similar properties belongs to same group. In a document different topics are arranged in a particular sequence. In cluster based summaries sentence selection is based upon the cluster Ci. The second factor is location of a particular sentence Li. The factor which increases the score of a particular sentence is its similarity to already present sentence in a document. The overall score of a sentence depend upon these three factors.

$$S_i = W1 * C_i + W2 * F_i + W3 * L_i \quad \ldots (2.5)$$

In this Si is score of sentence, C_i is cluster to which sentence belongs, Fi is document to which sentence belongs and L_i is location of a particular sentence.W1, W2 and W3 represents the weights [15].

Machine Learning Technique

Machine learning techniques are very effective for automatic text summarization. The some of the machine learning approaches are discuss as follows.

a) Naive Bayes Approach

Kupiec (1995) described a method for summarization. He described a classification function known as naïve Bayes classifier which is responsible for the each sentence to be a part of summary. If S denotes the total number of sentences and s denotes a particular sentence with a features F1 to Fk [16]. The formula of naïve Bayes is shown in 2.6:

$$P(s \in S|F1,F2,\ldots.Fk) = \frac{\left(\pi_1^k P(F_i s \in S).P(s \in S)\right)}{\pi_1^k P(F_i)}$$

The new features like sentence length and the uppercase words were added. Score is calculated for each sentence and based upon that top most n sentences were chosen. Aone *et al*. (1999) also describe a naïve Bayes classifier with more additional features. He introduced the terms "frequency" and "inverse document frequency" in plain text summaries. The corpus used in the experimental analysis was from newswire. The inverse document frequency was computed from a large corpus of the same area.

b) Rich features and Decision Trees

Lin and Hovy (1997) describe the importance of a feature "sentence position". According to this a weight is provided to sentence based upon its position in the text [17]. This method also called as position method. A newswire corpus was used for experimental

analysis. The authors measured the yield of every sentence position. They ranked the different sentence positions to produce the "Optimal Position Policy (OPP). They performed the two kinds of evaluations. They test on the unseen text. The first evaluation was exactly like the training documents and the second evaluation considered the word overlap for the manual abstracts was measured. Abstract windows and selected sentence windows were compared and precision, recall values were measured.

Lin (1999) broke away the assumption that the features are independent and tried to model the problem using decision trees instead of naïve-Bayes classifier. The system described lot of features in sentence extraction and their effects. The data set used was publicly available texts classified into various topics. The data set is divided into text fragments which are evaluated by human judges. Some important features were query signature (normalized score of the sentences depending on the number of query words), IR Signature (the salient word like the signature word), numerical data, proper name (Boolean value 1 is given to sentence that had a proper name), pronoun or adjective (Boolean value 1 is given if they appeared), weekday or month, Quotation, query and signature. The system experimented with different baselines like positional feature, simple combination of features. When machine extracted and human extracted sentences were matched, the decision tree was clearly the winner.

c) Log Linear Models

Osbrone (2002) described the Log Linear model approach for the plain text summarization. This approach is different than the previous approaches which always assumed feature independence. The system showed that this approach is better than naïve Bayes classifier approach [18]. The model can be stated mathematically as follows.

$$P(c|s) = \frac{1}{Z(s)} \exp\left(\Sigma_i \, \lambda_i F_i(c,s) \right) \qquad \ldots (2.7)$$

Where Z(S) is

$$Z(s) = \sum c \, (\sum i \, \lambda_i F_i(c,s))$$

In these equations c is a label, s is a item to be labeled, f_i is a feature (i-th feature) and λ_i is weight of the feature. There are two possible labels regarding whether the sentence is to be extracted from the document or not. The weights given to sentences are calculated from conjugate gradient descent. The non uniform prior is added to the model by authors. This model rejects too many sentences during processing. The features included by the authors were word paring, length and position of the sentence and discourse features like inside the introduction, part of conclusion.

d) Neural Networks

DUC (2001) applied the neural network technique for plain text summarization. The system produced a summary of single news article in 100 words. However the best systems in evaluations of experiments could not outperform the baseline analyzed by Nenkova in 2005. After 2002 the task for the single document summarization was dropped by DUC. Svore (2007) produced an algorithm based upon neural networks and used the third party features like dataset to resolve the problem of extractive summarization. The data set consists of 1365 documents collected from CNN.com. The datasets consists of human generated stories, articles, title and timestamp *etc*. For the evaluation two metrics were considered. The first one is to combine the system produced three highlights, combine the human generated three highlights and comparison of these two. The second take care about the ordering and the individual level comparison of the sentences [19].

Strove (2007) trained this model on the basis of labels and featured for each sentence that referred the ranking of each sentence in source document. Ranking was provided to the sentences on the basis of RankNet which was a paired based neural network algorithm. ROUGE-1 is used as a training set. The authors concluded that if a sentence contains keywords regarding new search engines and Wikipedia articles then the probability of a sentence in highlight is more.

ROUGH-1, ROUGH-2 was used for the evaluation purpose and statistically improvements were shown over baseline [20].

e) Other Approaches

Deep Natural Language Analysis Methods

Barzilay and Elhadad (1997) also described the technique for summarization. It is called Deep NLP analysis. The system described lexical chain that is formed using sequenced words in a given text, neighbor words called as spanning short and long distances. The following steps were used by them. First of all segmentation of the whole text, lexical chains identification, strong lexical chains are used for identification of the sentences. The system described cohesion in the document means togetherness of the different parts of the text. In the lexical cohesion semantically related words are used. For example, consider the sentence:

Amit bought a jag. *He loves the car* ... (2.9)

In (2.9) the word "car" refers to the word in the previous sentence "jag". It is lexical cohesion. The cohesion formula phenomenon occurs at word level as well as sentence level too. This results into lexical chains which are building blocks for summarization. Relation of the different words and their sequence was also find out which result into several chains and responsible for document representation. For the lexical chain determination Word net was used and three steps were applied.

1. Selection of candidate word set.
2. Find an appropriate chain for each candidate word.
3. If a chain is found word is inserted in a chain and then further updation is carried out.

Word net distance is used to measure relatedness. To find a desired set of candidate's simple nouns and compound nouns were used initially. At last summary is created by using strong lexical chains. The score was provided to them on the basis of length and homogeneity. Significant sentences were chosen on the basis of some heuristics.

Ono *et al.* (1994) put forward a computational model. They elaborated the procedure for extraction of rhetorical structure. To represent a relation between the different chunks of sentences binary trees were used. There were series of NLP steps used for the structure extraction they are analysis of sentences, rhetoric relation, text segmentation, generation of candidate and judgement. Evaluation was done on the basis of importance of different relations. Nodes of trees were pruned to reduce the sentence but the important information was always the part of the tree. Same procedure was applied to the paragraphs for summary generation. The data set used was 30 articles of Japanese newspaper [21].

Marco (1998) described a new approach without assuming as sentences in a document form a sequence which is flat. This approach was really different from other approaches. Rhetorical Structure Theory (RST) is used in the approach [22]. According to the system there are two non overlapping pieces of text spans known as nucleus and satellite. The purpose of the nucleus is to express the more essential part according to the writer but satellite did not do this. Satellite was dependent on nucleus but not vice versa.

The numbers inside the nodes represents the sentence number in the provided text. The text below the number is rhetorical relations in the selected text. There were different metrics used they were cluster based (in which each node was given a score on the basis of internal or external node. external nodes were always zero score nodes and the score of internal nodes depends upon the immediate children. Discourse tree is chosen when it has more score than other tree.), metric based upon the marker (a discourse structure A is chosen than B if A used more rhetorical relations than B), technique based upon clusters, metric based upon shape (prefer tree which is more skewed in shape),metric based upon connectivity (discourse structure is chosen if connectivity is higher than others), metrics based upon titles and positions.

There is also more different type of approaches for summary generation like short summaries, sentence compression and sequence document representation. Short summaries are headline style summaries. In a system there is a statistical relationship between headline and source text units. Statistical approach is used for sentence compression. The basic idea behind the system is compressing the sentence may be useful concept in single or multi document summarization

Research Gap

It has been analyzed that lot of research work has been done in plain text summarization. The various algorithm techniques have been proposed with different methodologies. Multilingual text summarization is a new research area in automatic summarization. A little work has been done in this research area. Mostly UNL is used for language translation. The various plain text summary techniques can be used to multilingual summarization for higher efficiency.

Conclusion and Future Scope

In this paper various approaches of summarization like single document, multi document, extractive, abstractive, generic and query based *etc*. has been described. The various summary techniques like machine learning, semantic and syntactic, clustering and statistical also have been discussed. Machine learning includes naïve Bayes, decision trees, and neural networks. Multilingual text summary approaches also have been described with detail procedures.

REFERENCES:

[1] Lal, Partha, and Stefan Ruger. "Extract-based summarization with simplification." Proceedings of the ACL. 2002

[2] Das, Dipanjan, and André FT Martins. "A survey on automatic text summarization." Literature Survey for the Language and Statistics II course at CMU 4 (2007): 192-195

[3] Cheung, Jackie CK. Comparing Abstractive and Extractive Summarization of Evaluative Text: Controversiality and Content Selection. Diss. UNIVERSITY OF BRITISH COLUMBIA, 2008

[4] Gupta, Vishal, and Gurpreet Singh Lehal. "A survey of text summarization extractive techniques." Journal of Emerging Technologies in Web Intelligence2.3 (2010): 258-268.

[5] Karmakar, Lad, and Chothani Hiten. "A Review Paper on Extractive Techniques of Text Summarization." (2015).

[6] Gong, Yihong, and Xin Liu. "Generic text summarization using relevance measure and latent semantic analysis." Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2001.

[7] Lal, Partha. "Text Summarization." (2002).

[8] Baxendale, Phyllis B. "Machine-made index for technical literature: an experiment." IBM Journal of Research and Development 2.4 (1958): 354-361.

[9] Edmundson, Harold P. "New methods in automatic extracting." Journal of the ACM (JACM) 16.2 (1969): 264-285.

[10] Martins, Camilla Brandel, and Lucia Helena Machado Rino. "Revisiting UNLSumm: Improvement through a case study." the Proceedings of the Workshop on Multilingual Information Access and Natural Language Processing. Vol. 1. 2002.

[11] Mangairkarasi, S., and S. Gunasundari. "Semantic based text summarization using universal networking language." Int. J. Appl. Inf. Syst 3.8 (2012): 18-23.

[12] Kalpana, S. "UNL based Document Summarization based on Level of Users."International Journal of Computer Applications 66.24 (2013).

[13] Sornlertlamvanich, Virach, Tanapong Potipiti, and Thatsanee Charoenporn. "UNL Document Summarization." Proceedings of the First International Workshop on Multimedia Annotation. 2001.

[14] Sherry, Parteek Kumar," Multilingual Text Summarizer" in "International Conference on Advances in Computer Engineering and Applications (ICACEA) IMS Engineering College, Ghaziabad, India 2015.

[15] Karmakar, Lad, and Chothani Hiten. "A Review Paper on Extractive Techniques of Text Summarization." (2015).

[16] Conroy, John M., and Dianne P. O'leary. "Text summarization via hidden markov models." Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2001.

[17] Das, Dipanjan, and André FT Martins. "A survey on automatic text summarization." Literature Survey for the Language and Statistics II course at CMU 4 (2007): 192-195.

[18] Osborne, Miles. "Using maximum entropy for sentence extraction."Proceedings of the ACL-02 Workshop on Automatic Summarization-Volume 4. Association for Computational Linguistics, 2002.

[19] Kaikhah, Khosrow. "Automatic text summarization with neural networks." (2004).

[20] Svore, Krysta Marie, Lucy Vanderwende, and Christopher JC Burges. "Enhancing Single-Document Summarization by Combining RankNet and Third-Party Sources." EMNLP-CoNLL. 2007.

[21] Barzilay, Regina, and Michael Elhadad. "Using lexical chains for text summarization." Advances in automatic text summarization (1999): 111-121.

[22] Flowerdew, Lynne. "An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies." English for Specific Purposes 24.3 (2005): 321-332