

**MULTI-DOCUMENT TEXT SUMMARIZATION USING TEXT
CLUSTERING FOR ARABIC LANGUAGE**

SAMER ABDULATEEF WAHEEB

**SCHOOL OF COMPUTING
UUM COLLEGE OF ARTS AND SCIENCES
UNIVERSITI UTARA MALAYSIA
2014**

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstract (English)

The process of multi-document summarization is producing a single summary of a collection of related documents. In this work we focus on generic extractive Arabic multi-document summarizers. We also describe the cluster approach for multi-document summarization. The problem with multi-document text summarization is redundancy of sentences, and thus, redundancy must be eliminated to ensure coherence, and improve readability. Hence, we set out the main objective as to examine multi-document summarization salient information for text Arabic summarization task with noisy and redundancy information. In this research we used Essex Arabic Summaries Corpus (EASC) as data to test and achieve our main objective and of course its subsequent sub-objectives. We used the token process to split the original text into words, and then removed all the stop words, and then we extract the root of each word, and then represented the text as bag of words by TFIDF without the noisy information. In the second step we applied the K-means algorithm with cosine similarity in our experimental to select the best cluster based on cluster ordering by distance performance. We applied SVM to order the sentences after selected the best cluster, then we selected the highest weight sentences for the final summary to reduce redundancy information. Finally, the final summary results for the ten categories of related documents are evaluated using Recall and Precision with the best Recall achieved is 0.6 and Precision is 0.6.

Keywords: Multi-document text summarization, Arabic text summarization, Automatic text summarization, Text clustering.

Abstrak (Bahasa Malaysia)

Proses ringkasan multi-dokumen adalah menghasilkan ringkasan tunggal daripada beberapa dokumen yang berkaitan. Dalam disertasi ini kami memberi tumpuan kepada ringkasan multi-dokumen generik ekstraktif dalam Bahasa Arab. Kami juga menghuraikan pendekatan kluster bagi ringkasan berbilang dokumen. Permasalahan yang berkaitan dengan ringkasan multi-dokumen ialah lebih ayat yang berulang, dan dengan itu, ianya mesti dikeluarkan daripada ringkasan bagi memastikan kepaduan dan meningkatkan kebolehbacaan ringkasan yang dihasilkan. Oleh itu, objektif utama disertasi ini ialah untuk memeriksa maklumat penting ringkasan pelbagai dokumen untuk teks Bahasa Arab dengan mengambilkira maklumat asing dan lebih ayat. Dalam kajian ini kami menggunakan Essex Arabic Summaries Corpus (EASC) sebagai data bagi menguji dan mencapai matlamat utama kami dan seterusnya mencapai sub-objektif berikutnya. Kami menggunakan proses pertimbangan untuk mengasingkan teks asal ke dalam perkataan, dan kemudian mengeluarkan semua perkataan yang tidak signifikan, dan kemudian kata akar bagi setiap perkataan diekstrak, dan seterusnya mewakili teks dalam bentuk bag perkataan dengan TFIDF tanpa maklumat yang tidak diperlukan. Dalam langkah kedua kami menggunakan algoritma K-means dengan persamaan kosinus dalam percubaan untuk memilih kluster terbaik berdasarkan susunan kluster oleh prestasi jarak. Kami menggunakan SVM untuk menyusun ayat selepas memilih kluster yang terbaik, dan kemudian memilih ayat dengan pemberat paling tinggi bagi ringkasan akhir untuk mengurangkan maklumat lebihan. Akhirnya, keputusan ringkasan akhir bagi sepuluh kategori dokumen berkaitan dinilai menggunakan *Recall* dan *Precision* dengan *Recall* yang terbaik dicapai adalah 0.6 dan *Precision* ialah 0.6.

Kata kunci: Ringkasan teks multi-dokumen, ringkasan teks Bahasa Arab, ringkasan teks automatik, pengklusteran teks.

Acknowledgements

To the memory of my father ...

To my mother ...

To my family and brothers ...

Thanks to Allah for giving me the help and power to accomplish this research. Without the grace of Allah, I was not able to accomplish this work.

Writing this master thesis would not have been possible without the help and continuous with people support me during this wonderful journey.

This thesis would not have seen the light without the endless support and enormous, guidance and patience of my supervisors, Dr. Husniza binti Husni, and A.P. Dr. Faudziah Ahmad for which my mere thanks expression likewise does not suffice.

My time as a master student at Utara University was great with lots of wonderful memories. The time when I started my master I was a bit worried of losing my social life but luckily that was not the case. I was able to split my time between my master, connecting with my friends, playing football, and travelling.

Last but not the least, I would like to offer thanks to all my friends at University Utara Malaysia. They each helped make my time during this master program more fun and interesting.

I would like to thank to all Universiti Utara Malaysia management especially College of Arts and Sciences staff and those who involved directly or indirectly in the dissertation. May Allah bless all of you.

Table of Contents

Permission to Use	i
Abstract (English)	ii
Abstrak (Bahasa Malaysia)	iii
Acknowledgements	iv
List of Figures	viii
List of Tables	ix
Glossary of Terms	x
CHAPTER ONE INTRODUCTION	1
1.1 Introduction	1
1.2 Background of the Study	1
1.3 Problem Statement	5
1.4 Research Questions	7
1.5 Research Objectives	8
1.6 Scope of Study	8
1.7 Significance of Study	8
1.8 Thesis Organization	9
1.9 Summary	10
CHAPTER TWO LITERATURE REVIEW	11
2.1 Introduction	11
2.2 Multi-Document Summarization	11
2.3 Multi-document text summarization approaches	12
2.3.1 Machine -based approach	12
2.3.1.1 Clustering based approach	15

2.3.2 Sentence co-relation based approach	21
2.3.3 Time based approach	24
2.3.4 Graph based approach.....	25
2.3.5 Statistical based approach	27
2.4 The Arabic Natural Language Processing (ANLP)	36
2.4.1 Arabic text Summarization Systems	39
2.5 Multi-document Summarization Evaluation.....	40
2.6 Corpora	43
2.6.1 CCA corpus.....	43
2.6.2 BBC Arabic corpus	43
2.6.3 Aljazeera corpus.....	44
2.6.4 Khaleej-2004 corpus	44
2.6.5 CNN Arabic corpus.....	44
2.6.6 Open Source Arabic Corpus (OSAC)	45
2.6.7 Essex Arabic Summaries Corpus (EASC).....	45
2.7 Summary	46
CHAPTER THREE METHODOLOGY	48
3.1 Introduction.....	48
3.2 Research Methodology	48
Phase I: Data collection	49
Phase II: Text Analyzing	49
Phase III: Cluster Experiment.....	53
Phase IV: Summary Generation.....	56
Figure 3.6: Phase IV steps.	57
Phase V: Evaluation of the final summarization result.....	57

3.4 Summary	58
CHAPTER FOUR RESULT AND DISCUSSION	59
4.1 Introduction.....	59
4.2 Experimental Results	59
4.2.1 Phase I – Data collection.....	59
4.2.2 Phase II – Text Analyzing.....	60
4.2.3 Phase III: Cluster Experiment.....	63
4.2.4 Phase IV: Summary Generation.....	66
4.2.5 Phase V: Evaluation.....	70
4.4 Summary	73
CHAPTER FIVE CONCLUSION AND FUTURE WORK.....	75
5.1 Conclusion	75
5.2 Future Work	77
REFERENCES.....	78
APPENDICES.....	87
Appendix A.....	87
Appendix B	91
Appendix C	94

List of Figures

Figure 1.1: Summary types (Gholamrezazadeh, et al., 2009; Lloret & Palomar, 2012)....	3
Figure 2.1: Techniques of Clustering.....	15
Figure 2.2: Cluster based summarization (Kumar & Salim, 2011).....	16
Figure 2.3: multi document node represents a sentence an example graph (Kumar & Salim, 2011).....	26
Figure 3.1: Methodology phases (Radev, et al., 2002).....	48
Figure 3.2: Phase II steps.....	51
Figure 3.3: K-means steps.....	54
Figure 3.4: The angle of cosine measure.....	54
Figure 3.5: Phase III steps.	55
Figure 3.6: Phase IV steps.....	57
Figure 4.1: Text tokenization.....	60
Figure 4.2: Stop words removal.....	61
Figure 4.3: Stemming words.....	62
Figure 4.4: TFIDF vector of words.....	63
Figure 4.5: Cluster Results.....	64
Figure 4.6: Sentences represented by binary map.....	68
Figure 4.7: Sentences labeled.....	68
Figure 4.8: Sentences Order.....	69
Figure 4.9: Sentences selection process.....	70

List of Tables

Table 2.1: Machine -based approach summary.....	15
Table 2.2: Cluster based approach summary.....	21
Table 2.3: Sentence co-relation based approach summary.....	23
Table 2.4: Time based approach summary.....	25
Table 2.5: Graph based approach summary.....	28
Table 2.6: Statistical based approach summary.....	34
Table 2.7 advantages and dis advantages for each approach.....	35
Table 2.8: Various words have different sub part and same root.....	39
Table 4.1: Statistics of EASC Corpus.....	60
Table 4.2: Order of clusters.....	65
Table 4.3: Recall and Precision results.....	71
Table 4.4: Comparison results based on Recall and Precision results.....	73

Glossary of Terms

Notation	Description
Natural Language Processing (NLP)	The science information branch that deals with natural language information.
Information extraction (IE)	A kind of information retrieval whose goal is to automatically extract structured information from unstructured documents.
Automatic Summarization	The creation of a shortened version of a text by a program of computer.
Extractive Summarization	Using IE for generating a system summary.
Generic-based Summary	A summary that presents an overall sense of a documents' contents.
Query-based Summary	A summary that presents the contents of a document that are related to a user's query.
Cluster	A similar group of objects growing closely together.
Clustering	The task of assigning a set of objects into groups (so called clusters) so that the objects in the same cluster are more similar to each other than to those in other clusters.
Hidden Markov Model	A statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved state.
Machine Learning	a scientific discipline concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data.

Unsupervised learning	A machine learning task of inferring a function from unlabeled data.
Supervised learning	A machine learning task of inferring a function from supervised (labeled) training data.

CHAPTER ONE

INTRODUCTION

1.1 Introduction

This chapter presents an overview of the whole study. The first section describes the background of the study that leads to the implementation of the whole research. This is followed by the statement of problem, research question, research objectives, the scope, and the significance of the study.

1.2 Background of the Study

The availability of electronic documents in Arabic language on the internet is increasing exponentially. So everyone should take advantage of this information revolution. The excellent way to gain access to these documents and get the basic thought is to be able to extract the main idea and take advantage of them. For this reason, automatic text summarization has rapidly grown into main research area as demonstrated by the Document Understanding Conference (DUC), which started in 2001 and the Text Analysis Conference (TAC) (McKeown, 2011).

Automatic Text Summarization (ATS) is a procedure of examining the maximum salient information of related documents and transporting them in less space from the original text. On the other hand, Text Summarization (TS), which goal is to take out a reductive source text transformation to summary text through content condensation by

The contents of
the thesis is for
internal user
only

REFERENCES

- Abbas, M., Smaili, K., & Berkani, D. (2009a). Comparing TR-Classifier and KNN by using Reduced Sizes of Vocabularies. *Culture, 1*, 210.
- Abbas, M., Smaili, K., & Berkani, D. (2009b). *A trigger-based classifier*. Paper presented at the The 2nd Int. Conf. on Arabic Language Resources and Tools (MEDAR 2009).
- Agarwal, N., Reddy, R. S., Gvr, K., & Rosé, C. P. (2011). Towards multi-document summarization of scientific articles: making interesting comparisons with SciSumm. *ACL HLT 2011*, 8.
- Al-Sulaiti, L., & Atwell, E. (2004). *Designing and developing a corpus of contemporary Arabic*. University of Leeds (School of Computing).
- Al-Sulaiti, L., & Atwell, E. (2006). The design of a corpus of Contemporary Arabic. *International Journal of Corpus Linguistics, 11*(2).
- Albared, M., Omar, N., & Ab Aziz, M. J. (2009). *Classifiers combination to arabic morphosyntactic disambiguation*. Paper presented at the Electrical Engineering and Informatics, 2009. ICEEI'09. International Conference on.
- Amini, M., & Usunier, N. (2007). A contextual query expansion approach by term clustering for robust text summarization.
- Azmi, A., & Al-thanyyan, S. (2009). *Ikhtasir—A user selected compression ratio Arabic text summarization system*. Paper presented at the Natural Language Processing and Knowledge Engineering, 2009. NLP-KE 2009. International Conference on.
- Azmi, A. M., & Al-Thanyyan, S. (2012). A text summarizer for Arabic. *Computer Speech & Language, 26*(4), 260-273.

- Berger, A., & Mittal, V. O. (2000). *Query-relevant summarization using FAQs*. Paper presented at the Proceedings of the 38th Annual Meeting on Association for Computational Linguistics.
- Boudabous, M. M., Maaloul, M. H., & Belguith, L. H. (2010). Digital learning for summarizing Arabic documents *Advances in Natural Language Processing* (pp. 79-84): Springer.
- Christensen, J., Mausam, S. S., & Etzioni, O. (2013). *Towards Coherent Multi-Document Summarization*. Paper presented at the Proceedings of Association for Computational Linguistics pages 1163–1173, Atlanta, Georgia.
- Conroy, J. M., O’Leary, D. P., & Schlesinger, J. D. (2006). CLASSY Arabic and English multi-document summarization. *Multi-Lingual Summarization Evaluation, 2006*.
- Das, D., & Martins, A. F. (2007). A survey on automatic text summarization. *Literature Survey for the Language and Statistics, 4*, 192-195.
- Davenport, M. (2012). Introduction to Modern Information Retrieval. *Journal of the Medical Library Association: JMLA, 100*(1), 75.
- Deshpande, A. R., & Lobo, L. (2013). Text Summarization using Clustering Technique. *International Journal of Engineering Trends and Technology (IJETT) - Volume4 Issue8*
- Douzidia, F. S., & Lapalme, G. (2004). Lakhas, an Arabic summarization system. *Proceedings of Document Understanding Conferences 2004*.
- El-Haj, M., Kruschwitz, U., & Fox, C. (2011). *Multi-document Arabic text summarisation*. Paper presented at the Computer Science and Electronic Engineering Conference (CEEC), 2011 3rd.
- El-Haj, M. O., & Hammo, B. H. (2008). *Evaluation of query-based Arabic text summarization system*. Paper presented at the Natural Language Processing and Knowledge Engineering, 2008. NLP-KE’08.

- Erkan, G., & Radev, D. R. (2004). LexRank: Graph-based lexical centrality as salience in text summarization. (*JAIR*), 22(1), 457-479.
- Fan, J., Gao, Y., Luo, H., Keim, D. A., & Li, Z. (2008). *A novel approach to enable semantic and visual image summarization for exploratory image search*. Paper presented at the Proceedings of the 1st ACM international conference on Multimedia information retrieval.
- Fisher, S., & Roark, B. (2007). *Feature expansion for query-focused supervised sentence ranking*. Paper presented at the Document Understanding (DUC 2007) Workshop Papers and Agenda.
- Fizman, M., Demner-Fushman, D., Kilicoglu, H., & Rindfleisch, T. C. (2009). Automatic summarization of MEDLINE citations for evidence-based medical treatment: A topic-oriented evaluation. *Journal of biomedical informatics*, 42(5), 801-813.
- Fukumoto, F., Sakai, A., & Suzuki, Y. (2010). *Eliminating redundancy by spectral relaxation for multi-document summarization*. Paper presented at the Proceedings of the 2010 Workshop on Graph-based Methods for Natural Language Processing.
- Galanis, D., & Malakasiotis, P. (2008). *Aueb at tac 2008*. Paper presented at the Proceedings of the TAC 2008 Workshop.
- Gholamrezazadeh, S., Salehi, M. A., & Gholamzadeh, B. (2009). A comprehensive survey on text summarization systems. 9, 1-6.
- Ghwanmeh, S. H. (2005). Applying Clustering of hierarchical K-means-like Algorithm on Arabic Language. *International Journal of Information Technology*, 3(3).
- Giannakopoulos, G., Karkaletsis, V., Vouros, G., & Stamatopoulos, P. (2008). Summarization system evaluation revisited: N-gram graphs. *ACM Transactions on Speech and Language Processing (TSLP)*, 5(3), 5.
- Gupta, M. V., Chauhan, M. P., Garg, S., Borude, M. A., & Krishnan, S. (2012). An Statistical Tool for Multi-Document Summarization. *International Journal of Scientific and Research Publications* 2(5).

- Gupta, V., & Lehal, G. S. (2010). A survey of text summarization extractive techniques. *Journal of Emerging Technologies in Web Intelligence*, 2(3), 258-268.
- Habash, N. Y. (2010). Introduction to Arabic natural language processing. *Synthesis Lectures on Human Language Technologies*, 3(1), 1-187.
- Haboush, A., & Al-Zoubi, M. (2012). Arabic Text Summerization Model Using Clustering Techniques *World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 2, No. 3, 62 – 67, 2012.*
- Hammo, B., Abu-Salem, H., & Lytinen, S. (2002). *QARAB: A question answering system to support the Arabic language*. Paper presented at the Proceedings of the ACL-02 workshop on Computational approaches to semitic languages.
- Hariharan, S., Ramkumar, T., & Srinivasan, R. (2012). Enhanced Graph Based Approach for Multi Document Summarization. *The International Arab Journal of Information Technology*, 4460-4411.
- He, L., Sanocki, E., Gupta, A., & Grudin, J. (1999). *Auto-summarization of audio-video presentations*. Paper presented at the Proceedings of the seventh ACM international conference on Multimedia (Part 1).
- Ibrahim, A., & Elghazaly, T. (2012). *Arabic text summarization using Rhetorical Structure Theory*. Paper presented at the Informatics and Systems (INFOS), 2012 8th International Conference on.
- Ibrahim, A., Elghazaly, T., & Gheith, M. (2013). A Novel Arabic Text Summarization Model Based on Rhetorical Structure Theory and Vector Space Model.
- Jayashree, R., Murthy, S., & Anami, B. (2012). *Categorized Text Document Summarization in the Kannada Language by sentence ranking*. Paper presented at the Intelligent Systems Design and Applications (ISDA), 2012 12th International Conference on.

- Katragadda, R. (2010). GEMS: generative modeling for evaluation of summaries *Computational Linguistics and Intelligent Text Processing* (pp. 724-735): Springer.
- Kaur, R., & Bhathal, G. S. (2013). A Survey of Clustering Techniques. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(5).
- Kumar, Y. J., & Salim, N. (2011). Automatic multi document summarization approaches. *Journal of Computer Science*, 8(1), 133.
- Larson, M. (2011). Automatic Summarization. *Foundations and Trends® in Information Retrieval*, 5(3), 235-422.
- Lin, C.-Y., & Hovy, E. (2002). *From single to multi-document summarization: A prototype system and its evaluation*. Paper presented at the Proceedings of the 40th Annual Meeting on Association for Computational Linguistics.
- Liu, D.-X., He, Y.-X., Ji, D.-H., & Yang, H. (2006). *A Novel Chinese Multi-Document Summarization Using Clustering Based Sentence Extraction*. Paper presented at the Machine Learning and Cybernetics, 2006 International Conference on.
- Lloret, E., & Palomar, M. (2010). Challenging Issues of Automatic Summarization: Relevance Detection and Quality-based Evaluation. *Informatica (Slovenia)*, 34(1), 29-35.
- Lloret, E., & Palomar, M. (2012). Text summarisation in progress: a literature review. *Artificial Intelligence Review*, 37(1), 1-41.
- Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2), 159-165.
- McKeown, A. N. a. K. (2011). Automatic Summarization. *The Essence of Knowledge*, 5.

- Minaei-Bidgoli, B., Parvin, H., Alinejad-Rokny, H., Alizadeh, H., & Punch, W. F. (2014). Effects of resampling method and adaptation on clustering ensemble efficacy. *Artificial Intelligence Review*, 41(1), 27-48.
- Ouyang, Y., Li, W., Li, S., & Lu, Q. (2011). Applying regression models to query-focused multi-document summarization. *Information processing & management*, 47(2), 227-237.
- Owczarzak, K. (2009). *Depeval (summ): dependency-based evaluation for automatic summaries*. Paper presented at the Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1.
- Radev, D. R., Blair-Goldensohn, S., & Zhang, Z. (2001). Experiments in single and multi-document summarization using MEAD. *Ann Arbor*, 1001, 48109.
- Radev, D. R., Hovy, E., & McKeown, K. (2002). Introduction to the special issue on summarization. *Computational linguistics*, 28(4), 399-408.
- Rai, P., & Singh, S. (2010). A survey of clustering techniques. *International Journal of Computer Applications*, 7(12), 156-162.
- Saad, M. K. (2010). Open Source Arabic Language and Text Mining Tools. *International Conference on Electrical and Computer Systems (EECS'10)*.
- Saad, M. K., & Ashour, W. (2010). Arabic Morphological Tools for Text Mining. *Corpora*, 18, 19.
- Said, D., Wanas, N. M., Darwish, N. M., & Hegazy, N. (2009). *A study of text preprocessing tools for Arabic text categorization*. Paper presented at the The Second International Conference on Arabic Language.
- Schilder, F., Kondadadi, R., Leidner, J. L., & Conrad, J. G. (2008). *Thomson reuters at tac 2008: Aggressive filtering with fastsum for update and opinion summarization*. Paper presented at the Proceedings of the first Text Analysis Conference, TAC-2008.

- Schlesinger, J. D., O’leary, D. P., & Conroy, J. M. (2008). Arabic/English multi-document summarization with CLASSY—the past and the future *Computational Linguistics and Intelligent Text Processing* (pp. 568-581): Springer.
- Sobh, I., Darwish, N., & Fayek, M. (2009). *Evaluation Approaches for an Arabic Extractive Generic Text Summarization System*. Paper presented at the proceeding of 2nd International Conference on Arabic Language Resource and Tools.
- Suanmali, L., & Salim, N. (2009). Literature Reviews for Multi-Document Summarization.
- Sun, J.-T., Shen, D., Zeng, H.-J., Yang, Q., Lu, Y., & Chen, Z. (2005). *Web-page summarization using clickthrough data*. Paper presented at the Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval.
- Tiedan Zhu, X. Z. (2012). An Improved Approach to Sentence Ordering For Multi-document Summarization. *IACSIT Press, Singapore, 25*.
- Vishal Gupta , G. S. L. (2010). A survey of text summarization extractive techniques. *Journal of Emerging Technologies in Web Intelligence, 2(3), 258-268*.
- Wadhvani, R., Pateriya, R., & Roy, D. (2013). A Topic-driven Summarization using K-mean Clustering and Tf-Isf Sentence Ranking. *International Journal of Computer Applications, 79*.
- Wan, X. (2007). *TimedTextRank: adding the temporal dimension to multi-document summarization*. Paper presented at the Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval.
- Wan, X. (2008). *An exploration of document impact on graph-based multi-document summarization*. Paper presented at the Proceedings of the Conference on Empirical Methods in Natural Language Processing.

- Wan, X., & Yang, J. (2008). *Multi-document summarization using cluster-based link analysis*. Paper presented at the Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval.
- Wang, L., Raghavan, H., Castelli, V., Florian, R., & Cardie, C. (2013). *A Sentence Compression Based Framework to Query-Focused Multi-Document Summarization*. Paper presented at the Proceedings of ACL.
- Yang, Z., Lin, Y., Wu, J., Tang, N., Lin, H., & Li, Y. (2011). Ranking support vector machine for multiple kernels output combination in protein–protein interaction extraction from biomedical literature. *Proteomics*, *11*(19), 3811-3817.
- Zechner, K., & Waibel, A. (2000). *DIASUMM: Flexible summarization of spontaneous dialogues in unrestricted domains*. Paper presented at the Proceedings of the 18th conference on Computational linguistics-Volume 2.