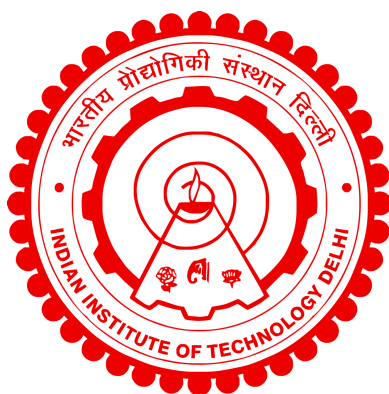


**MULTI-DOCUMENT ABSTRACTIVE TEXT
SUMMARIZATION USING MULTI-SENTENCE
COMPRESSION AND TEXT SIMPLIFICATION**

RAKSHA AGARWAL



**DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY DELHI
OCTOBER 2022**

© Indian Institute of Technology Delhi (IITD), New Delhi, 2022

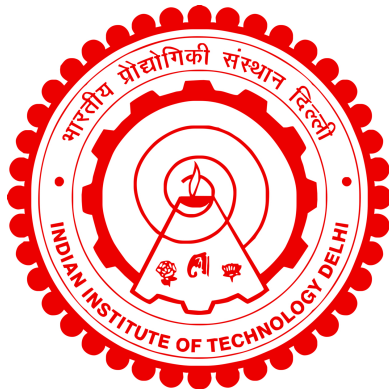
**MULTI-DOCUMENT ABSTRACTIVE TEXT
SUMMARIZATION USING MULTI-SENTENCE
COMPRESSION AND TEXT SIMPLIFICATION**

by

**RAKSHA AGARWAL
DEPARTMENT OF MATHEMATICS**

Submitted

in fulfillment of the requirements of the degree of
DOCTOR OF PHILOSOPHY
to the



**Indian Institute of Technology Delhi
October 2022**

CERTIFICATE

This is to certify that the thesis entitled, **Multi-Document Abstractive Text Summarization using Multi-Sentence Compression and Text Simplification** being submitted by **Ms. Raksha Agarwal** for the award of the degree of **Doctor of Philosophy** is a record of *bona fide* research work carried out by her in the Department of Mathematics of Indian Institute of Technology Delhi.

Ms. Raksha Agarwal has worked under my guidance and supervision. She has fulfilled the requirements for the submission of this thesis, which to my knowledge has reached the requisite standard. The results contained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree or diploma. .

Date:

Prof. Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology Delhi
New Delhi-110016, India

ACKNOWLEDGEMENTS

I am deeply indebted to Mother Saraswati; the Goddess of Knowledge. Due to Her divine grace I have been able to carry out this research work.

This thesis would not have been possible without the support of many people. It is impossible to acknowledge all of them here, but there are few people who deserve a special mention. I am forever indebted to Prof. Niladri Chatterjee for providing me an opportunity to carry out the Ph.D. work under his supervision at IIT Delhi. It is a privilege to work with a person who can guide your research and also train you for all the aspects which Ph.D. entails. His unwavering support has been the driving force of motivation even during the most difficult times. He taught me research writing, research methodologies and most importantly he helped me out in dealing with failures in experiments that came several times during this duration.

I am extremely grateful to the Department of Mathematics and IIT Delhi authorities for providing me with excellent state of the art facilities and infrastructure to work in during my Ph.D. program. I thank Head, Department of Mathematics and DRC Chairperson for their help at various stages of my research work. I am especially thankful to my Student Research Committee members: Prof. B. S. Panda, Prof. Aparna Mehra and Prof. Sumeet Agarwal for their insightful inputs, which strengthened my research work.

I would like to acknowledge the Council of Scientific and Industrial Research, India, for providing financial support through their Shyama Prasad Mukherjee (SPM) fellowship program during my research period.

I would like to use this platform to thank the wonderful people I met during my Ph.D. and who made this entire learning experience in IIT Delhi productive and fun. I would like to thank my senior colleagues, Dr. Shradhha Chaudhary, Dr. Nidhika Yadav, Dr. Susmita Gupta, Dr. Kartikay Gupta, Dr. Anubha Goel, Dr. Puneet Parsicha, Dr. Lipy and Dr. Sanjay Kumar for the academic as well as non-academic guidance they provided to me. I would also like to thank Dr. Priyamvada, Ms. Soniya Takshak, Ms. Pooja, Dr. Pooja Goyal, Dr. Hariom, Mr. Navnit Yadav, Dr. Pradeep Singh, Dr. Vaibhav and my juniors Ms. Shivani Chaudhary, Ms. Kushagri Tandon, Mr. Aayush Singha Roy. I am likewise thankful to all those who have directly or indirectly helped me to finish my dissertation study.

No amount of thanks and credit is sufficient enough to reflect the support I received from Mr. Pratik Sinha in this academic endeavour. I would like to extend my special thanks to Mrs. Archana Kumari. She was my pillar of strength through thick and thin of the Ph.D. journey. I would also like to thank Ms. Surbhi for the unfailing support and motivation she provided me. I thank Ms. Ambika Sharma for unconditional support whenever I approached her for the same.

All my achievements are outcomes of the silent sacrifice and patience of my family. First and foremost, I thank my father Late Mr. Kailash Kumar Agarwal and my sister Late Mrs. Rashmi Aggarwal. You left this earth too soon but I know you are looking upon me from heaven. I do not find words to thank my mother Mrs. Manju Agarwal for the love, care, and support that she has given me all my life. I would like to thank her for allowing me to progress on the path of research while not being bothered by my duties as a daughter. I wish to thank my brothers Mr. Rajiv Sanwatia and Mr. Ravi Sanwatia for being immensely supportive and having my back at all time. I also thank my brother-in-law Mr. Amit Aggarwal and sister-in-laws Mrs. Kiran Sanwatia and Mrs. Suruchi Sanwatia for encouraging me throughout this journey. I thank Yuvika, Toshank, Tishika, AaroHi, Vedika and Aarav for the joy they have brought in our life. I express my gratitude to Mr. Rajesh Agarwal and his family for supporting me selflessly. It would not have been possible for me to pursue this academic life without their help.

At last, I feel extremely thankful to almighty for giving me such a life where I have been able to make choices that I have made and pursue the goal of higher knowledge and wisdom.

Raksha Agarwal

ABSTRACT

Automatic Text Summarization systems are of high demand due to the overwhelming amount of textual information generated on the internet every minute. Historically sentence ranking approach used in Extractive text summarization has received huge attention from NLP researchers. However, different shortcomings associated with extraction has of late created a lot of enthusiasm among the researchers towards Abstractive summarization. But this task is quite challenging because it requires deep understanding of the input text and intensive natural language generation techniques for preserving the information and semantics of the input. The focus of the present research is on performing multi-document abstractive text summarization using two sentence rewriting techniques, namely, multi-sentence compression and sentence simplification.

Our research tries to find answer to several queries pertaining to Abstractive text summarization: a) Effective representation of sentences to facilitate summary generation. b) Examining the usefulness of syntactic simplification as a pre-processing step before employing summarization techniques c) Developing an appropriate sentence selection process from the simplified text both in a deterministic way, and in a fuzzy way. In particular, the latter is used in the event of query-focused summarization.

A Word Graph-based representation of sentence clusters is used for generation of multi-sentence compression. In the proposed *AlignFuse* algorithm node alignment between semantically similar words is performed to facilitate phrase stitching from similar sentences. Sentences of the input text often contain additional clauses that are not relevant enough to be included in the summary. As a consequence, DEPSYM++, a syntactic simplification algorithm based on dependency tree structure is proposed to split long sentences into simpler and smaller sentences before applying abstract generation strategies. Input text representation using a multilayer weighted sentence network along with VoteRank sentence node scoring has been used in *VoteSumm* summarization system. Its efficacy has been examined with both raw and simplified inputs. A comparative study with different sentence node ranking schemes is also conducted. In order to perform query-focused text summarization, *QSummFuzzy* has been developed in this research. It uses a combination of FuzzyRank, derived from a Mamdani fuzzy inference system, network based TextRank and keyphrase information for sentence ranking. Coreference resolution has been used for sentence rewriting.

बहु-वाक्य संक्षिप्तीकरण और वाक्य सरलीकरण से बहु-दस्तावेज़ अब्स्ट्रक्टिव पाठ सारांशीकरण

शोध प्रबंध सार

हर मिनट इंटरनेट पर भारी मात्रा में पाठ्य सूचना उत्पन्न होने के कारण स्वचालित टेक्स्ट सारांश प्रणाली की अत्यधिक मांग है। ऐतिहासिक रूप से एक्स्ट्रेक्टिव टेक्स्ट सारांशीकरण में प्रयुक्त वाक्य रैंकिंग दृष्टिकोण ने एनएलपी (NLP) शोधकर्ताओं का बहुत ध्यान आकर्षित किया है। हालांकि, एक्सट्रैक्शन से जुड़ी विभिन्न कमियों ने हाल ही में अब्स्ट्रक्टिव सारांशीकरण के प्रति शोधकर्ताओं में अति उत्साह पैदा किया है। लेकिन यह कार्य काफी चुनौतीपूर्ण है क्योंकि इसमें इनपुट टेक्स्ट की गहरी समझ और इनपुट की सूचना और शब्दार्थ को संरक्षित करने के लिए गहन प्राकृतिक भाषा निर्माण (NLG) तकनीकों की आवश्यकता होती है। वर्तमान शोध का ध्यान दो वाक्य पुनर्लेखन तकनीकों अर्थात् बहु-वाक्य संक्षिप्तीकरण और वाक्य सरलीकरण का उपयोग करके बहु-दस्तावेज़ अब्स्ट्रक्टिव पाठ सारांशीकरण करने पर है।

हमारा शोध अब्स्ट्रक्टिव टेक्स्ट सारांशीकरण से संबंधित कई प्रश्नों के उत्तर खोजने का प्रयास करता है:

अ) सारांश संरचना को सुगम करने के लिए वाक्यों का प्रभावी प्रतिनिधित्व।

ब) संक्षेपण तकनीकों को नियोजित करने से पहले पूर्व-प्रसंस्करण चरण के रूप में वाक्यात्मक सरलीकरण की उपयोगिता की जांच करना।

स) नियतात्मक तरीके से और फ़ज़ी (Fuzzy) तरीके से एक उपयुक्त सरलीकृत वाक्य चयन प्रक्रिया का विकास करना।

विशेष रूप से, बाद वाले का उपयोग क्वेरी-केंद्रित सारांशीकरण की स्थिति में किया गया है।

बहु-वाक्य संक्षिप्तीकरण उत्पन्न करने के लिए वाक्य समूहों के वर्ड ग्राफ़ (Word Graph) -आधारित प्रतिनिधित्व का उपयोग किया गया है। प्रस्तावित AlignFuse एल्गोरिथ्म में समान वाक्यों से बने व्याख्यांशों के मेल के लिए समान शब्दार्थ वाले शब्दों के बीच नोड एकीकरण किया गया है। इनपुट टेक्स्ट के वाक्यों में अक्सर अतिरिक्त खंड होते हैं जो सारांश में शामिल किए जाने के लिए पर्याप्त प्रासंगिक नहीं होते हैं। एक परिणाम के रूप में निर्भरता वृक्ष पद विच्छेदन पर आधारित, DEPSYM++, एक वाक्यात्मक सरलीकरण एल्गोरिथ्म प्रस्तावित किया गया है। इसका इस्तेमाल सारांशीकरण की रणनीतियों को लागू करने से पहले लंबे वाक्यों को सरल और छोटे वाक्यों में विभाजित करने में किया गया है। वोटसम (VoteSumm) सारांश प्रणाली में इनपुट टेक्स्ट प्रतिनिधित्व के लिए एक बहुपरत भारित वाक्य नेटवर्क का उपयोग करते हुए वोटरैंक (VoteRank) वाक्य नोड स्कोरिंग का प्रयोग किया गया है। अनिर्मित और सरल इनपुट दोनों के साथ इसकी प्रभावकारिता की जांच की गई है। विभिन्न वाक्य नोड रैंकिंग योजनाओं के साथ एक तुलनात्मक अध्ययन भी किया जाता है। क्वेरी-केंद्रित टेक्स्ट सारांशीकरण करने के लिए इस शोध में QSummFuzzy को विकसित किया गया है। यह वाक्य रैंकिंग के लिए ममदानी फ़ज़ी इनफरेंस (Mamdani Fuzzy Inference) सिस्टम से प्राप्त फ़ज़ीरैंक, नेटवर्क-आधारित टेक्स्टरैंक, की-फ्रेज़ जानकारी के संयोजन का उपयोग करता है। इसमें वाक्य पुनर्लेखन के लिए कोरफेरेंस रिज़ॉल्यूशन का उपयोग किया गया है।

TABLE OF CONTENTS

	Page
Certificate	i
Acknowledgments	ii
Abstract	iv
List of Figures	ix
List of Tables	xi
List of Abbreviations	xiii
List of Symbols	xvi
CHAPTER I – INTRODUCTION	1
1.1 Introduction to Text summarization	1
1.2 Genres of Summarization	1
1.2.1 Extractive vs. Abstractive	1
1.2.2 Single document vs. Multi-document	2
1.2.3 Generic vs. Query-focused	2
1.3 Motivation	2
1.4 Objectives of the Thesis	4
1.5 Organisation of the Thesis	5
CHAPTER II – LITERATURE REVIEW	10
2.1 Past Works on Extractive Summarization	10
2.2 Past Works on Abstractive Summarization	15
2.2.1 Fully Abstractive Summarization	15
2.2.2 Sentence Fusion	19
2.2.3 Sentence Simplification	22
2.2.3.1 Sentence Compression	22
2.2.3.2 Syntactic Simplification	24
2.3 Summary Evaluation	26
CHAPTER III – MULTI-SENTENCE COMPRESSION USING WORD GRAPH AND NODE ALIGNMENT FOR MULTI-DOCUMENT AB- STRACTIVE SUMMARIZATION	31
3.1 Introduction	31
3.2 Sentence Clustering	33
3.3 Proposed MSC Generation Through Word Graph	35
3.3.1 Node Alignment	36
3.3.1.1 Broad Syntactical Categories	40
3.3.1.2 Removal of Non-transitive Alignments	41

3.3.1.3	Filtering based on Distributional Similarity of Context	42
3.3.2	FaceWord Selection for Aligned Nodes	43
3.3.3	Edge Weights	46
3.3.4	Top-k Shortest Path	46
3.4	Sentence Scoring	47
3.4.1	Grammatical Quality	47
3.4.2	Information Content Score	47
3.4.3	Intensifier Function	48
3.4.4	Computation of Final Score	49
3.5	Sentence Selection for Summary Generation	49
3.6	Experiments for Sentence Fusion	51
3.6.1	Sentence Fusion Evaluation	51
3.6.2	Results for Sentence Fusion	53
3.7	Experiments for Multi-Document Summarization	54
3.7.1	Results and Discussion	56
3.8	Concluding Remarks	59
CHAPTER IV – SYNTACTIC SIMPLIFICATION OF TEXT USING DEPENDENCY TREES AND ITS APPLICATION TO TEXT SUM- MARIZATION		61
4.1	Introduction	61
4.2	Proposed Syntactic Simplification Approach:	
	DEPSYM++	62
4.2.1	Appositive Clauses	62
4.2.2	Relative Clauses	64
4.2.3	Conjoint Clauses	68
4.2.4	Passive Voice	70
4.3	Evaluation of DEPSYM++	73
4.3.1	Test Dataset	73
4.3.2	Baseline Simplification Systems	74
4.3.3	Evaluation Methods	75
	4.3.3.1 Automatic Evaluation	75
	4.3.3.2 Human Evaluation	75
4.3.4	Results and Analysis	76
4.4	Studying the Effect of DEPSYM++ on Single Document Summarization	80
4.4.1	Text Summarization Techniques	80
4.4.2	Experimental Details	82
4.4.3	Results and Analysis	83
4.5	Concluding Remarks	85

CHAPTER V – A MULTILAYER SENTENCE NETWORK BASED APPROACH FOR MULTI-DOCUMENT TEXT SUMMARIZATION	88
5.1 Introduction	88
5.2 VoteSumm: Proposed Multilayer Sentence Network Approach for Summarization	89
5.2.1 Sentence Similarity	89
5.2.2 Network Creation	90
5.2.2.1 Multilayer Network	90
5.2.3 VoteRank Node Ranking	91
5.2.4 Summary Generation	92
5.3 Experimental Details	93
5.3.1 Baseline Network Measures	93
5.3.1.1 Weighted Degree	93
5.3.1.2 PageRank	93
5.3.1.3 Closeness Centrality	94
5.3.1.4 Betweenness Centrality	94
5.3.2 Summary Evaluation	95
5.4 Results and Discussion	95
5.4.1 Unilayer Network	95
5.4.1.1 Unilayer with TopRank Approach	95
5.4.1.2 Unilayer with OnePerCluster Approach	98
5.4.2 Multilayer Network	99
5.4.2.1 Multilayer with TopRank Approach	99
5.4.2.2 Multilayer with OnePerCluster Approach	100
5.4.3 Comparison with Existing Works	102
5.5 Comparison of VoteSumm+ with AlignFuse	103
5.6 Concluding Remarks	105
CHAPTER VI – QUERY-FOCUSED MULTI-DOCUMENT TEXT SUMMARIZATION USING FUZZY INFERENCE	106
6.1 Introduction	106
6.2 QSummFuzzy: The Proposed Approach	108
6.2.1 Text Similarity	109
6.2.2 Keyphrase Extraction	109
6.2.3 Preliminaries of Fuzzy Sets	109
6.2.4 Mamdani Fuzzy Inference	111
6.2.5 Fuzzy Ranking	113
6.2.5.1 Antecedent Variables	114
6.2.5.2 Fuzzification	117

6.2.5.3 Rule Base	117
6.2.6 TextRank	120
6.2.7 Sentence Clustering	120
6.2.8 Summary Generation	121
6.2.9 Coreference Resolution	122
6.3 Experimental Details	123
6.4 Results and Discussion	123
6.5 Concluding Remarks	127
CHAPTER VII – CONCLUSION AND SUGGESTIONS FOR FUTURE WORK	129
7.1 Summary of the Research Work	130
7.1.1 Word Graph-based Multi-Sentence Compression	130
7.1.2 Syntactic Simplification using Dependency Trees	131
7.1.3 Sentence Network Approach for Text Summarization	132
7.1.4 Fuzzy Inference for Query-focused Text Summarization	133
7.2 Future Research Directions	134
REFERENCES	136
APPENDICES	156
LIST OF PUBLICATIONS	167
ABOUT THE AUTHOR	169

LIST OF FIGURES

Fig. 3.1	AlignFuse summarization model architecture	33
Fig. 3.2	Hierarchical Agglomerative Clustering method	35
Fig. 3.3	Precision, Recall and F1-Score for SICK dataset	36
Fig. 3.4	Proposed Word Graph based sentence representation	37
Fig. 3.5	Dependency Representation	38
Fig. 3.6	Aligned words for the sentences picked from DUC-2004 dataset (d30036t)	
		39
Fig. 3.7	Alignment of words with dissimilar POS tag	40
Fig. 3.8	Non-transitive alignments	41
Fig. 3.9	Word Graph with aligned nodes and a possible multi-sentence com-	
	pression path	43
Fig. 3.10	Intensifier function	49
Fig. 3.11	ROUGE-2 scores for different values of τ_{sim} with $Score_{sum}$	57
Fig. 3.12	ROUGE-4 scores for different values of τ_{sim} with $Score_{sum}$	57
Fig. 3.13	ROUGE-2 scores with $score_{sum}$ when M is varied	57
Fig. 3.14	ROUGE-4 scores with $Score_{sum}$ when M is varied	58
Fig. 4.1	Simplification of sentence containing appositive clause	63
Fig. 4.2	Simplification of relative clause with relative pronoun <i>where</i>	65
Fig. 4.3	Simplification of relative clause with relative pronoun <i>whose</i>	66
Fig. 4.4	Simplification of sentence containing relative clause where relative	
	token has one left child	66
Fig. 4.5	Simplification of sentence containing relative clause where relative	
	token has more than one left child	67
Fig. 4.6	Simplification of sentences containing conjoint clause with conjunc-	
	tion because	69
Fig. 4.7	Simplification of sentences containing conjoint clause with conjunc-	
	tion <i>after</i>	70
Fig. 4.8	Passive to active conversion	71
Fig. 4.9	Rules for sentence tense identification	71
Fig. 4.10	Application of DEPSYM++ as preprocessing for Text Summariza-	
	tion	80
Fig. 4.11	Percentage change in ROUGE Scores after application of DEPSYM++	
	Simplification for extractive summarization schemes	84
Fig. 4.12	Percentage change in ROUGE Scores after application of DEPSYM++	
	Simplification for abstractive summarizer BART	84
Fig. 5.1	Architecture of the proposed VoteSumm method	89

Fig. 5.2	ROUGE values for summary generated using <i>TopRank</i> approach for different values of K with raw (unsimplified) inputs in unilayer network	96
Fig. 5.3	ROUGE values for summary generated using <i>TopRank</i> approach for different values of K with simplified inputs in unilayer network	96
Fig. 5.4	ROUGE values for summary generated using <i>TopRank</i> approach for different values of K in unilayer network	97
Fig. 5.5	ROUGE values for summary generated using <i>OnePerCluster</i> approach for different values of λ with raw inputs in unilayer network	98
Fig. 5.6	ROUGE values for summary generated using <i>OnePerCluster</i> approach for different values of λ with simplified inputs in unilayer network	98
Fig. 5.7	ROUGE values for summary generated using <i>TopRank</i> approach for different values of α with raw inputs in multilayer network	99
Fig. 5.8	ROUGE values for summary generated using <i>TopRank</i> approach for different values of α with simplified inputs in multilayer network	100
Fig. 5.9	ROUGE values for summary generated using <i>OnePerCluster</i> approach for different values of α and $K = 0.6$ with raw inputs in multilayer network	100
Fig. 5.10	ROUGE values for summary generated using <i>OnePerCluster</i> approach for different values of α and $K = 0.6$ with simplified inputs in multilayer network	101
Fig. 6.1	QSummFuzzy: The Proposed Approach	108
Fig. 6.2	Implication in Mamdani fuzzy inference system	112
Fig. 6.3	Rule aggregation in Mamdani fuzzy inference system	113
Fig. 6.4	Center of Maxima defuzzification	114
Fig. 6.5	Fuzzy Ranking using Mamdani fuzzy inference system	115
Fig. 6.6	Fuzzification of Variables	119

LIST OF TABLES

Table. 3.1	Examples of sentence pairs from SICK Dataset	34
Table. 3.2	Spearman correlation coefficient	34
Table. 3.3	Values of $-\log(p(a b))$ as reported in PPDB	44
Table. 3.4	Comparison with baselines of the proposed system ($\tau_{\text{sim}} = 0$) across different evaluation metrics (Average Scores are reported)	54
Table. 3.5	Mean scores for sentence fusion generated by the proposed model	54
Table. 3.6	Sentence Fusion generated by the proposed system	55
Table. 3.7	Data Statistics for DUC-2004 dataset	56
Table. 3.8	Comparison of the proposed scheme ($M = 12$) with various extractive summarization methods.	58
Table. 3.9	Comparison of the proposed system ($M = 12$, $Score_{\text{sum}}$) with Word Graph based systems	59
Table. 3.10	ROUGE scores when sentence scores are calculated without applying intensifier function	59
Table. 3.11	Summaries generated using AlignFuse ($\tau_{\text{sim}} = 0.8$, $M = 12$, $Score_{\text{sum}}$) for document set d30037t of DUC-2004 dataset	60
Table. 4.1	Examples	62
Table. 4.2	Pronoun form transform	72
Table. 4.3	Questions for human evaluation of simplification outputs	76
Table. 4.4	Results for PWKP/WikiSmall test set	77
Table. 4.5	Results of automatic evaluations for the TurkCorpus test set	77
Table. 4.6	Results of human evaluations for the TurkCorpus test set	77
Table. 4.7	Outputs of different simplification systems	79
Table. 4.8	Data Statistics for DUC-2001, DUC-2002 and CNN/DM datasets	83
Table. 4.9	ROUGE scores for different summarization methods	84
Table. 4.10	Summary for document la082889 -0067 of DUC2002	86
Table. 5.1	Best performance for different node scoring techniques	101
Table. 5.2	Comparison with existing techniques	103
Table. 5.3	Example of summaries generated using VoteSumm and VoteSumm+ for document set d30038t	103
Table. 5.4	Comparison with AlignFuse	104
Table. 5.5	Example of summaries generated using VoteSumm++ and AlignFuse ($\tau_{\text{sim}} = 0.8$) for document set d30038t	104
Table. 6.1	Examples of Queries from DUC-2006 dataset	115
Table. 6.2	Fuzzy Variable Modelling	118
Table. 6.3	Examples of Fuzzy Rules	120
Table. 6.4	Data Statistics for DUC 2006	123

Table. 6.5	Experimental Results	124
Table. 6.6	Analysis of individual components for generation of Extractive Sum-	
	mary	125
Table. 6.7	Analysis of individual components for generation of Abstractive Sum-	
	mary	125
Table. 6.8	Ablation results for analyzing the importance of individual antecedent	
	variables for generation of Extractive Summary	125
Table. 6.9	Ablation results for analyzing the importance of individual antecedent	
	variables for generation of Abstractive Summary	125
Table. 6.10	Summary generated using QSummFuzzy for document set D0623E	126
Table. A.1	Description of dependency labels in dependency trees	156
Table. B.1	Reference Summaries for document set d30037t of DUC2004	158
Table. B.2	Reference Summaries for document la082889-0067 of DUC2002	160
Table. B.3	Reference Summaries for document set d30038t of DUC2004	161
Table. B.4	Reference Summaries for document set D0623E of DUC2006	163

LIST OF ABBREVIATIONS

ATS	Automatic Text Summarization
AMR	Abstractive Meaning Representation
BART	Bidirectional and Auto-Regressive Transformer
BERT	Bidirectional Encoder Representations from Transformers
BLEU	BiLingual Evaluation Understudy
CNN/DM	CNN/DailyMail
CR	Compression Ratio
DEPSYM++	DEpendency Parse based SYntactic siMplification
DUC	Document Understanding Conferences
EACS	Embedding Average Cosine Similarity
EW	English Wikipedia
FKGL	Flesch-Kincaid Grade Leve
GMS	Greedy Matching Score
HAC	Hierarchical Agglomerative Clustering
ILP	Integer Linear Programming
LSA	Latent Semantic Analysis
LSTM	Long Short Term Memory Networks
METEOR	Metric for Evaluation of Translation with Explicit ORdering
ML	Machine Learning
MSC	Multi-Sentence Compression
NLP	Natural Language Processing
NLE	Natural Language Engineering
POS	Part Of Speech
PPDB	ParaPhrase DataBase
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
RI	Random Indexing
RNN	Recursive Neural Network
RSG	Rich Semantic Graphs
SARI	System output Against References and Input sentence
SCU	Summary Content Unit

SEW	Simple English Wikipedia
SVD	Singular Value Decomposition
SVM	Support Vector Machines
TAC	Text Analysis Conferences
TC	Turk Corpus
TF-IDF	Term Frequency-Inverse Document Frequency
TMF	Triangular Membership Function
WSM	Word Space Model

LIST OF SYMBOLS

d	Damping Factor
r	Sentence Relatedness score in SICK Dataset
λ	Sentence Clustering Threshold for HAC
τ_{sim}	Threshold for cosine similarity between BERT word vectors
T	Total number of paths between START and END node of the Word Graph
k	Number of paths extracted from the Word Graph
M	Minimum summary sentence length
W	Summary length limit
K	Threshold for cosine similarity between sentences
α	Layering constant