HYBRID DIFFERENTIAL EVOLUTION BASED AUTOMATIC SINGLE DOCUMENT TEXT SUMMARIZATION

ALBARAA ABUOBIEDA MOHAMMED ALI ABUOBIEDA

UNIVERSITI TEKNOLOGI MALAYSIA

HYBRID DIFFERENTIAL EVOLUTION BASED AUTOMATIC SINGLE DOCUMENT TEXT SUMMARIZATION

ALBARAA ABUOBIEDA MOHAMMED ALI ABUOBIEDA

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Computer Science)

Faculty of Computing
Universiti Teknologi Malaysia

SEPTEMBER 2013

To my beloved parents, brothers and sisters

ACKNOWLEDGEMENT

First and foremost, all praises be to the Almighty Allah (SWT), the Cherisher and the Sustainer of the world, praise be to Him who taught by the pen, taught man that which he does not know. A good number of people have been used by the Almighty God to achieve this giant stride. All these people deserved being commended for their efforts.

First on the list is my highly respected supervisor, Professor Dr. Naomie Binti Salim. She used her wealth of experience to give a very good direction in the work. Her regular motivation and encouragement coupled with her meticulous attention to details made this thesis a reality.

I am indebted a great deal to UTM and Faculty of Computing for providing a good office space and much needed resources for this research. In the same vain, I would like to thank my University, International University of Africa for providing the financial support for overseas training.

I would also like to thank my both parents, my brothers and sisters for their love, support and supplications. I so much appreciate your being there for me during trial periods.

Finally, I would like to thank all my friends and colleagues for their support and assistance.

ABSTRACT

Automatic single document text summarization is a process of condensing an input text document. In this process, a summary extraction approach summarizes a document by extracting the most informative sentences in a document. To select such sentences, a sentence scoring approach is used to assign a score for each input sentence before ranking them accordingly. Based on user defined summary ratio, only top ranked sentences are selected to be part of the summary and selecting the most informative sentences is a challenge for extractive based automatic text summarization researchers. Thus, this research proposed extraction based automatic single document text summarization methods by investigating a single meta-heuristic evolutionary algorithm called Differential Evolution (DE) to generate high quality summaries. The DE algorithm is used (i) to find out the best feature weight score to discriminate between important and non-important features, (ii) to perform as a cluster machine learning method using Normalized Google Distance and Jaccard similarity measures to generate a highly diversed summary, (iii) to employ opposition-based learning (OBL) approach to improve the performance of the DE algorithm and (iv) to develop a hybrid model used to investigate the adavantages of the combination of feature weighting, diversity and OBL approaches. To evaluate the proposed methods, the standard dataset from Document Understanding Conference (DUC) 2002 and the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) as the standard evaluation measurement toolkit were used. Experimental results showed that the hybrid models as well as all the proposed individual methods performed well for text summarization as compared to four benchmark methods: Microsoft Word, Copernic, the best DUC 2002, the worst DUC 2002 summarizers and a human against another human summarizer. In addition, the proposed methods in the DE algorithm outperformed Genetic Algorithm and fuzzy swarm diversity based methods evolutionary based algorithms. The results of the experiments have proven that the proposed hybrid models generate better quality text-summaries.

ABSTRAK

Peringkasan teks dokumen tunggal secara automatik merupakan proses mengkondensasikan teks dokumen input. Dalam proses ini pendekatan pengekstrakan ringkasan berfungsi meringkaskan dokumen dengan mengekstrak ayat-ayat yang penting dalam dokumen. Untuk memilih ayat-ayat penting satu pendekatan penskoran ayat digunakan untuk menetapkan skor bagi setiap ayat sebelum memberikan susunan kedudukan ayat-ayat tersebut. Berdasarkan nisbah ringkasan yang ditetapkan oleh pengguna hanya ayat-ayat yang berada pada susunan kedudukan tertinggi akan dipilih menjadi sebahagian daripada ringkasan. Pemilihan ayat-ayat penting ini merupakan satu cabaran kepada penyelidik bidang peringkasan teks secara ekstraktif. Untuk itu kajian ini mencadangkan peringkasan teks dokumen tunggal secara ekstraktif dengan mengkaji algoritma evolusi meta-heuristik yang dikenali sebagai Pembezaan Evolusi (DE) bagi menghasilkan ringkasan yang berkualiti tinggi. Algoritma DE digunakan untuk (i) mengetahui skor terbaik setiap pemberat ciri bagi membezakan ciri-ciri penting dan yang tidak penting, (ii) melaksanakan kaedah pembelajaran mesin secara gugusan menggunakan Jarak Google Ternormal dan ukuran kesamaan Jaccard untuk menjana pelbagai ringkasan, (iii) menggunakan pembelajaran berasaskan tentangan (OBL) untuk meningkatkan prestasi algoritma DE, dan (iv) membangunkan model hibrid untuk mengkaji kebaikan gabungan pemberat ciri, kepelbagaian dan pendekatan OBL. Untuk menilai kaedah-kaedah yang dicadangkan set data daripada Persidangan Pemahaman Dokumen (DUC) 2002 dan alat pengukuran piawai yang dikenali sebagai Recall-Oriented Understudy for Gisting Evaluation (ROUGE) digunakan. Hasil kajian menunjukkan bahawa model hibrid dan semua kaedah individu yang dicadangkan mempunyai prestasi lebih baik berbanding dengan empat kaedah tanda aras piawai, iaitu Microsoft Word, Copernic, kaedah-kaedah terbaik dan paling lemah dalam pertandingan DUC 2002 dan bandingan hasil ringkasan manusia sesama manusia. Selain itu penggunaan kaedah algoritma DE mengatasi kaedah-kaedah algoritma evolusi yang lain seperti algoritma genetik dan kaedah kerumunan kepelbagaian kabur. Keputusan eksperimen telah membuktikan bahawa model hibrid yang dicadangkan menghasilkan ringkasan teks yang lebih berkualiti.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Recently, several natural-language processing applications have been designed using intelligent and soft computing techniques to enable the computer systems to mimic the human text processing practices such as plagiarism detection, pattern recognition and machine translation. Intelligence techniques such as genetic algorithms, swarm intelligence, evolutionary algorithms, fuzzy logic and neural networks are often employed. Some of the main reasons behind enabling such mimic are because that computer systems are more precise and perform faster compared to human performance. Automatic text summarization is one of these natural language applications that use such techniques to optimizes it's performance.

Text summarization is a process of summarizing texts into condensed forms Saggion and Poibeau (2013). If the summary is generated by a human, it is called "manual text summarization", whereas if a summary is generated using the computer system it is called "automatic text summarization" (ATS). As this research concerns automatic text summarization, the rest of this section discusses automatic text summarization (ATS) approaches, styles, input size, and evaluation techniques.

The research on ATS can be divided into two approaches: extraction-based summary and abstraction-based summary. The extraction-based approach generates a summary by selecting (copy-paste) the important sentences. These sentences are evaluated based a on scoring mechanism called "features" where each sentence is assigned a score. The top scored sentences are selected as summary candidates sentences. The abstraction-based approach composes summaries by editing the most important text units (sentences or phrases) such as: removing, appending, segmenting and paraphrasing some parts of those text units. The abstraction-based approach is more complex compared to extraction-based approach (Armano *et al.*, 2011).

The target summary can be employed and written in one of the following two styles (Gholamrezazadeh *et al.*, 2009): "indicative summary" or "informative summary". Indicative summary presents brief information of what is contained in the original document focusing on a certain topic. The generated summary of this kind is usually compressed between five to ten percent of the original text. An informative summary covers the most topics that arise within the original document. The generated summary of this kind is usually compressed between twenty to thirty percent of original text content. In addition, ATS researches cover two types of document input size i.e. single-document, and multi-document summarization.

Based on the processing level, the summarization techniques can be classified into three approaches: the surface, entity, and discourse (Mani and Maybury, 1999, Saggion and Poibeau, 2013). The surface level approach uses a shallow feature set to extract the most relevant sentences in a document to be included in the summary. The methods in entity level approach first extract entities and their relationships from the text, then model the extraction. In order to identify the salient entity-to-entity relationship from the text, there are several approaches that can be used such as a graph-based representation and a vector space model. The Discourse-Level approach concerns modelling the global structure of the text and its relationships such as: the rhetorical structure of the text (e.g., narrative and argumentation structure), document format (e.g., document outlines, hypertext mark-up) and topics threads (as they are exposed in the text).

A recent survey written by Saggion and Poibeau (2013) states that summarization evaluation still represents a big challenge in computer natural-language processing. There are several difficulties being faced by the automatic summarization researchers such as the deep understanding of linguistic issues, language modelling and computer-based problem solving techniques. In addition, comparing manually generated human summaries with automatically generated summaries also poses hard issues for the purpose of evaluation. However, there are two main classes of evaluation methods used in automatic text summarization: intrinsic and extrinsic (Jing *et al.*, 1998, Mani and Maybury, 1999, Afantenos *et al.*, 2005). Extrinsic evaluation is a task-oriented based facility that measures how the summaries are used for a given task. Whereas the intrinsic evaluation method compares generated system summaries to reference summaries i.e. human generated summaries. There are many evaluation tools proposed such as the Recall-Oriented Understanding for Gisting Evaluation

(ROUGE) (Lin, 2004), and PYRAMID (Nenkova *et al.*, 2007). Both methods represent intrinsic automatic evaluations tools, and ROUGE is found to highly correlate with the results of human judgments (Lin, 2004).

The methods proposed in this research are for single document extraction based text summarization that produce informative summaries using techniques of surface-level processing approach.

1.2 Problem Background

Most research in the area of information retrieval (IR) aim to relieve high information load (e.g., Internet, documents) that users potentially face by proposing methods that are precisely targeted for retrieved results. Searching for the information of interest in a wide scope of knowledge is a very difficult task, and if the retrieval systems are imprecise (designed with poor quality) they may return missed-information or zero-sum results. In addition, exploring many documents one by one is time consuming. Accordingly, ATS researchers aim to mitigate or solve this problem by proposing methods that produce high quality summaries. The goal of the summary as a part of the IR system acts as a rapid guide to information of interest through presenting a condensed form of each document within the field of search.

The initially proposed methods for text summarization research are surface level (feature-scoring) approaches (Luhn, 1958, Baxendale, 1958, Edmundson, 1969). (Luhn, 1958) proposed a term-frequency approach to indicate term-importance within the context. (Baxendale, 1958) proposed a sentence position approach to enable the summarizer to identify the sentence importance within the document. Ten years later, (Edmundson, 1969) included the above two approaches and proposed a feature of pragmatic words (cue words such as “significant”, “key”, “idea” and so on). Since feature scoring approach presented significant results, researchers worked on proposing additional features to enhance the summarization quality.

The literature demonstrates that the text features approach plays an observable role in generating qualified summaries (Ferreira *et al.*, 2013, Haque *et al.*, 2013). Therefore, other researchers tried to enclose feature weighting to adjust feature scores in summarization problems (Fattah and Ren, 2009, Binwahlan *et al.*, 2009a, Suanmali *et al.*, 2011b). Empirically, the feature selection methods lead to high quality solution generation. Similarly, the quality of the text summary is sensitive to these features as to how they are scored and weighted. Therefore, the need for a mechanism to differentiate between high and low importance features has emerged. To this end, many feature selection mechanism were proposed, but there is further need to design and build strong mechanisms in order to obtain higher qualified results. The Differential Evolution (DE) algorithm is an evolutionary algorithm that is able to carry out such a role and acts as a feature weighting machine learner. The DE has not been previously proposed as feature-weighting mechanism in text summarization problem; however, it has been employed in related fields such as document clustering, image classification and web data extraction (Abraham *et al.*, 2006, Omran *et al.*, 2005c). The following are the reasons of why the DE was chosen to solve the problem of automatic text summarization. The DE is a powerful algorithm for real parameter optimization (Storn and Price, 1997). A recent work published by (Das *et al.*, 2009) reported that the DE algorithm has become quite popular in the machine intelligence and cybernetics communities. It has successfully been applied to different domains of science and engineering, such as mechanical engineering design (Joshi and Sanderson, 1999), signal processing (DAS and KONAR, 2006) and machine intelligence (Omran *et al.*, 2005a). Section 2.4.1.1 provides in details the characteristics of the DE algorithm which make it strong and robust compare to other heuristic methods.

Another challenge that needs to be addressed concerns capturing most of the document subtopics. This leads to generate a summary that covers most of the themes presented in the text. To solve this problem, the cluster-based (or diversity) approach is used to diversify the sentence selection mechanism whereby selected sentences cover most topics in the document. There are several approaches employed for the diversity-based approach in text summarization (Carbonell and Goldstein, 1998, Filippova *et al.*, 2007, Gong and Liu, 2001b,a, Kraaij *et al.*, 2001, Mori *et al.*, 2005, Steinberger *et al.*, 2005, Binwahlan *et al.*, 2009c). The diversity is used in the summarization to control sentence redundancy in the summary which generates a higher quality summary. The DE algorithm presented previously has been used to optimize the sentence clustering process in order to optimize the diversity within the generated summary text (Alguliev and Aliguliyev, 2009). This research implemented the same method presented by (Alguliev and Aliguliyev, 2009) and discovered two limitations. First, the selection of the similarity measure called “Normalized Google Distance (NGD) (Cilibrasi and Vitanyi, 2007) is improper. Second, sentence centrality is computed independently from other sentences within the document. The NGD is a similarity measure that was successfully implemented to extract a similarity score between two terms in large databases such as Google (Cilibrasi and Vitanyi, 2007) in which billion number of web pages are processed; Table 1.1 shows an example

Table 1.1: Number of retrieved web-pages using Google search engine for certain keywords.

Keyword	Number of retrieved web-pages
“Text”	4,260,000,000
“Summarization”	13,100,000
“Text Summarization”	221,000

of retrieved numbers of web pages with 3 keywords searched at the moment of writing this thesis. Employing the NGD in a small database search space resulted in improper score calculation (Alguliev and Aliguliyev, 2009). Hence, selecting a proper similarity measure plays an important role in adjusting data clustering (Jain *et al.*, 1999). (Alguliev and Aliguliyev, 2009) computed the sentence centrality score after clustering the sentences. However, computing such a score prevents the method from capturing the full relationship between sentences in the document (Shen *et al.*, 2007). Therefore, this study tries to utilize other similarity measures and different sentence scoring mechanisms; then, building a new diversity-based method using the DE algorithm. The method optimizes the clustering process by using an alternative similarity measure, the “Jaccard coefficient” (Jaccard, 1901) , and a “feature-scoring” mechanism for diverse sentence extraction in text summarization.

Naturally, the proposed methodologies are exposed to advantages and disadvantages. Although the optimization techniques are used to overcome some limitations of other proposed methods, they suffer many defects. (Jun *et al.*, 2011) surveyed some evolutionary computing algorithms (ECAs). The survey discussed how the ECAs search performance could be optimized using machine learning techniques. This trend of research direction treated the term “Machine Learning for Evolutionary Computing (MLEC)” for the discussed purpose. The ECAs agreed in a general structure which includes the following stages: population initialization, fitness evaluation and selection, population reproduction and variation, algorithm adaptation, and local search. The survey viewed the algorithm defects and the successful solutions. Most of the techniques used to enhance the search performance of the ECAs are machine learning techniques; they have been initially used to train algorithms before addressing a targeted problem solution. Machine Learning (ML) techniques were used to optimize all stages of the ECAs. In the initial population stage, the machine learning (ML) techniques were used to:

1. Organize the initial solution position.
2. Improving the initial solution quality.

3. Incorporate historical search performance.

The Opposition-based learning (OBL) is a machine learning technique that has been widely used to enhance the DE search performance by adjusting the following: initial population, the next generation of the population, and maintenance of population diversity (Rahnamayan and Tizhoosh, 2008). The OBL was tested on a numerical dataset (Rahnamayan and Wang, 2008) but neither tested on text data nor used for text summarization problems. In addition, the-state-of-the-art review exposed that fact that none of the proposed automatic text summarization studies built based on optimization techniques is included in the concept of the MLEC. So this study investigates the incorporation of OBL (Rahnamayan and Tizhoosh, 2008) to enhance the DE algorithm and test its performance in non-numerical datasets (text data).

This current work considers four important issues in text summarization: feature-weighting mechanism, diversity-based optimization and machine learning for ECAs, and a combination of these issues in a single hybrid model.

1.3 Problem Statement

By understanding the problem background, we found that designing a robust feature weighting mechanism is necessary for generating high quality summary. On the other hand, generating a summary with high diversity could lead to the inclusion of most of the topics existing in the input document. Furthermore, by not relying on the random estimation, it could improve the performance of the evolutionary algorithm which could subsequently enhance the quality of the generated summary. This research concerns sentence extraction to answer the following research questions:

- Can the evolutionary algorithm produce optimal weights for the selected features that produce a high quality summary?
- Can sentence selection based on an optimized cluster-based approach achieve better diversity in the summarization?
- Can an opposition-based learning technique enhance the search performance of an evolutionary algorithm and obtain better qualified results compared to traditional versions for text summarization purposes?

- Can the combination of the previous multiple techniques could exploit the advantages of each method then interact together to produce a summary best than previous methods?

1.4 Research Objectives

The main goal of this research is to introduce text summarization methods designed to solely use a functional approximation (randomized search) approach attached to different learning techniques. Several proposed text summarization research works have been designed using different and hybrid techniques such as functional approximation and approximate reasoning. The functional approximation includes evolutionary algorithms and neural networks, and the approximate reasoning includes probabilistic models and fuzzy logic. Therefore, this research investigates the following hypothesis: "Is a single meta-heuristic based method integrated with learning approaches able to generate higher quality summarization." To achieve this goal, the following objectives have been fixed:

1. To investigate DE-based feature weighting method for text summarization.
2. To improve an existing diversity summary generation method using the term-weighting approach and Jaccard similarity measure. In addition, to design a new real-to-integer data modulator for solving the discrete problem (clustering) for generating a high diverse summaries.
3. To investigate the opposition-based learning technique to optimize the summarization solutions generated from the Differential-Evolution algorithm.
4. To investigate a hybrid approach of differential evolution algorithm with cluster-based approach to select diverse contents from the text for summarization purposes.

1.5 Research Scope

This research was designed using a single meta-heuristic "Differential Evolution" algorithm integrated with learning approaches (feature scoring, cluster based and opposition based learning) in order to examine its ability compared to other summarization applications that are designed using several hybrid techniques. The following aspects are the scope of this research:

1. The methods proposed in this research are for single document extraction based text summarization that produces informative summaries using techniques for surface-level processing. These methods were designed to use a Differential Evolution algorithm with feature-weighting approach, cluster-based approach, the opposition-based learning approach and hybrid-based approach.
2. For the evaluation of data, the DUC 2002 was selected as the test bed of each proposed method. DUC 2002 was chosen because it is the last dataset designed for single document summarization.
3. The Recall-Oriented Understanding for Gisting Evaluation (ROUGE) toolkit was selected to measure and evaluate the system"s generated summaries with reference summaries. In addition, the statistical significance test "Pearson Correlation Coefficient" is used to measure the agreement level between the proposed methods and the human method.
4. The proposed methods in automatic text summarization are evaluated and compared with well-known benchmark methods such as Microsoft Word summarizer and Copernic summarizer. The best and worst systems from the DUC 2002 summarization competition are also compared. In addition to these four methods, similar methods were selected from the literature that may have the same/similar structure and/or methodological functions to some of this study"s proposed methods.

1.6 Research Significance

Since the beginning of research in automatic text summarization by (Luhn, 1958) all proposed methods aim to increase the quality of the summarization results via designing a single technique or by combining models of other techniques. This research desires to make a significant contribution by presenting a novel "Differential Evolution Based Automatic Text Summarization Model" to get higher quality summary. First: the proposed model generates optimal weighting for selected features embedded within the model. Second: to achieve diversity in the summary, this model clusters similar sentences through the same evolutionary algorithm to avoid data redundancy problem. Third: the proposed model enhances the search performance of the evolutionary algorithm (DE) in order to extract optimized results instead of relying on random estimated solutions using Opposition-Based Learning (OBL) concept. Fourth: a hybrid model of all three techniques above are integrated together to utilize their advantages in a single model. The additional significance of this proposed model is that it was built by using a single optimization algorithm (DE); the DE was then appended and supplied with different learning techniques such as feature selection approach, cluster-based approach and opposition-based learning approach.

1.7 Contribution of the Study

The expected contributions of this research are as follows:

1. Extraction of the most important sentences can be obtained by identifying optimal feature weighting using differential evolution algorithm.
2. Achieving diversity in summarization and avoiding the data redundancy problem as obtained through the optimized cluster-based method. Mainly, this structured contribution encloses three sub-contributions as follow:
   (a) NGD + DE-based Term-Weighting: A novel integration of the term-weighting approach with NGD similarity measure.
   (b) DE + Jaccard: A novel integration of the Jaccard similarity measure with the DE algorithm
   (c) Real-to-Integer Modulator: A novel real-to-integer value modulator is designed. This modulator aims to amend real values generated by the DE to integer-base values that enable the DE search in discrete space and fine-tune the cluster based problem.
3. Avoid reliance on random estimated solutions and produce more qualified summaries to exceed those of former approaches (1 and 2).
4. A hybrid model designed to integrate the advantages of all proposed contributions in order to improve the quality of summary generation.

1.8 Thesis Organization

This thesis is organized into eight chapters as follows:

Chapter 1, Introduction: this chapter discusses general issues concerning this research. It also states the problem background, the problem statement, research objectives, research scope, research significance, and expected contribution, respectively.

Chapter 2, this chapter reviews state-of-the-art approaches in the field of automatic text summarization. The chapter reviews recent surveys introduced in the field. Since this research proposes evolutionary algorithm based solutions, the chapter also reviews, in particular, most summarization research based on similar or other evolutionary algorithms. In addition, it reviews machine learning methods that have been presented to enhance the search performance of evolutionary algorithms. The chapter covers available datasets utilized in methodology evaluation as well as the evaluation of tools-kit.

Chapter 3, Research Methodology: this chapter defines the methodology followed in this research to achieve the study"s objectives. The main experiments of this study are: binary differential evolution based text summarization; diversity based differential evolution text summarization; opposition diversity based differential evolution text summarization, and hybrid model based differential evolution text summarization.

Chapter 4, Binary Differential Evolution Based Text Summarization: this chapter presents features the weighting method which uses the evolutionary algorithm Differential Evolution (DE). The DE is configured in binary mode in order to control probable score calculation. The method will be compared against "state of the art" methods and similar systems based on particle swarm optimization and generic algorithms.

Chapter 5, Diversity Based Differential Evolution Text Summarization: this chapter proposes cluster-based methods enhanced with the DE algorithm for generating a highly diverse summary. The methods also aim to avoid falling into a problem of data redundancy. Three methods have been proposed: the first is to improve the Normalized Google Distance (NGD) (Cilibrasi and Vitanyi, 2007) similarity measure performance by incorporating the term-weighting approach. Secondly, to investigate the proper selection of similarity measure that is more suitable to the dataset. Thirdly, is to design a novel "real-to-integer" value modulator instead of adopting an external genetic mutation operator.

Chapter 6, Opposition Diversity Based Differential Evolution Text Summarization: the main goal of this chapter is to avoid the problem of generating solutions based on random estimates. The problem of the application based on random estimates (guesses) is that it may give different solutions each time that are far from the optimal points. This chapter investigates the use of Opposition-based learning (OBL) to solve this issue for automatic text summarization. The OBL is proposed to make sure it is able to enforce DE generating solutions that are closer to optimal points than are traditional versions in text summarization.

Chapter 7, Hybrid model based on DE algorithm: this chapter aims to integrate the advantages of all proposed methods in one single (hybrid) model. The model initially extracts the optimal feature-weights assigned, then, calls on the methods proposed in Chapter Five to explore sentence diversity. Methods in Chapter Five are then automatically improved with a method component presented in Chapter Six. Finally, the optimized feature scores are employed for selecting the top “n” sentences.

Chapter 8, Conclusion and Future Work: this chapter concludes the research and attempts to give an overall discussion regarding all contributions presented in this research as well as recommendations and suggestions for future research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ATS	–	Automatic Text Summarization
DE	–	Differential Evolution
DUC	–	Document Understanding Conference
EAs	–	Evolutionary Algorithms
EC	–	Evolutionary Computing
GA	–	Genetic Algorithm
ML	–	Machine Learning
MLEC	–	Machine Learning for Evolutionary Computing
NGD	–	Normalized Google Distance
OBL	–	Opposition Based Learning
PSO	–	Particle Swarm Optimization
ROUGE	–	Recall-Oriented Understudy for Gisting Evaluation
TF-ISF	–	Term Frequency - Inverse Sentence Frequency

LIST OF SYMBOLS

α	–	Alpha

LIST OF APPENDICES

REFERENCES

Abraham, A., Das, S. and Konar, A. (2006). Document clustering using differential evolution. In *Evolutionary Computation, 2006. CEC 2006. IEEE Congress on*. IEEE, 1784–1791. doi:10.1109/CEC.2006.1688523.

Achananuparp, P., Hu, X. and Shen, X. (2008). The evaluation of sentence similarity measures. *Data Warehousing and Knowledge Discovery*, 305–316.

Afantenos, S., Karkaletsis, V. and Stamatopoulos, P. (2005). Summarization from medical documents: a survey. *arXiv preprint cs/0504061*.

Ahandani, M. A. and Alavi-Rad, H. (2012). Opposition-based learning in the shuffled differential evolution algorithm. *Soft Comput.* 16(8), 1303–1337.

Al-Qunaieer, F. S., Tizhoosh, H. R. and Rahnamayan, S. (2010). Opposition based computing - A survey. In *Neural Networks (IJCNN), The 2010 International Joint Conference on*. 1–7.

Alguliev, R. and Aliguliyev, R. (2007). Experimental investigating the F-meaure as similarity measure for automatic text summarization. *Applied and Computational Mathematics*. 6(2), 278–287.

Alguliev, R. M. and Aliguliyev, R. M. (2009). Evolutionary Algorithm for Extractive Text Summarization. *Intelligent Information Management*. 1(2), 128–138.

Alguliev, R. M., Aliguliyev, R. M. and Isazade, N. R. (2012). DESAMC+DocSum: Differential evolution with self-adaptive mutation and crossover parameters for multi-document summarization. *Knowledge-Based Systems*.

Alguliev, R. M., Aliguliyev, R. M. and Mehdiyev, C. A. (2011a). An Optimization Model and DPSOEDA for Document Summarization. *International Journal of Information Technology and Computer Science(IJITCS)*. 3(5), 59–68.

Alguliev, R. M., Aliguliyev, R. M. and Mehdiyev, C. A. (2011b). pSum-SaDE: a modified p-median problem and self-adaptive differential evolution algorithm for text summarization. *Appl. Comp. Intell. Soft Comput.* 2011, 11–11.

Aretoulaki, M. (1994). Towards a hybrid abstract generation system. In *Proceedings of the International Conference on New Methods in Language Processing*. Citeseer, 220–227.

Armano, G., Giuliani, A. and Vargiu, E. (2011). Studying the impact of text summarization on contextual advertising. In *Database and Expert Systems Applications (DEXA), 2011 22nd International Workshop on*. IEEE, 172–176. Retrievable at `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6059812`.

Bandyopadhyay, S. and Maulik, U. (2002). Genetic clustering for automatic evolution of clusters and application to image classification. *Pattern Recognition*. 35(6), 1197–1208.

Baxendale, P. B. (1958). Machine-made index for technical literature: an experiment. *IBM J. Res. Dev.* 2(4), 354–361.

Binwahlan, M., Salim, N. and Suanmali, L. (2009a). Swarm based features selection for text summarization. *International Journal of Computer Science and Network Security IJCSNS*. 9(1), 175–179.

Binwahlan, M., Salim, N. and Suanmali, L. (2009b). Swarm based text summarization. In *Computer Science and Information Technology-Spring Conference, 2009. IACSITSC'09. International Association of*. IEEE, 145–150.

Binwahlan, M., Salim, N. and Suanmali, L. (2009c). Swarm Diversity Based Text Summarization. In *Neural Information Processing*. Springer, 216–225.

Binwahlan, M. S., Salim, N. and Suanmali, L. (2010). Fuzzy swarm diversity hybrid model for text summarization. *Information processing & management*. 46(5), 571–588.

Black, W., Elkateb, S. and Vossen, P. (2006). Introducing the Arabic WordNet Project. In *In Proceedings of the third International WordNet Conference (GWC-06*.

Boros, E., Kantor, P. and Neu, D. (2001). A clustering based approach to creating multi-document summaries. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Boudabous, M. M., Chaaben Kammoun, N., Khedher, N., Belguith, L. H. and Sadat, F. (2013). Arabic WordNet semantic relations enrichment through morpho-lexical patterns. In *Communications, Signal Processing, and their Applications (ICCSPA), 2013 1st International Conference on*. IEEE, 1–6.

Brest, J., Greiner, S., Boskovic, B., Mernik, M. and Zumer, V. (2006). Self-Adapting Control Parameters in Differential Evolution: A Comparative Study on Numerical Benchmark Problems. *Evolutionary Computation, IEEE Transactions on*. 10(6), 646–657.

Carbonell, J. and Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 335–336.

Chang-Yong, L. and Xin, Y. (2004). Evolutionary programming using mutations based on the Levy probability distribution. *Evolutionary Computation, IEEE Transactions on*. 8(1), 1–13.

Chaves, R. P. (2001). WordNet and Automated Text Summarization. In *NLPRS*. 109–116.

Cilibrasi, R. L. and Vitanyi, P. M. B. (2007). The Google Similarity Distance. *Knowledge and Data Engineering, IEEE Transactions on*. 19(3), 370–383.

Cui, X., Potok, T. E. and Palathingal, P. (2005). Document clustering using particle swarm optimization. In *Swarm Intelligence Symposium, 2005. SIS 2005. Proceedings 2005 IEEE*. 185–191.

Cunningham, H. (2002). GATE, a general architecture for text engineering. *Computers and the Humanities*. 36(2), 223–254.

Da Cunha, I., Fernandez, S., Velzquez Morales, P., Vivaldi, J., SanJuan, E. and Torres-Moreno, J. (2007). A new hybrid summarizer based on vector space model, statistical physics and linguistics. *MICAI 2007: Advances in Artificial Intelligence*, 872–882.

Dancey, C. P. and Reidy, J. (2008). *Statistics without maths for psychology*. Pearson Education.

Dantzig, G. B. (1957). Discrete-variable extremum problems. *Operations Research*. 5(2), 266–288.

Das, S., Abraham, A., Chakraborty, U. and Konar, A. (2009). Differential Evolution Using a Neighborhood-Based Mutation Operator. *Evolutionary Computation, IEEE Transactions on*. 13(3), 526–553. ISSN 1089-778X. doi:10.1109/TEVC.2008.2009457.

Das, S., Abraham, A. and Konar, A. (2008). Automatic Clustering Using an Improved Differential Evolution Algorithm. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*. 38(1), 218–237.

DAS, S. and KONAR, A. (2006). Two-Dimensional IIR Filter Design With Modern Search Heuristics: A Comparative Study. *International Journal of Computational Intelligence and Applications*. 06(03), 329–355. doi:10.1142/S1469026806001848. Retrievable at `http://www.worldscientific.com/`

doi/abs/10.1142/S1469026806001848.

Das, S. and Suganthan, P. (2011). Differential evolution: A survey of the state-of-the-art. *Evolutionary Computation, IEEE Transactions on*. 15(1), 4–31.

de Castro, L. N. (2007). Fundamentals of natural computing: an overview. *Physics of Life Reviews*. 4(1), 1–36.

Deng, C., Zhao, B., Yang, Y. and Deng, A. (2009). Novel binary differential evolution algorithm for discrete optimization. In *Natural Computation, 2009. ICNC'09. Fifth International Conference on*, vol. 4. IEEE, 346–349. Retrievable at http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5366344.

Dorigo, M., Maniezzo, V. and Colorni, A. (1996). Ant system: optimization by a colony of cooperating agents. *Trans. Sys. Man Cyber. Part B*. 26(1), 29–41.

Edmundson, H. P. (1969). New Methods in Automatic Extracting. *J. ACM*. 16(2), 264–285.

Engelbrecht, A. and Pampara, G. (2005). Binary differential evolution strategies. In *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on*. IEEE, 1942–1947.

Fattah, M. A. and Ren, F. (2009). GA, MR, FFNN, PNN and GMM based models for automatic text summarization. *Computer Speech and Language*. 23(1), 126–144.

Fellbaum, C. (2010). WordNet: An electronic lexical database. 1998. *WordNet is available from http://www. cogsci. princeton. edu/wn*.

Ferreira, R., Lins, R. D., Freitas, F., Cavalcanti, G. D., Lima, R., Simske, S. J., Favaro, L. *et al.* (2013). Assessing Sentence Scoring Techniques for Extractive Text Summarization. *Expert Systems with Applications*.

Filippova, K., Mieskes, M., Nastase, V., Ponzetto, S. and Strube, M. (2007). Cascaded filtering for topic-driven multi-document summarization. In *Proceedings of the Document Understanding Conference*. 26–27.

Fuentes, M., González, E., FERRes, D. and RODRiguez, H. (2005). QASUM-TALP at DUC 2005 Automatically Evaluated with a Pyramid based Metric. In *Proceedings of Document Understanding Workshop (DUC). Vancouver, BC, Canada*.

Gholamrezazadeh, S., Salehi, M. A. and Gholamzadeh, B. (2009). A Comprehensive Survey on Text Summarization Systems. In *Computer Science and its Applications, 2009. CSA '09. 2nd International Conference on*. 1–6.

Gong, Y. and Liu, X. (2001a). Creating generic text summaries. In *Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on*. IEEE, 903–907.

Gong, Y. and Liu, X. (2001b). Generic text summarization using relevance measure and latent semantic analysis. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 19–25.

Gupta, P., Pendluri, V. S. and Vats, I. (2011). Summarizing text by ranking text units according to shallow linguistic features. In *Advanced Communication Technology (ICACT), 2011 13th International Conference on*. IEEE, 1620–1625.

Haque, M. M., Pervin, S. and Begum, Z. (2013). Literature Review of Automatic Single Document Text Summarization Using NLP.

Harabagiu, S. M. and Lacatusu, F. (2002). Generating single and multi-document summaries with GISTEXTER. In *Proceedings of the workshop on text summarization DUC 2002*. 30–38.

Harman, D. and Liberman, M. (1993). Tipster complete. *Corpus number LDC93T3A, Linguistic Data Consortium, Philadelphia*.

Hasler, L., Orasan, C. and Mitkov, R. (2003). Building better corpora for summarization. In *Proceedings of Corpus Linguistics*. 309–319.

He, X. and Han, L. (2007). A novel binary differential evolution algorithm based on artificial immune system. In *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on*. IEEE, 2267–2272.

He, X., Zhang, Q., Sun, N. and Dong, Y. (2009). Feature selection with discrete binary differential evolution. In *Artificial Intelligence and Computational Intelligence, 2009. AICI'09. International Conference on*, vol. 4. IEEE, 327–330.

Hou, G. and Ma, X. (2010). A Novel Binary Differential Evolution for Discrete Optimization. *Key Engineering Materials*. 439, 1493–1498.

Hovy, E. and Lin, C.-Y. (1998). *Automated text summarization and the SUMMARIST system*.

Huang, A. (2008). Similarity measures for text document clustering. In *Proceedings of the Sixth New Zealand Computer Science Research Student Conference (NZCSRSC2008), Christchurch, New Zealand*. 49–56.

Hung, C. and Purnawan, H. (2008). A hybrid rough k-means algorithm and particle swarm optimization for image classification. *MICAI 2008: Advances in Artificial Intelligence*, 585–593.

Jaccard, P. (1901). tude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin del la Socit Vaudoise des Sciences Naturelles*. 37, 547–579.

Jain, A., Murty, M. and Flynn, P. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*. 31(3), 264–323.

Jing, R., H. Barzilay, McKeown, K. and Elhadad, M. (1998). Summarization evaluation methods: Experiments and analysis. In *In AAAI Symposium on Intelligent Summarization*. 60–68.

John McCrae, P. B. P. C. T. D. A. G.-P. J. G. L. H. E. M.-P. D. S. T. W., Guadalupe Aguado-de-Cea (2012). Interchanging lexical resources on the Semantic Web. *Language Resources and Evaluation*. 46(4), 701–719. ISSN 1574-020X. doi: 10.1007/s10579-012-9182-3. Retrievable at `http://dx.doi.org/10.1007/s10579-012-9182-3`.

Joshi, R. and Sanderson, A. (1999). Minimal representation multisensor fusion using differential evolution. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*. 29(1), 63–76. ISSN 1083-4427. doi:10.1109/3468.736361.

Julian Kupiec, F. C., Jan Pedersen (1995). *A trainable document summarizer*.

Jun, Z., Zhi-hui, Z., Ying, L., Ni, C., Yue-jiao, G., Jing-hui, Z., Chung, H. S. H., Yun, L. and Yu-hui, S. (2011). Evolutionary Computation Meets Machine Learning: A Survey. *Computational Intelligence Magazine, IEEE*. 6(4), 68–75.

Junhong, L. and Jouni, L. (2002). A fuzzy adaptive differential evolution algorithm. In *TENCON '02. Proceedings. 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering*, vol. 1. 606–611 vol.1.

Kan, M.-Y., Klavans, J. L. and McKeown, K. R. (2002). Using the annotated bibliography as a resource for indicative summarization. *arXiv preprint cs/0206007*.

Karaboga, D. (2005). *An idea based on honey bee swarm for numerical optimization*. Technical report. Erciyes University.

Kazantseva, A. (2006). An approach to summarizing short stories. In *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop*. (pp. 55–62).

Kennedy, J. and Eberhart, R. (1997). A discrete binary version of the particle swarm algorithm. In *Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on*, vol. 5. ISSN 1062-922X, 4104–4108 vol.5. doi:10.1109/ICSMC.1997.637339.

Khushaba, R. N., Al-Ani, A. and Al-Jumaily, A. (2008). Differential evolution based feature subset selection. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. 1–4.

Kiani, A. and Akbarzadeh, M. R. (2006). Automatic Text Summarization Using

Hybrid Fuzzy GA-GP. In *Fuzzy Systems, 2006 IEEE International Conference on*. 977–983.

Kim, M. and Choi, K. (1999). A comparison of collocation-based similarity measures in query expansion. *Information processing & management*. 35(1), 19–30.

Kogilavani, A. and Balasubramanie, P. (2010). Clustering based optimal summary generation using Genetic Algorithm. In *Communication and Computational Intelligence (INCOCCI), 2010 International Conference on*. 324–329.

Kong, M. and Tian, P. (2005). A binary ant colony optimization for the unconstrained function optimization problem. In *Proceedings of the 2005 international conference on Computational Intelligence and Security - Volume Part I*. CIS'05. Berlin, Heidelberg: Springer-Verlag. ISBN 3-540-30818-0, 978-3-540-30818-8, 682–687. doi:10.1007/11596448_101. Retrievable at `http://dx.doi.org/10.1007/11596448_101`.

Kraaij, W., Spitters, M. and Van Der Heijden, M. (2001). Combining a mixture language model and naive bayes for multi-document summarisation. In *Working notes of the DUC2001 workshop (SIGIR2001), New Orleans*. Citeseer.

Kumar, C., Pingali, P. and Varma, V. (2008). Generating personalized summaries using publicly available web documents. In *Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on*. (pp. 103–106). vol. 3.

Kuo, J.-J. and Chen, H.-H. (2008). Multidocument summary generation: Using informative and event words. *ACM Transactions on Asian Language Information Processing (TALIP)*. 7(1), 3.

Kwedlo, W. (2011). A clustering method combining differential evolution with the¡ i¿ K¡/i¿-means algorithm. *Pattern Recognition Letters*. 32(12), 1613–1621.

Ledeneva, Y., Gelbukh, A. and Garca-Hernndez, R. (2008). Terms derived from frequent sequences for extractive text summarization. *Computational Linguistics and Intelligent Text Processing*, 593–604.

Lerman, K. and McDonald, R. (2009). Contrastive summarization: an experiment with consumer reviews. In *Proceedings of human language technologies: The 2009 annual conference of the North American chapter of the association for computational linguistics, companion volume: Short papers*. (pp. 113–116).

Lin, C. and Hovy, E. (2003). Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. Association for Computational Linguistics, 71–78.

Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of summaries. In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*. 74 – 81.

Lin, C.-Y. and Hovy, E. (1997). Identifying topics by position. In *Proceedings of the fifth conference on Applied natural language processing*. 283–290.

Lin, C.-Y. and Hovy, E. (2002). From single to multi-document summarization: a prototype system and its evaluation. In *Association for Computational Linguistics*. 457–464.

Lloret, E. and Palomar, M. (2012). Text summarisation in progress: a literature review. *Artificial Intelligence Review*. 37(1), 1–41.

Lu, Y., Lu, S., Fotouhi, F., Deng, Y. and Brown, S. (2004). FGKA: a fast genetic K-means clustering algorithm. In *Proceedings of the 2004 ACM symposium on Applied computing*. ACM, 622–623.

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM J. Res. Dev.* 2(2), 159–165.

Luo, W., Zhuang, F., He, Q. and Shi, Z. (2010). Effectively Leveraging Entropy and Relevance for Summarization. In *Information Retrieval Technology: 6th Asia Information Retrieval Societies Conference, AIRS 2010, Taipei, Taiwan, December 1-3, 2010, Proceedings*, vol. 6458. Springer, 241.

Macqueen, J. B. (1967). Some Methods for classification and analysis of multivariate observations. In *Procedings of the Fifth Berkeley Symposium on Math, Statistics, and Probability*, vol. 1. University of California Press, 281–297.

Mallipeddi, R., Suganthan, P. N., Pan, Q. K. and Tasgetiren, M. F. (2011). Differential evolution algorithm with ensemble of parameters and mutation strategies. *Appl. Soft Comput.* 11(2), 1679–1696.

Mani, I. (2001). *Automatic summarization*. vol. 3. John Benjamins Publishing Company.

Mani, I. and Maybury, M. (1999). *Advances in automatic text summarization*. MIT press.

Marcu, D. (1999). Discourse trees are good indicators of importance in text. *Advances in automatic text summarization*, 123–136.

Maynard, D., Tablan, V., Ursu, C., Cunningham, H. and Wilks, Y. (2001). Named entity recognition from diverse text types. In *Recent Advances in Natural Language Processing 2001 Conference*. 257–274.

Miller, G. and Fellbaum, C. (1998). *Wordnet: An electronic lexical database*.

Mori, T., Nozawa, M. and Asada, Y. (2005). Multi-answer-focused multi-document summarization using a question-answering engine. *ACM Transactions on Asian Language Information Processing (TALIP)*. 4(3), 305–320.

Murad, M. and Martin, T. (2007). Similarity-based estimation for document summarization using Fuzzy sets. *International Journal of Computer Science and Security*. 1(4), 1–12.

Nenkova, A. (2006). Summarization evaluation for text and speech: issues and approaches. *INTERSPEECH-2006, paper*.

Nenkova, A., Passonneau, R. and McKeown, K. (2007). The pyramid method: Incorporating human content selection variation in summarization evaluation. *ACM Transactions on Speech and Language Processing (TSLP)*. 4(2), 4.

Neto, J. L., Freitas, A. A. and Kaestner, C. A. A. (2002). Automatic text summarization using a machine learning approach. *Advances in Artificial Intelligence, Proceedings*. 2507, 205–215.

Neto, S. A. D. K. C. A. A. F. A. A., J. L. (2000). Generating text summaries through the relative importance of topics. *Advances in Artificial Intelligence*. 1952, 300–309.

Nobata, C., Sekine, S., Murata, M., Uchimoto, K., Utiyama, M. and Isahara, H. (2001). Sentence Extraction System Assembling Multiple Evidence. In *Proceedings of the Second NTCIR Workshop Meeting*. 319–324.

Norusis, M. (1990). *SPSS advanced statistics user's guide*. SPSS Chicago.

Omran, M., Engelbrecht, A. and Salman, A. (2005a). Differential evolution methods for unsupervised image classification. In *Evolutionary Computation, 2005. The 2005 IEEE Congress on*, vol. 2. 966–973 Vol. 2. doi:10.1109/CEC.2005.1554795.

Omran, M., Salman, A. and Engelbrecht, A. (2005b). Dynamic clustering using particle swarm optimization with application in unsupervised image classification. In *Fifth World Enformatika Conference (ICCI 2005), Prague, Czech Republic*. 199–204.

Omran, M. G. H., Engelbrecht, A. P. and Salman, A. (2005c). Differential evolution methods for unsupervised image classification. In *Evolutionary Computation, 2005. The 2005 IEEE Congress on*, vol. 2. 966–973 Vol. 2.

Over, P., Dang, H. and Harman, D. (2007). DUC in context. *Information processing & management*. 43(6), 1506–1520.

Pampara, G., Engelbrecht, A. P. and Franken, N. (2006). Binary Differential Evolution. In *Evolutionary Computation, 2006. CEC 2006. IEEE Congress on*. 1873–1879.

Pampara, G., Franken, N. and Engelbrecht, A. (2005). Combining Particle Swarm

Optimisation with angle modulation to solve binary problems. In *Evolutionary Computation, 2005. The 2005 IEEE Congress on*, vol. 1. IEEE, 89–96.

Papineni, K., Roukos, S., Ward, T. and Zhu, W.-J. (2002). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. ACL '02. Stroudsburg, PA, USA: Association for Computational Linguistics, 311–318. doi: 10.3115/1073083.1073135. Retrievable at `http://dx.doi.org/10.3115/1073083.1073135`.

Passonneau, R. J. (2010). Formal and functional assessment of the pyramid method for summary content evaluation. *Natural Language Engineering*. 16(02), 107–131.

Pearson, K. (1920). Notes on the history of correlation. *Biometrika*. 13(1), 25–45.

Peng, C., Jian, L. and Zhiming, L. (2008). Solving 0-1 Knapsack Problems by a Discrete Binary Version of Differential Evolution. In *Intelligent Information Technology Application, 2008. IITA '08. Second International Symposium on*, vol. 2. 513–516. doi:10.1109/IITA.2008.538.

Peraki, R., Petrakis, E. G. and Hliaoutakis, A. (2012). An information retrieval system for expert and consumer users. In *Bioinformatics & Bioengineering (BIBE), 2012 IEEE 12th International Conference on*. IEEE, 145–150.

Pooya Khosraviyan Dehkordi, H. K., Farshad Kumarci (2009). Text Summarization Based on Genetic Programming. *Journal of Computing and ICT Research (IJCIR)*. Vol.3 No.1, 57–64.

Porter, M. F. (1997). *An algorithm for suffix stripping*, Morgan Kaufmann Publishers Inc. 313–316.

Pourvali, M. and Abadeh, M. S. (2012). Automated Text Summarization Base on Lexicales Chain and graph Using of WordNet and Wikipedia Knowledge Base. *arXiv preprint arXiv:1203.3586*.

Radev, D., Hovy, E. and McKeown, K. (2002). Introduction to the special issue on summarization. *Computational linguistics*. 28(4), 399–408.

Radev, D. R. and Tam, D. (2003). Summarization evaluation using relative utility. In *Proceedings of the twelfth international conference on Information and knowledge management*. ACM, 508–511.

Rahnamayan, S. and Tizhoosh, H. R. (2008). *Differential Evolution Via Exploiting Opposite Populations*, Springer Berlin Heidelberg, *Studies in Computational Intelligence*, vol. 155, chap. 8. 143–160.

Rahnamayan, S., Tizhoosh, H. R. and Salama, M. M. A. (2007). A novel population initialization method for accelerating evolutionary algorithms. *Computers and Mathematics with Applications*. 53(10), 1605–1614.

Rahnamayan, S., Tizhoosh, H. R. and Salama, M. M. A. (2008a). Opposition-Based Differential Evolution. *Evolutionary Computation, IEEE Transactions on*. 12(1), 64–79.

Rahnamayan, S., Tizhoosh, H. R. and Salama, M. M. A. (2008b). Opposition versus randomness in soft computing techniques. *Applied Soft Computing*. 8(2), 906–918.

Rahnamayan, S. and Wang, G. G. (2008). Solving large scale optimization problems by opposition-based differential evolution (ODE). *W. Trans. on Comp.* 7(10), 1792–1804.

Reed, J. W., Jiao, Y., Potok, T. E., Klump, B. A., Elmore, M. T. and Hurson, A. R. (2006). *TF-ICF: A New Term Weighting Scheme for Clustering Dynamic Data Streams*.

Ren Arnulfo Garca-Hernndez, G. M. M. n. H. D. J. C.-A. G. J. L. T. F., Yulia Ledeneva (2009). *Comparing Commercial Tools and State-of-the-Art Methods for Generating Text Summaries*.

Saggion, H. and Poibeau, T. (2013). Automatic Text Summarization: Past, Present and Future. In Poibeau, T., Saggion, H., Piskorski, J. and Yangarber, R. (Eds.) *Multi-source, Multilingual Information Extraction and Summarization*. (pp. 3–21). Theory and Applications of Natural Language Processing. Springer Berlin Heidelberg. ISBN 978-3-642-28568-4. doi:10.1007/978-3-642-28569-1_1. Retrievable at `http://dx.doi.org/10.1007/978-3-642-28569-1_1`.

Saggion, H., Radev, D., Teufel, S., Lam, W. and Strassel, S. (2002). Developing infrastructure for the evaluation of single and multi-document summarization systems in a cross-lingual environment. *Ann Arbor*. 1001, 48109–1092.

Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Inf. Process. Manage.* 24(5), 513–523.

Schilder, F. and Kondadadi, R. (2008). FastSum: fast and accurate query-based multi-document summarization. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*. (pp. 205–208).

Schilder, F., Kondadadi, R., Leidner, J. L. and Conrad, J. G. (2008). Thomson reuters at tac 2008: Aggressive filtering with fastsum for update and opinion summarization. In *Proceedings of the first Text Analysis Conference, TAC-2008*.

Schuler, K. K. (2005). VerbNet: A broad-coverage, comprehensive verb lexicon. Retrievable at `http://repository.upenn.edu/dissertations/AAI3179808`.

Shahzad, F., Baig, A., Masood, S., Kamran, M. and Naveed, N. (2009). *Opposition-Based Particle Swarm Optimization with Velocity Clamping (OVCPSO) Advances in Computational Intelligence*, Springer Berlin / Heidelberg, *Advances in Intelligent and Soft Computing*, vol. 116. 339–348.

Shen, D., Sun, J.-T., Li, H., Yang, Q. and Chen, Z. (2007). Document summarization using conditional random fields. In *IJCAI*, vol. 7. 2862–2867.

Shi, Z., Melli, G., Wang, Y., Liu, Y., Gu, B., Kashani, M. M., Sarkar, A. and Popowich, F. (2007). Question answering summarization of multiple biomedical documents. In *Advances in Artificial Intelligence*. (pp. 284–295). Springer.

Song, W. and Park, S. C. (2008). *An Improved Genetic Algorithm for Document Clustering with Semantic Similarity Measure*.

Sridhara, G., Hill, E., Pollock, L. and Vijay-Shanker, K. (2008). Identifying Word Relations in Software: A Comparative Study of Semantic Similarity Tools. In *Program Comprehension, 2008. ICPC 2008. The 16th IEEE International Conference on*. ISSN 1092-8138, 123–132. doi:10.1109/ICPC.2008.18.

Steinbach, M., Karypis, G. and Kumar, V. (2000). A comparison of document clustering techniques. In *KDD workshop on text mining*, vol. 400. Boston, 525–526.

Steinberger, J. and Jezek, K. (2004). Text summarization and singular value decomposition. *Advances in Information Systems, Proceedings*. 3261, 245–254.

Steinberger, J. and Jezek, K. (2009a). Evaluation Measures for Text Summarization. *Computing and Informatics*. 28(2), 251–275.

Steinberger, J. and Jezek, K. (2009b). *Text Summarization: An Old Challenge and New Approaches Foundations of Computational, IntelligenceVolume 6*, Springer Berlin / Heidelberg, *Studies in Computational Intelligence*, vol. 206. 127–149.

Steinberger, J., Kabadjov, M., Poesio, M. and Sanchez-Graillet, O. (2005). Improving LSA-based summarization with anaphora resolution. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 1–8.

Storn, R. and Price, K. (1997). Differential Evolution A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *J. of Global Optimization*. 11(4), 341–359.

Suanmali, L., Salim, N. and Binwahlan, M. (2011a). Fuzzy Genetic Semantic Based Text Summarization. In *Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on*. IEEE, 1184–1191.

Suanmali, L., Salim, N. and Binwahlan, M. S. (2011b). Genetic Algorithm Based for Sentence Extraction in Text Summarization. *International Journal of Innovative Computing*. 1(1).

Svore, K., Vanderwende, L. and Burges, C. (2007). Enhancing single-document summarization by combining RankNet and third-party sources. In *Proceedings of the EMNLP-CoNLL*. 448–457.

Teufel, S. and Van Halteren, H. (2004). Evaluating Information Content by Factoid Analysis: Human annotation and stability. In *EMNLP*. 419–426.

Van Rijsbergen, C. (1974). Foundation of evaluation. *Journal of Documentation*. 30(4), 365–373.

Vesterstrom, J. and Thomsen, R. (2004). A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems. In *Evolutionary Computation, 2004. CEC2004. Congress on*, vol. 2. 1980–1987 Vol.2.

Vossen, P. (1998). *EuroWordNet: a multilingual database with lexical semantic networks*. Kluwer Academic Boston.

Wan, X. and Xiao, J. (2007). Towards a unified approach based on affinity graph to various multi-document summarizations. In *Research and Advanced Technology for Digital Libraries*. (pp. 297–308). Springer.

Wang, H., Wu, Z., Rahnamayan, S. and Wang, J. (2010a). *Diversity Analysis of Opposition-Based Differential EvolutionAn Experimental Study Advances in Computation and Intelligence*, Springer Berlin / Heidelberg, *Lecture Notes in Computer Science*, vol. 6382. 95–102.

Wang, L., Fu, X., Menhas, M. and Fei, M. (2010b). A modified binary differential evolution algorithm. *Life System Modeling and Intelligent Computing*, 49–57.

Wilks, Y., Galescu, L., Allen, J. and Dalton, A. (2013). Automatic Metaphor Detection using Large-Scale Lexical Resources and Conventional Metaphor Extraction. *Meta4NLP 2013*, 36.

Xin, Y., Yong, L. and Guangming, L. (1999). Evolutionary programming made faster. *Evolutionary Computation, IEEE Transactions on*. 3(2), 82–102.

Yeh, J. Y., Ke, H. R., Yang, W. P. and Meng, I. H. (2005). Text summarization using a trainable summarizer and latent semantic analysis. *Information processing*

& management. 41(1), 75–95.

Zajic, D., Dorr, B. and Schwartz, R. (2002). Automatic headline generation for newspaper stories. In *Proceedings of the workshop on text summarization*. 78–85.

Zha, H. (2002). Generic summarization and keyphrase extraction using mutual reinforcement principle and sentence clustering. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 113–120.

Zhou, L. and Hovy, E. (2004). Template-Filtered headline summarization. *Text Summarization Branches Out: Pr ACL-04 Wkshp, July*.