Volume 4, Issue 3, March 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Theme Encapsulation and Content Framework using Lexical Chaining

A.V.Seetha Lakshmi
Asst.Prof, Department of IT
G.T.N Arts College, Dindigul, India

Dr.S.R.Victor
Head,Department of Computer Science
St.Xavier's College (Autonomous) ,PalayamKottai, India

Abstract—A developing event which contains many related events and activities is defined as a topic. There is a difficulty in exploration as there are sequence of documents published by different authors for the particular keyword and the convenience of storage in internet. This phenomenal growth has made the users to read the entire contents and conclude what exactly present in the document. In this paper, we propose a summarization technique in which the core content is summarized in the chronological order. The summarization process involves four steps 1. The original text is segmented, 2. Lexical chains are constructed using new algorithm, 3. Strong chains are identified, 4. Significant sentence are extracted. The extracted sentences are associated to find the temporal closeness with the help of evolution graph.

Keywords— Data mining, Text mining, Encapsulation, Lexical chains, Summarization

I. INTRODUCTION

The internet provides an abundant source of information through the number of documents. The improvisations in the technology have paved the way for efficient search requests to satisfy the keyword search request. But, still readers have much difficulty in obtaining the required document from the abundant resources. The time related events add more controversy to the situation. The project, "Topic Detection and Tracking" initiated by the DARPA (Defence Advanced Research Project Agency) defines the topic as the *"semantic event or activity along with all directly related events and activities".* In this paper, a computational algorithm that detects all topics and track related documents from several documents using the keyword search technique. A prominent text mining research pattern known as Topic anatomy is used to summarize the essential document in chronological order. It involves three major tasks such as *theme generation, event segmentation and summarization, evolution graph construction.*

The process of condensing a source text into shorter version is known as summarization. The content information is preserved in this method. The summarization process serves several goals such as survey analysis of a scientific field, quick notes on the general topic of the text, etc. The quality informative summary can be produced only by the full understanding of the text. The quick indicative summaries are used to decide whether the text is worth reading. The indicative summaries are easy to produce than the informative summaries. Such indicative summaries are produced to select the particular topic for reading in chronological order. Jones in 1993 explained summarization as a two steps process that involves 1. Building source representation from the source text, 2. Summary generation. There may be controversy to select the information to be included to create the summary. There are three types of source text information such as linguistic, domain and communicative. Summaries can be built using different techniques such as deep semantic analysis of source text, shallow linguistic analysis of the text. These techniques can be easily computed and relies on the clues available in the text. For example, the deep semantic analysis of the source text can be done by investing the ways to produce the coherent summary of many texts that describes the same event. This can be done when the full semantic representation of the source text is available. This type can be expressive but they are hard to compute as they are domain dependent. The shallow linguistic analysis can be done by various methods such as abstracting the source text into frequency table, cue phrase method and location method. The latter two methods produce better results than the first method and can be computed easily. The most severe limitation of location and cue phrase method is their dependency on the text genre.

In the proposed technique, the first task is to identify themes of the topic from the several document that are related. Each document may reflect its importance than the other while reading it. These events must be defined in a unique way. Later, the task of event segmentation and summarization process which extracts the events and presents the event in the chronological order is done as in Google news service that arrange documents related to the news topic from online news website. While our system TECF(Theme encapsulation and content framework) using lexical chains detect the core content of the document in the effective manner. The lexical chains are used as a model of source text to produce a summary. The source text needs to be integrated with the text representation to produce quality summaries. In this paper, we describe how lexical chains are used to identify significant sentences within the source text to produce the summary.

II. RELATED WORK

A. *Text Segmentation*

The document related to the topic is divided into segments that are related to the topics. Non overlapping segments are formed in this technique. The segmentation can be classified into two types based on the input text as story boundary detection and document sub topic identification. The text stream is given as input to the story boundary detection. In general, cue phrases are used to identify the boundaries between the documents. For example, there are no distinct boundaries between the documents from online news documents. In document sub topic identification, with a single input document the blocks of document that are relevant to certain sub topic are identified. The cue phrase approach does not suit this technique as the subtopics in the document are similar. Therefore, the cue phrases for the document about the subtopic boundaries do not exist virtually. For example, Search engine can retrieve the documents and return the most relevant blocks segmented from the search results. The document can be decomposed into a set of consecutive sentences and the word usage in every block is analyzed in finding the subtopic boundaries. The major problem in this technique is that the block interrelationships cannot be determined with the information in the block.

Brants et al and Choi et al [3] to enrich the information in a consecutive set of sentences applied the latent semantics concepts. The training data is used to create a domain dependent construct in this method. Blei and Moreno [4] utilized Hidden Markov models to detect the subtopics of document and used it to model as states in an HMM. In this method, every document is treated as a series of blocks, which is used to calculate the best state transition. When two successive states in the best state transition sequence are different, then the boundary occurs in the documents.

Ji and Zha[5] proposed a domain-independent segmentation method that models the block of the document using a square matrix. It considers the matrix as a gray scale image. then some image processing method is applied to sharpen the boundaries in the image .Finally the significant and diagonal segments are selected as a block of the document.

B. *Text Summarization*

Text summarization creates one or more documents that capture the list of documents automatically. A document's content may consist of many themes, so generic summarization methods are used to extend the summary diversity to provide wider coverage of the content of the documents[6]. In text summarization, the informative sentence is extracted from the actual documents by composing the summaries. Extraction-based text summarization methods can be classified as supervised and unsupervised. In supervised methods, the document is summarized by labeling the sentences of the document as either informative or non informative. Shen et al[7] proposed a supervised summarization method that uses conditional random fields (CRF) to train a classification model. It calculates the information of the sentences. Top ranked sentences are selected as a summary with suitable training corpora. The supervised summarization methods perform as well as unsupervised summarization methods. Domain-dependency is a drawback of supervised summarization methods. The trained summarization model is specific to a certain document domain. Deploying a supervised summarization method in a new domain involves explaining another manual training corpus and it is a time consuming task. In general, the number of studies show that inter agreement between explanator is low, which affects the quality of the training corpus and the acquired summarization model .We recently proposed that the summarization method are unsupervised. Next, we consider the method and discuss the limitations of applying them to the topic anatomy task. Gong and Liu applied Singular value Decomposition (SVD) to a document. The term-sentence association matrix can be used to perform extraction-based generic summarization. This method uses the decomposed singular vectors for the themes of the document and composes diverse summaries by selecting the informative sentences from important themes. Nomoto and Matsumoto [8] proposed the X means algorithm ,which is used to find the sentences that contains more useful information from the clusters. Allan et al temporal summarization method [9] processes topic sentences in a chronological order. This method weights the information of the sentences depending on the usefulness and novelty of the sentences. A sentence is useful for readers to comprehend a topic if its content is similar to the main themes of the topic .To avoid the extraction of redundant summary sentences, the sentence should be different to all previously extracted sentences. Nowadays, graph based summarization methods are used to model the relationships between the sentence and terms in the document. This model considers a sentence informative if it connects with many informative terms and the reinforcement procedure updates the informative scores of the terms and sentences. Finally, the summaries are composed by selecting the informative sentences.

Erkan and Radev[10] represent the set of documents as a graph in the sentence are represented as nodes and the edges connects the content similarity between the sentences. A sentence is more informative, if it connects with many sentences, hence, the connected sentences are also informative. From the informative scores of the sentences, the informative sentence can be taken as the summary. Topic summarization differs from existing text summarization because of its temporal properties .The topic summaries should describe the storylines of the topics.

C. *Topic Evolution Mining*

Kleinberg [11] developed a technique to construct the hierarchical tree from a series of documents. This technique uses an HMM-based transition diagram to model the status of the topics and splits a topic into diverse themes. These topics are modeled as tree branches, if the topic contains busty information. Nallapati et al [12] formalized the problem of topic evolution mining as a text clustering task in which the identified clusters (the events of a topic) are connected chronologically to form the evolution graph of the topic .Yang and shi [13] focused on the temporal properties of a topic and showed that evolution graphs can be obtained by using the temporal information about topics. Feng and Allan [14] proposed an incident threading method that is similar to the proposed method. This method first identifies incidents from news documents and then identifies the semantic dependencies between the incidents. Swan and Allan [15] proposed a timeline system to detect the topic's importance graphically at the specific time.

III. IMPLEMENTATION

A. Theme Model

A Theme is a real world incident that contains one or more sentences that are related to a particular incident. In the entire theme, one sentence may be more important than the rest of sentences and it may be repeated in many sets of documents. We specify an event as an important theme enhancement that carries out a sequence for time period. By nature, all the sentences are merged to form the corpus of the theme. Even though the events of the subject are chronologically disjoint, they are assumed to be language-dependent in order to enhance the sentence. The proposed model finds the inter-relation between the theme and the event to make the theme's progression graph.

The major tasks involved in the process are crawling and extraction, Lexical chain computation, Event segmentation and summarization and finally the Story-boundary detection.

B. Crawling and extraction

The crawling is the basic operation performed in every web search Engines. The process of crawling takes place with the help of crawler. The crawler is a program that is already developed in the search engine. It isolates the citation from the World Wide Web (WWW) with respect to the particular theme through which we get the backlink. Normally, every backlink has a set of documents. Such documents may contain set of sentences, which helps to develop the storyline of the theme/topic. This storylines helps to understand the topic by making the user to comprehend the topic but it involves more time consumption. Hence, this complexity is reduced in our system by introducing extraction technique. The crawler can be represented as

In early days, single crawler can be used. Now a day parallel and multiple crawlers are used to enhance the optimization of the search engine. Hence, the user can benefit with the increased speed in displaying the output to the query. The main function of the crawler is to download the citation related to the theme or query [16].

The processing steps involved in crawler are as follows,

i) The seed lists that are maintained by the crawler should be cleared first to make sure that the old data will not be provided to the user.
ii) Every webhosts has an IP Address, which is determined in this step.
iii) The crawler can download the document based on the user's search.
iv) The backlinks present in the document are extracted by the crawler.
v) The user can perform any operations on the downloaded document.
vi) While processing the document, the user may switch on to other citation and if the citation is new to the user, such citations are added to the seedlists .
vii) The process is repeated from (i) for every query requested by the user.

Extraction

In the Existing technique such as TDT, after performing the crawling operations, they directly apply some of the summarization technique to produce the summary of the original document which sometimes did not satisfy the internet user since it can produce summary up to two lines and these two lines will be from the first two lines or last two lines of the original document.

Some of the summarization techniques are as follows,

i) forward method:

The summary is produced by extracting the first paragraph of the Original document and the first two lines of the paragraph are displayed.

ii) **backward method:**

The summary is produced by extracting the last paragraph of the Original document and the last two lines of the paragraph are displayed.

iii) k-means method:

The documents are formed into clusters to produce the summary by randomly picking up the cluster. It sometimes may not produce the exact summary [8].

iv) **frequent-content word:**

It can produce the summary by extracting the frequently arriving words in the paragraph [17].

A new technique called SVD (single value decomposition) [6] is used which enhances the summarization results. Our technique resembles SVD but we adopt soundix technique for the entries of vector. We adopt two types of extraction called BLOCK EXTRACTION and THEME EXTRACTION.

Block extraction

In this, the blocks (the paragraphs) are extracted from the Original document using Indexing Technique**.** To ensure the chronological order, the blocks are extracted from the original document. While indexing, we assign index value to each tag used in source code. With that index value, we can identify the paragraph tag(i.e)<p> and with this, the paragraph are extracted.

Extracting blocks form the inbound links in the websites. In this block extraction, extracting blocks and image are extracted for the topic. We obtain n number of blocks from the web sites.

Theme extraction

The Themes (sentences) from the identified block are extracted using the Cue Phrase Identification Approach. The themes are extracted to reduce the complexity involved in Lexical chaining. After extracting blocks, Events are extracted from the blocks. A set of events are extracted for every blocks.

The cue phrase identification extracts the sentence by identifying the full stop at the end of each sentence.

C. Lexical chaining

Lexical cohesion is an easily recognizable relation that enables the lexical chain computation. It is created between the two terms. The lexical cohesion was classified as reiteration category and collocation category by Halliday and Hasan. The repetition, synonyms and hyponyms results in Reiteration whereas the collocation specify the relation between the words that occur in the same contexts. For example, the sentence *She works as a teacher in the school* results in collocation. The collocation method is more problematic than the reiteration even though the text is identified on its surface.

Morris and Hirst in the year 1991 proved that lexical cohesion also occurs among the sequence of related words and presented the first computational model for the lexical chains. The lexical chains represent the lexical cohesive structure of the text. These lexical are commonly used for information retrieval and also for the correction of malapropisms. The computational model explained the relations in terms of categories, index entries and pointers in the Roget's Thesaurus and covered over 90% of the intuitive lexical relations. The chains are created according to the relatedness criteria to take a new chain and find the related chains.

Morris and Hirst introduce the notions of "activated chain" and "chain returns", to take into account the distance between occurrences of cognate words. They withal analyze factors contributing to the vigor of a chain — repetition, density and length. Morris and Hirst did not implement their algorithm, because there was no machine-readable version of Roget's Thesaurus at that time. One of the drawbacks of their approach was that they did not require the same word to appear with the same sense in its different occurrences for it to belong to a chain. For semantically equivocal words, this can lead to perplexities (e.g., commixing two senses of table as a piece of furniture or an array). Note that culling the opportune chain for a word is identically tantamount to disambiguating this word in context, which is a well-kenned conundrum in text understanding. More recently, two algorithms for the calculation of lexical chains have been presented in (Hirst & St-Onge 1998 to appear) and (Stairmand 1996). Both of these algorithms utilize the WordNet lexical database for determining relatedness of the words (Miller et al.1990). Senses in the WordNet database are represented relationally by synonym sets ('synsets')—which are the sets of all the words sharing a prevalent sense. For example two senses of "computer" are represented as: {calculator, reckoner, figurer, estimator, computer} (i.e., a person who computes) and {computer, dataprocessor, electronic computer, information processing system}. WordNet contains more than 118,000 different mword forms. Words of the same category are linked through semantic cognations like synonymy and hyponymy. Polysemous words appear in more than one synsets (for example, computer occurs in two synsets). Approximately 17% of the words in WordNet are polysemous. But, as noted by Stairmand, this figure is very bamboozling: "a paramount proportion of WordNet entities are Latin labels for biological entities, which by their nature are monosemous and our experience with the news-report texts we have processed is that approximately a moiety of the entities encountered are polysemous." (Stairmand 1996).

Generally, a procedure for constructing lexical chains follows three steps:

1. Cull a set of candidate words;
2. For each candidate word, find a felicitous chain relying on a relatedness criterion among members of the chains;
3. If it is found, insert the word in the chain and update it accordingly.

An example of such a procedure is represented by Hirst and St-Onge (henceforth, H&S). In the preprocessing step, all words that appear as an entity ingress in WordNet are opted for. Relatedness of words is tenacious in terms of the distance between their occurrences and the shape of the path connecting them in the WordNet thesaurus. Three kinds of cognations are defined: extra-vigorous (between a word and its repetition), vigorous (between two words connected by a Word-Net cognation) and medium-vigorous when the link between the synsets of the words is longer than one (only paths gratifying certain restrictions are accepted as valid connections). The maximum distance between cognate words depends on the kind of cognation: for extra-vigorous cognations, there is not inhibit in distance, for vigorous cognations, it is constrained to a window of seven sentences; and for mediumstrong cognations, it is within three sentences back. To find a chain in which to insert a given candidate word, extra-vigorous cognations are preferred to strongrelations and both of them are preferred to mediumstrong cognations. If a chain is found, then the candidate word is inserted with the felicitous sense, and the senses of the other words in the receiving chain are updated, so that every word connected to the incipient word in the chain relates to its culled senses only. If no chain is found, then a newc hain is engendered and the candidate word is inserted with all its possible senses in WordNet.The greedy disambiguation strategy implemented in this algorithm has some limitations illustrated by the following example:

Mr. *Kenny is the* **person** *that invented an anesthetic* **machine** *which uses* **micro-computers** *to control the rate at which an anesthetic is pumped into the blood. Such* **machines** *are nothing new. But his* **device** *uses two* **micro-computers** *to achieve much closer monitoring of the* **pump** *feeding the anesthetic into the patient.*

D. Building Lexical Chains

The calculation of lexical chains can be done by two algorithms by Hirst et al and Starimind. These algorithms uses the WordNet lexical database which are represented through synonym sets (set of all words sharing common sense). For example two senses of "*computer*" are represented as: *{*calculator, reckoner, figurer, estimator, computer*}*. Similarly more than 118,000 word forms are available in the WordNet. The Words are linked through semantic relations like synonymy and hyponymy. Polysemous words appear in more than one synsets (synonym sets).

Three steps to construct the lexical chains:

1. Select a set of candidate words;

2. For each candidate word, find an appropriate chain relying on a relatedness criterion among members of the chains;
3. If it is found, insert the word in the chain and update it accordingly.

Relatedness of the words should be tenacious by the distance between their occurrences. Three kinds of cognations such as extra-vigorous (between a word and its repetition), vigorous (between two words connected by a Word-Net cognation) and medium-vigorous when the link between the synsets of the words is longer than one subsists. The candidate word is inserted in the congruous sense and the chain is updated if the chain is found else the incipient chain is engendered and inseted in the wordNet.For example:

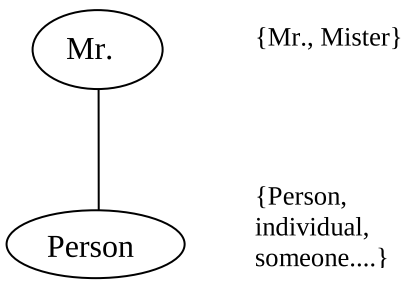

Fig:1: Step1

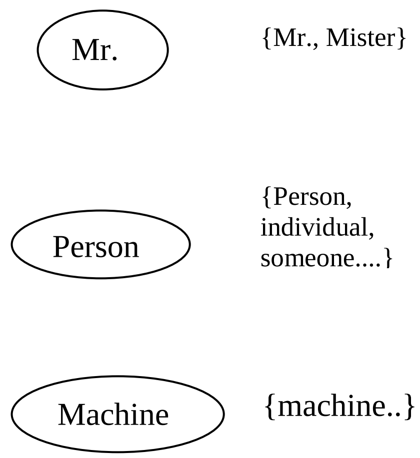

Fig: 3: Step 2

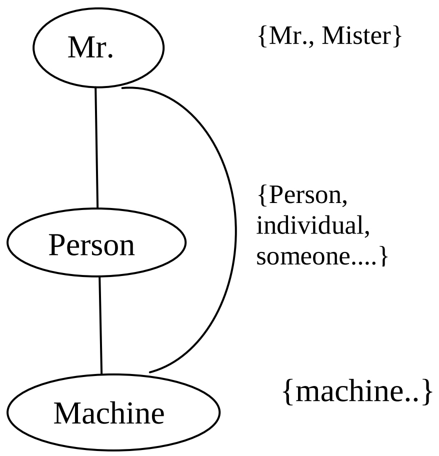

Fig: 2: Step 2

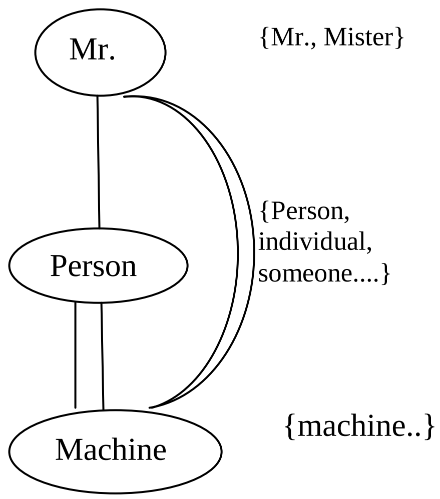

Fig: 4: Step 3

Once the lexical chain is constructed, the summaries must be built using lexical chains. The chain length, distribution of text in the span, etc are identified to build the chain strongly. After the selection of strong chains, the next step is to extract full sentences from the original topic based on the chain distribution. Three alternative method can be used for this technique.

1) For each chain in the summary representation, choose the sentence that contains the first appearance of a chain member in the text.
2) For each chain in the summary representation, choose the sentence that contains the first appearance of a representative chain member in the text.
3) For each chain, find the text unit where the chain is highly concentrated. Extract the sentence with the first chain appearance in this central unit.

Concentration is computed as the number of chain members occurrences in a segment divided by the number of nouns in the segment.

Algorithm for computing Lexical chains:
Start.
For all candidate words do
Expand the words into possible senses (S_1, S_2,….S_n).
Determine the "offset " for each sense in WordNet .
End for
For all senses do
Insert the sense into the respective element of the synsetlD list
If inserted synset has relations with already inserted synsets then
Identify the relation and determine their score
End if
End for
For all relations do
Identify the chains compatible with the current relation
If compatible chain is found then
Insert into chain by looking out for repetition
Update the chainscore
Else create a new chain for the relation.
End if
End for
Sort the chains in descending order based on the chainscores
For all chains do
For all chainmembers do
If chain member already assigned a sense then
If assigned sense is not equal to the current chainmember sense then
FLG $\leftarrow$ FALSE
End if
else
Assign the chain member the sense temporarily
End if
End for
If FLG equals to FALSE then
Discard the chain
else
assign the chain members their respective senses from the temporarily stored values
retain the chain
end if
end for
Stop

E. Event Segmentation and summarization

A unique feature of summarization approach is the introduction of the event segmentation process to extract the semantic construct "event" before summarization.

F. Story Boundary Detection

Story boundary detection (or story segmentation) is used to identify where one story ends and the other story begins in a stream of text. It serves as a necessary precursor to various tasks, such as topic detection and tracking, information extraction, indexing, retrieval and summarization. A typical broadcast news retrieval system can locate the particular positions in a repository that match the user's query, but lack the ability of determining where the user-interested stories begin and end [20]. The data corpus consists of a collection of information, then using the classification techniques to create the boundary between two different news documents. The input from the event segmentation is detected core parts and from that, we have to identify the endpoints between the documents.

The story boundary detection is used to create the boundary between the two different news documents. Initially the end points are identified in the summarized documents and the boundary is created. This step is useful to display the news documents efficiently and helps the user to know different news.

IV. CONCLUSIONS

Many news documents related to same topic are posted by different authors and their opinions vary during the topic life span. The summarization method is used to help the user to obtain the news from different documents.

In this paper, we have presented a theme encapsulation and content frame work (TECF) using lexical chains, which extracts the themes, events and connects the associated events to form evolution graph..TECF can produce highly representive summaries that are composed by the experts.

REFERENCES

[1] A.V.Seetha Lakshmi and Dr. S.P.Victor, " TECF:Accomplishment of Content Framework Tactic to Temporal Theme Condensation", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 11, November 2013.

[2] Chien Chin Chen and Meng Chang Chen,"*TSCAN: A Content* Anatomy Approach to Temporal Topic Summarization",IEEE Transactions onKnowledge and Data EngineeringVolume:24Issue:1

[3] F.Y.Y.Choi,P.Wiemer-Hastings, and J.Moore,"Latent Semantic Analysis for Text Segmentation," Proc. Conf. Empirical Methods in Natural Language Processing, pp.109-117,2001.

[4] D.M. Blei and P.J. Moreno, "Topic Segmentation with an Aspect Hidden Markov Model," Proc. 24th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval,pp. 343-348, 2001.

[5] X. Ji and H. Zha, "Domain-Independent Text Segmentation Using Anisotropic Diffusion and Dynamic Programming," Proc.26th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 322-329, 2003.

[6] Y. Gong and X. Liu, "Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis," Proc. 24th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 19-25, 2001.

[7] D. Shen, J.T. Sun, H. Li, Q. Yang, and Z. Chen, "Document Summarization Using Conditional Random Fields," Proc. 20th Int'lJoint Conf. Artificial Intelligence (IJCAI '07), pp. 2862-2867, 2007.

[8] T. Nomoto and Y. Matsumoto, "A New Approach to Unsupervised Text Summarization," Proc. 24th Ann. Int'l ACM SIGIR Conf.Research and Development in Information Retrieval, pp. 26-34, 2001.

[9] J. Allan, R. Gupta, and V. Khandelwal, "Temporal Summaries of News Topic," Proc. 24th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 10-18, 2001.

[10] G.Erkan and D.R. Radev, "LexRank:Graph-Based Centrality as Salience in Text Summarization," J. Artificial Intelligence Research, vol. 22, pp. 457-479, 2004.

[11] J. Kleinberg, "Bursty and Hierarchical Structure in Streams," Proc.Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and DataMining, pp. 91-101, 2002.

[12] R. Nallapati, A. Feng, F. Peng, and J. Allan, "Event Threading within News Topics," Proc. 13th ACM Int'l Conf. Information and Knowledge Management, pp. 446-453, 2004.

[13] C.C. Yang and X. Shi, "Discovering Event Evolution Graphs from Newswires," Proc. 15th Int'l Conf. World Wide Web, pp. 945-946,2006.

[14] A. Feng and J. Allan, "Finding and Linking Incidents in News,"Proc. 16th ACM Conf. Information and Knowledge Management,pp. 821-830, 2007.

[15] R. Swan and J. Allan, "Automatic Generation of OverviewTimelines," Proc. 23rd Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 49-56, 2000.

*[16] Web Crawlers & Hyperlink Analysis.*Albert Sutojo. CS267 – Fall 2005. Instructor :Dr. T.YLin.

[17] A. Nenkova, L. Vanderwende, and K. Mckeown, "A Compositional Context Sensitive Multi-Document Summarizer: Exploringthe Factors that Influence Summarization," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval,pp. 573-580, 2006.

[18] L E. Spence, A.J. Insel, and S.H. Friedberg, Elementary Linear Algebra, a Matrix Approach.Prentice Hall, 2000.

[19] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval.Addison Wesley, 1999.

[20] Broadcast News Story Boundary Detection Using Visual, Audio And TextfeaturesMaryamDaneshi, Matt Yu

[21] Kleinberg,J. "Authoritative Sourcesin a Hyperlinked Envi- ronment," in Proceedings of the ninth annual ACM-SIAMsymposiumon Discretealgorithms,1998,668-677.

[22] Nicholas, C.,andDahlberg,R.,"Spotting TopicswiththeSin- gularValue Decomposition," Lecture Notes in ComputerScience, Vol.1481, 1998,82-91.

[23] Rabiner,L.R.,andSambur,M.R."AnAlgorithmfor DeterminingtheEndpointsforIsolated Utterances,"intheBellSystem Technical Journal, Vol.54,No.2, Feb.1975,pp297-315.

[24] Rabiner,L.R., andSchafer,R. W. "DigitalProcessingofSpeech Signals," Prentice-Hall, 1978.

