Context-Based Similarity Analysis for Document Summarization

[1]S.Prabha, [2]Dr.K.Duraiswamy, [3]B.Priyanga
[1]Associate Professor, Department of Information Technology
K.S.Rangasamy College of Technology, Tiruchengode – 637215, Tamil Nadu, India
[2]Dean Academic,
K.S.Rangasamy College of Technology, Tiruchengode – 637215, Tamil Nadu, India
[3]PG Scholar, Department of Information Technology
K.S.Rangasamy College of Technology, Tiruchengode – 637215, Tamil Nadu, India

Abstract— **Context-Based similarity analysis for document summarization extracts a condensed version of the original document in the information retrieval task. The document summarization mainly uses the similarity between sentences in the document to extract the most salient sentences. A document summary is useful to give an overview of the original document in a shorter period of time. The sentence similarity values remain independent of the context. The context is not taken into consideration for the document as well as the sentences are indexed using traditional term indexing measures. Context sensitive document indexing model based on the Bernoulli model of randomness is used for document summarization process. The lexical association between terms is used to produce a context sensitive weight to the document terms. The context sensitive indexing weights are used to compute the sentence similarity matrix and as a result, the informative sentences are presented on the top of the summary. The quality of the summary is to make a positive impact.**

Index Terms—**Document Summarization, Document indexing, Lexical association, Bernoulli model of randomness.**

I. Introduction

Modern text retrieval systems principally rely on orthographic, semantic, and statistical analysis. The main goal of a summary is to present the main ideas in a document/set of documents in a short and readable paragraph. Multidocument summarization is the task of producing summary from many documents. The usual approach is to use white space to identify the boundaries of the words, stemming is followed to conflate words with similar surface forms into a common term. Topic summarization deals with the evolution of topics in addition to providing the informative sentences. A weight is then computed for each term in every document using the frequency of the term in the document, the selectivity of the term, and the length of the document. In document the queries are represented in similar manner is the vector space text retrieval, and the collection of the query is the similarity of each document is then computed as the normalized document of the inner product and query term weight vectors. The probabilistic text retrieval with its term weight is treated as the probability of a document relevance to a query. In Boolean text retrieval probabilistic and vector space techniques are often combined, with the presence or absence of a term or the combination of terms can be explicitly required in the query specification. Boolean approach is that lists of documents that are ranked in order of decreasing probability of relevance allow users to interactively decide how many documents are worth examining with the principal advantage of vector space and probabilistic text retrieval. When no user interaction is possible before the next processing stage the unranked Boolean technique is used.

The text retrieval system depends on the utility of a query in which it is constructed, and that depends on the user understands the collection and the way in which the indexed features can be used for selecting the documents. The relevant documents are usually straight forward, but the relevant documents with the interactive inspection by the user are generally needed and will be more carefully separated from the irrelevant ones. Simulated Nucleation can be used to speed the process for an iterative query reformulation process, to produce a query with the better separates relevant and irrelevant documents with the leveraging inspection of an few documents.

To operate without user interaction the information extraction algorithm is designed, making the sharp decisions is rather than producing ranked lists of candidates. Results from the Message Understanding Conferences (MUC) suggest with the simple tasks such as the date extraction is done with the large documents and named the entity recognition which can be done quite accurately, but for the more complex tasks relationships among data items must also be extracted are still somewhat error-prone for the documents. The document summarization is making the large document into small, where it saves the time for typing and reading the documents. The document summarization is done with the analysis of term and concept. The term analysis with the analysis of term frequency and inverse document frequency with the term weight. The semantic analysis is

Manuscript received April, 2014.
S.Prabha,*Department of Information Technology, K.S.Rangasamy College of Technology) Namakkal, TN, India,*
Dr. K. Duraiswamy, Dean, *K.S.Rangasamy College of Technology) Namakkal, TN, India,*

B.Priyanga, *Department of Information Technology, K.S.Rangasamy College of Technology) Namakkal, TN, India,*

done with the analysis of a semantic weight. The semantic analysis is done by creating the ontology. The ontology with the concept relations of synonym, meronym, hypernym.

II. Proposed System

The main goal of a summary is to present the document with the main ideas in a short and readable paragraph. The Summaries can be produced from a single document or many documents. The task of producing summary from many documents is called multi-document summarization. Summarization can also be specific to the information needs of the user, thus called "query-biased" summarization. For instance, the QCS system (query, cluster, and summarize) retrieves relevant documents in response to a query, clusters these documents by topic and produces a summary for each cluster. Opinion summarization is another application of text summarization. Topic summarization deals with the evolution of topics in addition to providing the informative sentences.

This paper focuses on sentence extraction-based single document summarization. The previous studies on the sentence extraction is mainly based on the text summarization task in which it uses a graph-based algorithm to calculate the saliency of each sentence in a document and the most salient sentences are extracted to build the document summary. The sentence extraction techniques give an indexing weight to the document terms and use these weights to compute the sentence similarity and/or document centroid and so on. The sentence similarity calculation remains central to the existing approaches.

The indexing weights of the document terms are utilized to compute the sentence similarity values. Elementary document features are used to allocate an indexing weight to the document terms, which include the document length, term frequency, occurrence of a term in a background corpus. Therefore, the indexing weight of the other terms appearing in the document remains independent and the context in which the term occurs is overlooked in assigning its indexing weight for the documents. This results in "context independent document indexing." To the authors' knowledge, no other work in the existing literature addresses the problem of "context independent document indexing" for the document summarization task.

A document contains both the background terms as well as the content-carrying terms. In the sentence similarity analysis the traditional indexing schemes cannot distinguish between these terms. The higher weight is given by the context sensitive document indexing model to the topical terms where it is compared with the nontopical terms and thus influences the sentence similarity values in a positive manner.

Using the lexical association between document terms the system considers the problem of "context independent document indexing. The content carrying words will be highly associated with each other in a document, while the background terms will have very low in association with the other terms in the document. The association between terms is stated in this paper by the lexical association and is computed through the corpus analysis.

The context in which a word appears provides useful information about its meaning is the main motivation which is behind using the lexical association of the central assumption. The Co-occurrence measures observe the distributional patterns of a term in the document with other terms in the vocabulary and have applications in many tasks pertaining to natural language understanding such as word classification, knowledge acquisition, word sense disambiguation, information retrieval sentence retrieval and word clustering. In this paper, we derive a novel term association metric using the Bernoulli model of randomness. Multivariate Bernoulli models have previously been applied to document indexing and information retrieval. The Bernoulli model of randomness is used to find the probability of the co-occurrences of two terms in a corpus and use the classical semantic information theory to quantify the information contained in the co-occurrences of these two terms.

The lexical association metric, thus, derived is used to propose a context-sensitive document indexing model. The PageRank-based algorithm is applied for implementing the iteratively compute and how informative is each document term. Sentence similarity is calculated using the context sensitive indexing where it should reflect the contextual similarity between two sentences in the documents. Depending on the context the two sentences have different similarity values. The improvements in the document summarization are the hypothesis of the similarity measure.

For the single document summarization task the text summarization experiments have been performed over the DUC01 and DUC02 data sets. It has been shown that the proposed model which consistently improves the performance of the baseline sentence extraction algorithms under the various settings and, thus, can be used as an enhancement over the baseline models. The theoretical foundations along with the empirical results confirm that the proposed model advances the state of the art in document summarization. The main contributions of the system are summarized as follows:

1. To propose the novel idea of using the context-sensitive document indexing to improve the sentence extraction-based document summarization task.
2. To implement the idea by using the lexical association between document terms in a PageRank-based framework. A novel term association metric using the Bernoulli model of randomness has been derived for this purpose. Empirical evidence has been provided to show that using the derived lexical association metric, average lexical association between the terms in a target summary is higher compared to the association between the terms in a document.
3. Experiments have been conducted over the benchmark document understanding conference (DUC) data sets to empirically validate the effectiveness of the proposed model.

The underlying hypotheses of our approach in this section [31].

1. A document summary is centered around the topical terms (content-carrying terms) encountered in the

document. In other words, H1: "The ratio of topical words is higher in a summary of a document than in the original document."

2. The nontopical terms appear randomly across all the documents while topical terms appear in bursts. H2: "For a carefully chosen lexical association metric, lexical association between two topical terms should be higher than the lexical association between two nontopical terms or a pair of topical and nontopical terms." This lexical association can be calculated using a large corpus.
3. Once the lexical association is calculated, we can construct the document graph, with the terms appearing in the document as the vertices and the lexical association between these terms as the edges of the graph.H3: "A PageRank-based algorithm can be used to determine the context-sensitive indexing weights, resulting in performance improvement for a document summarization task."

The co-occurrence patterns in a corpus can be used to derive the lexical association measure. Assuming that the terms are distributed according to the Bernoulli distribution, divergence from the randomness behavior can provide a measure of the lexical association.

III. Sentence Similarity And Word Indexing

A. Bernoulli Model of Randomness

By using the PMI measure the lexical association between documents terms is higher than between the summary terms. Therefore, the PMI measure may not be a suitable choice for the possible application in document summarization. Using the MI and Bernoulli measure, on the other hand, the average lexical association between the terms in human summary is higher than that in the original document. As verified by the two different statistical tests, the difference is statistically significant using both these association measures and therefore, the hypothesis holds true for both the MI and Bernoulli measures. However, the significance level as well as the ratio of average lexical association between the target summary and original document is much higher for the Bernoulli measure as compared to the MI measure. Thus, the proposed Bernoulli measure is a better fit for H_2.

B. Context-Based Word Indexing

Given the lexical association measure between two terms in a document from hypothesis H_2, the next task is to calculate the context sensitive indexing weight of each term in a document using hypothesis H_3. A graph-based iterative algorithm is used to find the context sensitive indexing weight of each term. Given a document D_i, a document graph G is built. Let G = (V,E) be an undirected graph to reflect the relationships between the terms in the document D_i. V = $\{v_j|1 \leq j \leq\_ |V|\}$ denotes the set of vertices, where each vertex is a term appearing in the document. E is a matrix of dimensions |V| × |V|. Each edge e_{jk} ε E corresponds to the lexical association value between the terms corresponding to the vertices v_j and v_k. The lexical association between the same terms is set to 0.

C. Sentence Similarity Using the Context-Based Indexing

The model described above gives a context-sensitive indexing weight to each document term. The next step is to use these indexing weights to calculate the similarity between any two sentences. Given a sentence s_j in the document D_i, the sentence vector is built using the index$W_{t()}$. The sentence vector s_j is calculated such that if a term v_k appears in s_j, it is given a weight index$W_t(v_k)$, otherwise, it is given a weight 0. The similarity between two sentences sj and sl is computed using the dot product, i.e., sim(s_j, s_l) = $\hat{s}_j$ - $\hat{s}_l$. Besides using the new sentence similarity measure, the paradigm as presented in Wan and Xiao [1], described is used for calculating the score of the sentences. The proposed method will be denoted by "bern" corresponding to the "Bernoulli" measure.

IV. System Model

The document indexing and summarization scheme is enhanced with semantic analysis mechanism. Context sensitive index model is improved with semantic weight values. Concept relationship based lexical association measure estimation is performed for index process. Bernoulli lexical association measure is used to perform the document classification process.

The document summarization system is enhanced with document classification process. Concept relationship based semantic weight estimation mechanism is used for document relationship analysis. Ontology based semantic index scheme is used to perform the classification process. The system is divided into five major modules. They are document preprocess, term index process, semantic index process, document summarization, document classification.

The document preprocess module is designed to perform token separation and frequency estimation process. Term indexing process module is designed to estimation term weights and index process. Concept relationship is analyzed under semantic index process. Document summarization module is designed to prepare document summary. Document category assignment is performed under the document classification process.

A. Document Preprocess

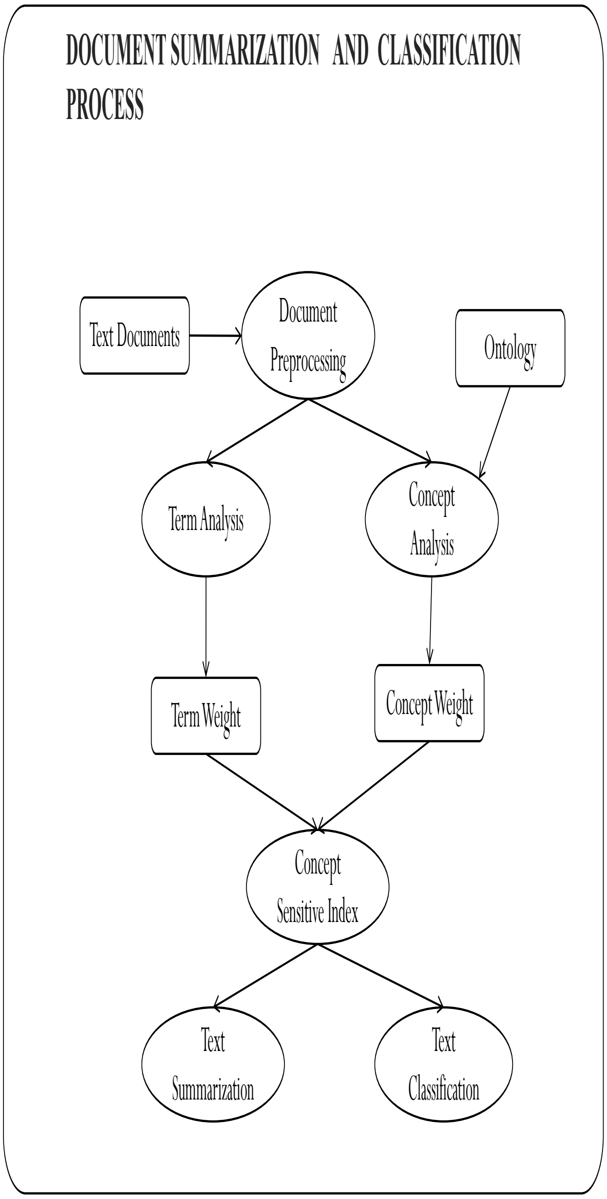
The document preprocess is performed to parse the documents into tokens. Stop word elimination process is applied to remove irrelevant terms. Stemming process is applied to carry term suffix analysis. Document vector is constructed with terms and their count values.

- Stop words:

Commonly used words that are excluded from searches to help index and search web pages faster. Some examples of **stop words** are **a**, **and**, **but**, **how**, **or**, and **what**

- Stemming:

A stemming algorithm reduces the words "fishing", "fished", and "fisher" to the root word, "fish"



B. Term Index Process

Statistical weight estimation process is applied with term and its count values. Term weight estimation is performed with Term Frequency (TF) and Inverse Document Frequency (IDF) values. Context sensitive index model uses the term weights for term index process. Latent semantic analysis is applied to estimate relationship values.

$$A.\ TF = \frac{No.of\ word\ count\ in\ docum}{Total\ no.of\ words\ in\ docum}$$

$$A.\ IDF = \frac{Total\ no.of\ documents}{Terms\ appear\ in\ a\ docum}$$

$$A.\ Term\ Weight = TF * I$$

C. Semantic Index Process

Ontology is a repository that maintains the concept term relationships. Semantic weights are estimated using concept relations. Synonym, hypernym and meronym relationships are used in the concept analysis. Context sensitive index model uses the semantic weight values for index process.

synonym value
: 0.6

meronym value : 0.4

hypernym value : 0.2

D. Document Summarization

Lexical association between terms is used to produce context sensitive weight. Weights are used to compute the sentence similarity matrix. The sentence similarity measure is used with the baseline graph-based ranking models for sentence extraction. Document summary is prepared with sentence similarity values.

E. Document Classification

Document classification is carried out to assign document category values. Term weight and semantic weights are used for the classification process. Context sensitive index is used for the document classification process. Sentence similarity is used in classification process.

V. PERFORMANCE ANALYSIS

The document summarization system is developed to prepare the summary for the text documents. Term weight based context sensitive index and semantic weight based context sensitive index models are used for the document summarization process. Document contents are preprocessed and sentence based similarity analysis is performed to estimate the context sensitive index values. Most important sentences are summarized with reference to the index values. The system is tested with different document count and weight schemes. The Context Sensitive Index with Term weights (CSIT) and Context Sensitive Index with Semantic weights (CSIS) schemes are used for the summarization process.

S. No	Documents	CSIT	CSIS
1	200	73%	84%
2	400	75%	86%
3	600	78%	89%
4	800	80%	92%
5	1000	82%	94%

Table No: 1. Summarization Relevancy Analysis between Context Sensitive Index with Term weights (CSIT) and Context Sensitive Index with Semantic weights (CSIS) schemes

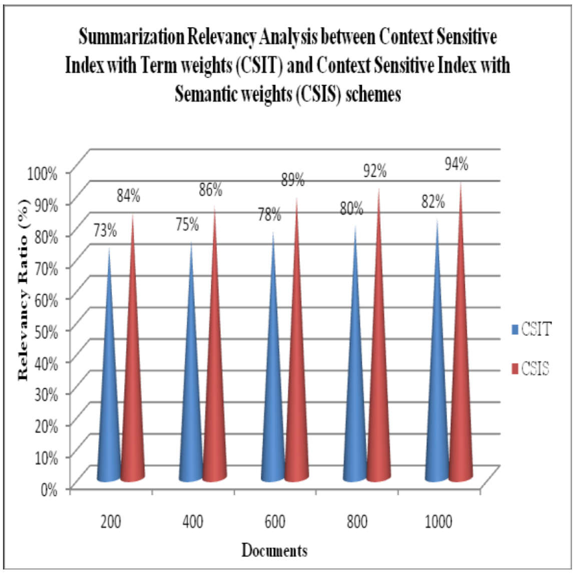


Figure No: 1. Summarization Relevancy Analysis between Context Sensitive Index with Term weights (CSIT) and Context Sensitive Index with Semantic weights (CSIS) schemes

S. No	**Documents**	**CSIT**	**CSIS**
1	200	68.42	81.22
2	400	70.78	82.89
3	600	71.96	84.36
4	800	74.13	86.13
5	1000	76.57	87.95

Table No: 2. Summarization Accuracy Analysis between Context Sensitive Index with Term weights (CSIT) and Context Sensitive Index with Semantic weights (CSIS) schemes

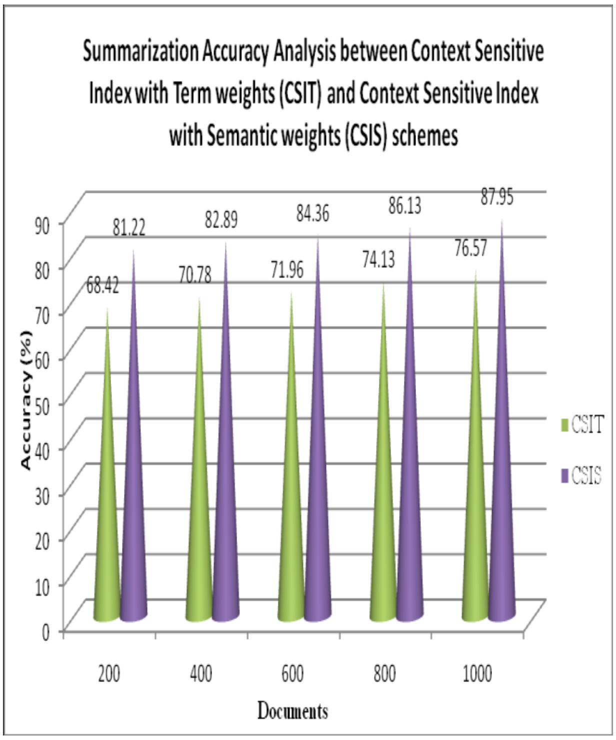


Figure No: 2. Summarization Accuracy Analysis between Context Sensitive Index with Term weights (CSIT) and Context Sensitive Index with Semantic weights (CSIS) schemes

S. No	Documents	CSIT	CSIS
1	200	66.73	82.47
2	400	69.05	85.14
3	600	71.75	87.98
4	800	74.23	91.23
5	1000	76.68	94.56

Table No: 3. Classification Accuracy Analysis between Context Sensitive Index with Term weights (CSIT) and Context Sensitive Index with Semantic weights (CSIS) schemes

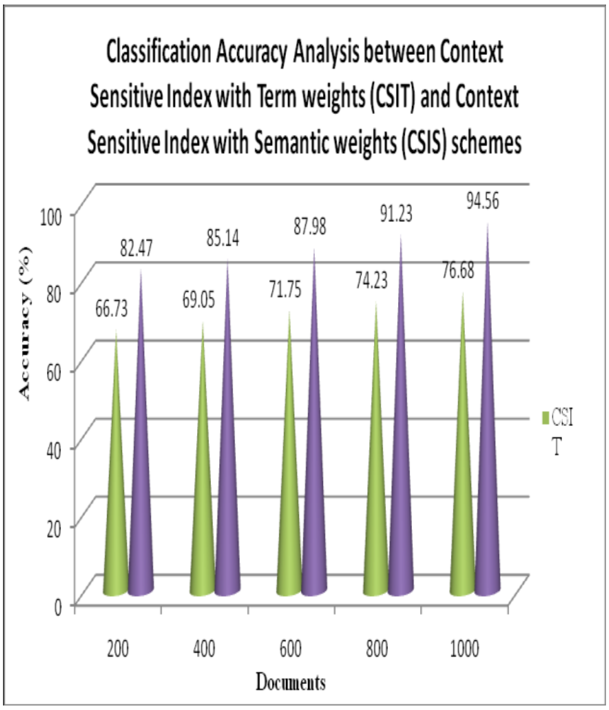


Figure No: 3. Classification Accuracy Analysis between Context Sensitive Index with Term weights (CSIT) and Context Sensitive Index with Semantic weights (CSIS) schemes

VI. Conclusion

In this paper Document summarization methods are used to extract the condensed version of the original document. Document classification methods are used to assign the category of the documents Bernoulli model of randomness is used for document summarization process. The Bernoulli model of randomness is used to find the probability of the co-occurrences of two terms in a large corpus. The lexical association between terms is used to produce a context sensitive weight to the document terms. The document indexing and summarization scheme is enhanced with semantic analysis mechanism. Context sensitive index model is improved with semantic weight values. Concept relationship based lexical association measure estimation is performed for index process. Bernoulli lexical association measure is used to perform the document classification process. The Java language and Oracle relational database are used for the system development process. The proposed model gives a higher weight to the content-carrying terms and as a result, the sentences are presented in such a way that the most informative sentences appear on the top of the summary, making a positive impact on the quality of the summary.

REFERENCES

[1] X. Wan and J. Xiao, “Exploiting Neighborhood Knowledge for Single Document Summarization and Keyphrase Extraction,” ACM Trans. Information Systems, vol. 28, pp. 8:1-8:34, http://doiacm.org/10.1145/1740592.1740596, June 2010.

[2] P. Goyal, L. Behera, and T. McGinnity, “Query Representation Through Lexical Assoc. for Information Retrieval,” IEEE Trans. Knowledge and Data Eng., vol. 24, no. 12, pp. 2260-2273, Dec. 2011.

[3] L.L. Bando, F. Scholer, and A. Turpin, “Constructing Query- Biased Summaries: A Comparison of Human and System Generated Snippets,” Proc. Third Symp. Information Interaction in Context.

[4] X. Wan, “Towards a Unified Approach to Simultaneous Single-Document and Multi-Document Summarizations,” Proc. 23rd Int’l Conf. Computational Linguistics,http://portal.acm.org/citation.cfm?id=1873781. 1873909 , 2010.

[5] Y. Ouyang, W. Li, Q. Lu, and R. Zhang, “A Study on Position Information in Document Summarization,” Proc. 23rd Int’l Conf. Computational Linguistics: Posters, pp. 919-927, http://portal.acmorg/citation.cfm?id=1944566.1944672, 2010.

[6] Q.L. Israel, H. Han, and I.-Y. Song, “Focuse Multi-Document Summarization: Human Summarization Activity vs. Automated Systems Techniques,” J. Computing Sciences in Colleges, vol. 25, pp.10http://portal.acm.org/citation.cfm?id=17417137 1747140, May 2010.

[7] H. Nishikawa, T. Hasegawa, Y. Matsuo, and G. Kikui, “Opinion Summarization with Integer Linear Programming Formulation for Sentence Extraction and Ordering,” Proc. 23rd Int’l Conf. Computational Linguistics: Posters, http://portal.acm.org/citation.cfm?id=1944566.1944671, 2010.

[8] C.C. Chen and M.C. Chen, “TSCAN: A Content Anatomy Approach to Temporal Topic Summarization,” IEEE Trans. Knowledge and Data Eng., vol. 24, no. 1, pp. 170-183, Jan. 2012.

[9] Pawan Goyal, Laxmidhar Behera, and Thomas Martin McGinnity, “A Context-Based Word Indexing Model for Document Summarization”,

IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 8, August 2013

[10] S. Harabagiu and F. Lacatusu, "Using Topic Themes for Multi-Document Summarization," ACM Trans. Information Systems, vol. 28, pp. 13:1-13:47, http://doi.acm.org/10.1145/1777432.1777436, July 2010.

[11] H. Daume´ III and D. Marcu, "Bayesian Query-Focused Summarization,"Proc. 21st Int'l Conf. Computational Linguistics and the 44th Ann. meeting of the Assoc. for Computational Linguistics,pp. 305-312, http://dx.doi.org/10.3115/1220175.1220214, 2006.

[12] D.M. Dunlavy, D.P. O'Leary, J.M. Conroy, and J.D. Schlesinger, "QCS: A System for Querying, Clustering and Summarizing Documents,"Information Processing and Management, vol. 43, pp. http://portal.acm.org/citation.cfm?id=1284916.

1285163, Nov. 2007.

[13] R. Varadarajan, V. Hristidis, and T. Li, "Beyond Single-Page Web Search Results,"IEEE Trans. Knowledge and Data Eng., vol. 20, no. 3, pp. 411-424, Mar. 2008.

[14] L.-W. Ku, L.-Y. Lee, T.-H. Wu, and H.-H. Chen, "Major Topic Detection and Its Application to Opinion Summarization,"Proc.28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval.

[15] E. Lloret, A. Balahur, M. Palomar, and A. Montoyo, "Towards Building a Competitive Opinion Summarization System: Challenges and Keys," Proc. Human Language Technologies: The 2009 Ann. Conference of the North Am. Ch. Assoc. for ComputationalLinguistics, Companion Vol. : Student Research Workshop and Doctoral Consortium, pp. http://portal.acm.org/citation.cfm?id=1620932.1620945, 2009.

[16] G. Conrad, J.L. Leidner, F. Schilder, and R. Kondadadi, "QueryBased Opinion Summarization for Legal Blog Entries,"Proc. 12th Int'l Conf. Artificial Intelligence and Law.

[17] H. Nishikawa, T. Hasegawa, Y. Matsuo, and G. Kikui, "Opinion Summarization with Integer Linear Programming Formulation for

Sentence Extraction and Ordering,"Proc. 23rd Int'l Conf. Computational Linguistics: Posters, pp. 910-918, http://portal.acm.org/ citation.cfm?id=1944566.1944671, 2010.

[18] C.C. Chen and M.C. Chen, "TSCAN: A Content Anatomy Approach to Temporal Topic Summarization,"IEEE Trans. Knowledge and Data Eng.,vol. 24, no. 1, pp. 170-183, Jan. 2012.

[19] D.R. Radev, H. Jing, M. Stys, and D. Tam, "Centroid-Based Summarization of Multiple Documents," Information Processing and Management, vol. 40, pp. 919-938, http://portal.acm.org/ citation.cfm?id=1036118.1036121, Nov. 2004.

[20] Z. Harris,Mathematical Structures of Language.Wiley, 1968.

[21] K. Morita, E.-S. Atlam, M. Fuketra, K. Tsuda, M. Oono, and J.-i. Aoe, "Word Classification and Hierarchy using Co-Occurrence Word Information,"Information Processing and Management, vol. 40, pp. 957-972, http://portal.acm.org/citation.cfm?id=1036118.1036123, Nov. 2004.

[22] T. Yoshinari, E.-S. Atlam, K. Morita, K. Kiyoi, and J.-i. Aoe, "Automatic Acquisition for Sensibility Knowledge Using CoOccurrence Relation,"Int'l J. Computer Applications in Technology,vol. 33, pp. 218-225, http://portal.acm.org/citation.cfm?id=1477782.1477797, Dec. 2008.

[23] B. Andreopoulos, D. Alexopoulou, and M. Schroeder, "Word Sense Disambiguation in Biomedical Ontologies with Term CoOccurrence Analysis and Document Clustering,"Int'l J. Data Mining and Bioinformatics,vol. 2, pp. 193-215, http://portal.acmorg/citation.cfm?id=1413934.1413935, Sept. 2008.

[24] P. Goyal, L. Behera, and T. McGinnity, "Query Representation Through Lexical Assoc. for Information Retrieval,"IEEE Trans.Knowledge and Data Eng.,vol. 24, no. 12, pp. 2260-2273, Dec. 2011.

[25] K. Cai, C. Chen, and J. Bu, "Exploration of Term Relationship for Bayesian Network Based Sentence Retrieval,"Pattern Recognition

Letters,vol. 30, no. 9, pp. 805-811, 2009.

[26] H. Li, "Word Clustering and Disambiguation Based on CoOccurrence Data,"Nat'l Language Eng.,vol. 8, pp. 25-42, http://portal.acm.org/citation.cfm?id=973860.973863, Mar. 2002.

[27] G. Amati and C.J. Van Rijsbergen, "Probabilistic Models of Information Retrieval Based on Measuring the Divergence from Randomness,"ACM Trans. Information Systems,vol. 20, pp. 357-389, http://doi.acm.org/10.1145/582415.582416, Oct. 2002.

[28] D.E. Losada and L. Azzopardi, "Assessing Multivariate Bernoulli Models for Information Retrieval,"ACM Trans. Information Systems, vol. 26, pp. 17:1-17:46, http://doi.acm.org/10.1145/1361684.1361690, June 2008.

[29] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web," technical report,

Stanford Digital Library Technologies Project, http://citeseer.ist.psu.edu/page98pagerank.html, 1998.

[30] J. Turner and E. Charniak, "Supervised and Unsupervised Learning for Sentence Compression,"Proc. 43rd Ann. Meeting on

Assoc. for Computational Linguistics,pp. 290-297,http://dx.doi.org/10.3115/1219840.1219876, 2005.

[31] Pawan Goyal, Laxmidhar Behera, "A Context-Based Word Indexing Model for Document Summarization", IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 8, August 2013.

[32] Kenton W. Murray, "Summarization by Latent Dirichlet Allocation: Superior Sentence Extraction through Topic Modeling", April 17, 2009.

[33] Karen Sprck Jones. Automatic summarising: The state of the art. Information Processing and Management, 43(6):1449 1481, 2007. Text Summarization.

[34] Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. In Proceedings of the Workshop on Text Summarization

B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (references)