

**THE REPUBLIC OF TURKEY
BAHCESEHIR UNIVERSITY**

**CONFLICT ANALYSIS FOR TURKISH DEBATES
USING TEXT MINING AND TEXT
SEGMENTATION TECHNIQUES**

Master Thesis

EFE BÜYÜK

ISTANBUL, 2016

**THE REPUBLIC OF TURKEY
BAHCESEHIR UNIVERSITY**

**GRADUATE SCHOOL OF NATURAL AND APPLIED
SCIENCES
INFORMATION TECHNOLOGIES**

**CONFLICT ANALYSIS FOR TURKISH DEBATES
USING TEXT MINING AND TEXT
SEGMENTATION TECHNIQUES**

Master Thesis

EFE BÜYÜK

ASSISTANT PROF.DR. DİLEK KARAHOCA

ISTANBUL, 2016

THE REPUBLIC OF TURKEY

BAHCESEHIR UNIVERSITY

**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
INFORMATION TECHNOLOGIES**

Name of the thesis: Conflict Analysis for Turkish Debates using Text Mining
and Text Segmentation Techniques
Name/Last Name of the Student: Efe Büyük
Date of the Defense of Thesis: 31.08.2016

The thesis has been approved by the Graduate School of _____.

Associate Prof. Nafiz ARICA
Graduate School Director
Signature

I certify that this thesis meets all the requirements as a thesis for the degree of
Master of Arts.

Prof. Dr. Adem KARAHOCA
Program Coordinator
Signature

This is to certify that we have read this thesis and we find it fully adequate in
scope, quality and content, as a thesis for the degree of Master of Arts.

Examining Committee Members

Signature

Thesis Supervisor
Assistant Prof. Dr. Dilek KARAHOCA

Member
Prof. Dr. Adem KARAHOCA

Member
Prof. Dr. Hüseyin UZUNBOYLU

DEDICATION

This thesis is dedicated to my beloved mother.
For her endless love, support and encouragement.



PREFACE

In Today's World, we have so many discussions and debates that seem as problems. However, when we examine them technically, it is easy to understand that they are actually not problems but there are lack of communication and misunderstandings between people. At this point, it is crucial to explain the aim of this work. The main purpose is to show that languages can be analyzed technically and meanings of words and sentences can be understood through sentiment points. These points can be accepted as almost same for every language so that this type of work can be adapted to any of them. In this work, Turkish language is selected for examination purpose because there is not much work related with this language. Moreover, this work should be counted as experimental because the database that we are using is from another language that is English. That is translated to Turkish for the sake of our work. It is important to know that these results should be increased by other works and experiments so that more accurate outputs will occur.

Finally, I would like to thank Assistant Prof. Dr. Dilek Karahoca for their support and help during my thesis. I had this topic in my mind to study as my thesis work but after many discussions with them, we could draw final boundaries of my study.

Istanbul / 30.05.2016

Efe Byk

ABSTRACT

CONFLICT ANALYSIS FOR TURKISH DEBATES USING TEXT MINING AND TEXT SEGMENTATION TECHNIQUES

Efe Büyük

Information Technologies

Thesis Supervisors: Assistant Prof. Dr. Dilek Karahoca

June 2016, 57 pages

Conflict Analysis is one of the most challenging issues in the world that many organizations and governments try to carry out perfectly. It is crucial to have a correct analysis to prepare a resolution for a problem. Thus, this thesis paper focuses on the ways that a software program can detect the reasons of arguments in a debate. The examples of debate dialogs are chosen from Turkish language because there is not much research in this area with this language. Moreover, the techniques which are applied in this work can also be applied to other languages, because a sentiment word dictionary is used and sentiments are almost the same in every language. This is a prepared dictionary from SentiWordNet with all the sentiment points for English words. It is translated and extended for the Turkish language. Furthermore, both machine learning and lexicon-based approaches are implemented in order to increase the diversity of results. This paper aims to show that languages can be processed in a technical manner and meanings can be extracted from sentences to understand the reasons of arguments. Likewise, the main contribution of this thesis is that conflict analysis for Turkish debates can be applied with the techniques which are examined here and they are also suitable for other languages.

Keywords: Text Mining, Opinion Mining, Sentiment Analysis, Machine Learning

ÖZET

METİN MADENCİLİĞİ VE METİN BÖLÜMLEME TEKNİĞİ KULLANARAK TÜRKÇE TARTIŞMALARDA UYUŞMAZLIK ANALİZİ

Efe Büyük

Bilgi Teknolojileri

Tez Danışmanı: Yrd. Doç. Dr. Dilek Karahoca

Haziran 2016, 57 sayfa

Uyuşmazlık Analizi birçok kurum ve devlet için önemli bir konu olmasının yanı sıra bunu çok iyi bir şekilde yapma ihtiyacı içerisindeyler. Bir problemin çözümünde doğru analizin önemi bunda büyük rol oynamaktadır. Bu sebeple bu tez çalışması bir bilgisayar yazılımının bir tartışmadaki sebeplerin tespiti üzerine odaklanmıştır. Bu çalışmadaki tartışma diyalogları Türkçe'den seçilmiştir, çünkü bu dilde bu konu özelinde yeterince çalışma yoktur. Dahası, buradaki çalışma yöntemi diğer dillere de uygulanabilir. Çünkü bu çalışma için bir duygu sözlüğü kullanılmıştır ve duygular hemen her dilde aynıdır. Faydalanılan duygu sözlüğü daha önce hazırlanmış olan SentiWordNet'den alınmıştır. Bu sözlük İngilizce özelinde yapılmıştır. Bu çalışma özelinde ise Türkçe'ye çevrilmiştir. Bununla birlikte, makine öğrenmesi ve sözlük tabanlı yaklaşımlar ile sonuçların çeşitliliği arttırılmaya çalışılmıştır. Bu çalışmanın amacı dillerin teknik yollar ile analiz edilebileceğini ve dolayısı ile tartışmaların sebeplerinin bulunabileceğini göstermektir. Aynı şekilde, bu çalışmanın ana katkılarından biri de Türkçe tartışmalardaki uyuşmazlık analizi için uygulanan buradaki tekniklerin diğer diller için de uygulanabilir olacaktır.

Anahtar Kelimeler: Metin Madenciliği, Düşünce Madenciliği, Duygu Analizi, Makine Öğrenimi

CONTENTS

TABLES	ix
FIGURES	x
ABBREVIATIONS.....	xi
1. INTRODUCTION	1
2. LITERATURE SEARCH.....	2
2.1. INFORMATION ABOUT TURKISH LANGUAGE AND SENTIMENT ANALYSIS.....	2
2.1.1. General Rules in Turkish Language.....	2
2.1.2. Background to The Proposed Research	5
2.1.3. Identifying The Scope of Work	8
2.1.4. Research Questions to Be Investigated.....	9
2.1.5. Expected Inputs and Outputs.....	9
2.2. WHAT IS THE ANTICIPATED SIGNIFICANCE AND/OR IMPACT? 10	
2.2.1. Business and Organizations.....	10
2.2.2. Individuals.....	10
2.2.3. Ads Placements	11
2.2.4. Opinion Retrieval/Search	11
3. DATA AND METHOD	12
3.1. WHICH TECHNIQUES WILL BE USED DURING CONDUCTING THE RESEARCH?	12
3.1.1. Word & Sentence Limitations.....	12
3.1.2. Preparation of Database and Python Functions.....	12
3.1.3. Baseline Algorithm	13
3.2. WHAT RESEARCH METHODS WILL YOU USE TO UNDERTAKE THE PROPOSED RESEARCH?.....	14
3.2.1. Research Methodologies	14
3.3. WHAT ARE THE INTENDED TIMELINES TO UNDERTAKE THE PROPOSED RESEARCH?	25
3.4. WHAT RESOURCES, FACILITIES, EQUIPMENT, FIELDWORK WILL BE REQUIRED TO UNDERTAKE THE PROPOSED RESEARCH?	25
3.4.1. Resources.....	25
3.4.2. Facilities.....	26
3.4.3. Equipment	26
3.4.4. Fieldwork.....	26

4. RESULTS	27
4.1. HOW WILL THE PROPOSED RESEARCH CONTRIBUTE TO THE RESEARCH OUTPUTS AND RESEARCH CULTURE IN THE RESEARCH AREA?	27
4.2. STRUCTURE OF THE SOFTWARE	27
4.3. MAIN FEATURES	29
4.4. ANALYZING THE PY FILES	30
4.4.1. Analysis.py	30
4.4.2. DB_operations.py	35
4.4.3. Description.py	37
4.4.4. File_operations.py	37
4.4.5. Greeting.py	38
4.4.6. IO_operations.py	38
4.4.7. Main.py	38
4.5. ANALYZING THE DATABASE STRUCTURE	39
4.5.1. Database Table 1: allwords	39
4.5.2. Database Table 2: turkce_kelimeler	40
4.6. EXECUTING THE SOFTWARE	41
5. FINDINGS	50
5.1. MEASUREMENT	50
5.2. HIGHLIGHTS	54
5.2.1. Comparison Between the Initial Hypothesis and the Current Results.. 54	
6. DISCUSSION	55
6.1. THEORETICAL IMPLICATIONS	55
6.2. PRACTICAL IMPLICATIONS	55
7. CONCLUSION	57
REFERENCES	58

TABLES

Table 2.1: Illustration of Result Screen	9
Table 3.1: Differences between Positivist and Phenomenology Paradigm	14
Table 3.2: Comparison between Deductive and Inductive Methods.....	15
Table 3.3: Differences between Exploratory and Conclusive Research	17
Table 3.4: Differences between Qualitative and Quantitative Research Methods	18
Table 5.1: Calculated scores and true states of sentences in the files	51
Table 5.2: Comparison between test results and true sentence states	53

FIGURES

Figure 3.1: Deductive Approach Flow	16
Figure 3.2: Inductive Approach Flow	16
Figure 3.3: Sampling Example	19
Figure 3.4: Master Thesis Timeline.....	25
Figure 4.1: Beginning of the software	41
Figure 4.2: Searching a word in the software	41
Figure 4.3: Result screen of searching a word.....	42
Figure 4.4: New search screen for a word	42
Figure 4.5: Represents the result of an unknown word	43
Figure 4.6: Selecting the type of the word.....	43
Figure 4.7: Typing the meaning of the word	44
Figure 4.8: Giving a score for the word	44
Figure 4.9: Saving the new word into the database	44
Figure 4.10: Program returns to the beginning.....	45
Figure 4.11: List of available file names	46
Figure 4.12: Choosing a file	47
Figure 4.13: Software prints the values of the next word in the sentence.....	47
Figure 4.14: Showing the first sentence's sentiment score	48
Figure 4.15: Showing the second sentence's sentiment score.....	48
Figure 4.16: Showing all sentences' scores with their other values.....	48
Figure 4.17: Program returns to the beginning.....	48

ABBREVIATIONS

FSR	: First Sentence Rate
FSTS	: First Sentence True State
SSR	: Second Sentence Rate
SSTS	: Second Sentence True State



1. INTRODUCTION

This master thesis aims to show that standings of people for a specific debate or discussion can be understood by the help of sentiment analysis. This means that when there are two opposite opinions for a topic, we can find who is positive or negative about the related debate. In the light of this information, it is clear to understand why this type of work is necessary in our today's world. It is important because there are lots of debates and some of them have huge history that people cannot find a common ground in order to solve the problem. The more a debate takes time, the less chance there is to solve it because the total amount of words and sentences that are spoken increase constantly. That is why we need software programs to examine debates and find a common ground for all parties. Furthermore, when it comes to understanding this master thesis, it is necessary for readers to know or get familiar with the following topics beforehand that they are text mining, opinion mining, sentiment analysis and machine learning. This work uses Turkish as language because there is not much work related with it. Most of work are done for English or other European languages. To sum up, this work's purpose is to make contribution for the technique of solving Turkish debates. In the future, people can create a more accurate mechanism to detect people's standings as positive and negative and make suggestions to them like showing the common ground. In the following sections of this master thesis, the chapters of result, discussion and conclusion can give more idea about the output of this work and future thoughts.

2. LITERATURE SEARCH

2.1. INFORMATION ABOUT TURKISH LANGUAGE AND SENTIMENT ANALYSIS

In this section, there are information about Turkish language and sentiment analysis. It is good to know some information about these in order to understand this study.

2.1.1. General Rules in Turkish Language

Turkish is one of the oldest languages in the world. According to the turkishlanguage.co.uk (2016) web site: “The Turkish Language originated in The Altai Mountain Range in Northern Siberia centuries ago. For this reason, it is called an Altaic Language. As nomads expanded further into Asia Minor, they brought their language with them to Turkmenistan, Kazakhstan, Uzbekistan, Azerbaijan and other countries. Many of these languages are mutually intelligible although local usage and vocabulary, spelling and alphabet may differ. However, they all exhibit the same grammatical structure of agglutination and vowel harmony.

Turkish being a language emanating from Central Asia, is spoken from the borders of Greece into the hinterland of Western China. While the Ottoman Empire flourished Turkish was spoken from Vienna to Arabia, Egypt and Northern Africa. The Turkish vocabulary contains many words from Arabic, Persian and European languages. These imported words mostly follow the basic grammar and vowel harmony of native Turkish.”

Of course, languages are not static and like the others, Turkish language also continue to evolve. This evolving changes in different time periods. For example, the gold generation who established the Republic of Turkey has learned read and write in an inductive way and they have got a very positive read and write habits (Gülsevin and Boz 2006). As can be understood from the above example, Turkish language is also a dynamic language and its today form has also other type of learning ways that emoji is

used in every language right now and this should be also considered in this type of study.

2.1.1.1. An Overview of the Turkish Language

There are many rules in Turkish like in every language but some of the rules are very different than the other languages. According to the turkishlanguage.co.uk (2016) web site, the below is a list of those rules in general.

- i. In Turkish word meanings are changed by fixing other words on to the root as direct suffixes.
- ii. These little important words show *motion towards/motion from* or *static location* of the suffixed word.
- iii. The suffixes change their spelling according to set rules and they must follow the same vowel pattern *Vowel Harmony* as the word to which they are being added. Sometimes they also have a consonant change *Consonant Mutation* for ease of pronunciation.
- iv. Those added to the stem of a verb may indicate its positive or negative form. Suffixes are then added for tense and person.
- v. Further *moods may, might, can, can't, must* can be added on to the original verb root, thus producing a new verb.
- vi. Nouns are suffixed with possessor and the motion or location words are then added.
- vii. There is no definite article "the" as a subject, but there is a specifier "the" as a direct object suffix.
- viii. There are no gender forms (no "le" or "la" as in French) in Turkish. One single word is used for "he, she, it"
- ix. As in English, adjectives describe their noun and remain in their basic form: there being no gender thus no gender agreement is required.
- x. The sentence form is SOV Subject, Object, Verb.

2.1.1.2. Affirmative Sentences in Turkish Language

Affirmative sentences in Turkish can be found in the four main sentence types and that is the sentences which are categorized by their meanings (Ertav 2014, p. 20). In Turkish, affirmative sentences are the sentences which their verbs declare that action is completed or done (Göker 1997). According to the Turkish Philology Encyclopedia, there is an almost same definition for affirmative sentences and that is if a sentence has a positive verb, then that sentence is an affirmative sentence (Kutlu 1977). Even though affirmative or positive sentences seem easy to spot and calculate their sentiment points, it is not. For instance, the sentence “Büyük zorlukların, yoklukların, olanaksızlıkların içerisinde, kanla ve terle çıkarılmış bir utkudur bizim Kurtuluş Savaşı.” in Turkish has a verb and a subject but their positions in this sentence are different. The subject is at the end of the sentence and that means it is an inverted sentence (Gökalp 1976, p.157). Moreover, according to Istanbul University Faculty of Literature there is another trouble which makes even harder this thesis problem that there are some sentences which their structures are affirmative but negative in their meanings (2006).

2.1.1.3. Negative Sentences in Turkish Language

Negativity in words and terms are different in every language and all of them are represented with different appendixes and structures. For example, the sentence “Ben okula gitmeyeceğim.” in Turkish language is formed differently in other languages like “I won’t go to school” in English, “Ya ne paydu v şkolu” in Russian, “ich gehe nicht in die Schule” in German (İlhan 2007). Moreover, sentences’ have many variations in the sense of their meanings. For example, sentence’s structure is positive but its meaning can be negative. That is why when classifying the sentences which are negative in meaning and affirmative in structure, it should be based on definitions of negativeness concept on both the action does not happen and the action which happens is not in the wanted and expected direction (Demirel 2009).

In addition to that, there are some words in Turkish which are negative in structure but they do not make the sentence’s meaning completely negative but only partially (Demirel 2002). Furthermore, there is another good example for showing the difficulties

in negative sentences in Turkish that affixed words like the one “-li” in Turkish makes harder in pronunciations when it comes to say “değildir / isn’t” (Akdeniz 1973). Last but not least, there is also some words that they have negative side meanings and these kinds of words should not be used in a positive way. For example, “Daha fazla şansınızın olması için her paketten iki tane alınız.” means that “If you want to increase your chance, you need to buy two of them for each package.”. In this sentence, instead of the word “fazla”, “çok” word could be used because word “fazla” can have negative side meanings like “excessive” (Özaydın 2015). According to the Turkish Language Association’s 2006 annual book, T. N. Gencan supports the idea that if sentences are long enough and some intensifier words or supplement words are added between the main words and verbs are at the beginning, then those sentences have probably negative meanings (Türk Dil Kurumu 2006). Situations like these make difficult to determine the sentence’s sentiment points.

2.1.2. Background to The Proposed Research

There are two main types of textual information on the Web: Facts and Opinions. Current search engines search for facts that they assume that information is true. Search engines do not search for opinions. Opinions are hard to express with topic keywords. Current search ranking strategy is not appropriate for opinion retrieval/search.

2.1.2.1. What is Sentiment Analysis?

Sentiment analysis’s past is more than most of people would think. According to Ahmad, the origin of the sentiment analysis lies in the mid-twentieth century political science and pioneers were Harold Lasswell and Philip Stone. They developed content-analysis systems for analyzing speeches by politicians and manifesto of political parties (2011).

Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. As also indicated in a research, the purpose of sentiment analysis is to determine the overall opinion of the user review in terms of positive and negative

(Mandal and Kar and Das and Panigrahi 2015). It's also known as opinion mining, deriving the opinion or attitude of a speaker. A common use case for this technology is to discover how people feel about a particular topic.

According to Liu, people's sentiments, opinions, attitudes, and emotions are for the computational study in sentiment analysis. This interesting problem is crucial in our daily business and society life. This area is a big challenge for interested researchers especially in opinion analysis and social media analysis (2015).

Additionally, there is another interesting study which was conducted by Steels, Kaplan, McIntyre and Looveren at Paris at the Sony Computer Science Laboratory in Paris. This study is important because it forms the basis of sentiment analysis. According to their study, they focused on why language has evolved (2000).

Last but not least, what is more important about Sentiment Analysis is that you can even learn why people think the idea is good or bad, by extracting the exact words that indicate why people did or didn't like it. For example, in the experiment of analyzing Twitter's data, when people mention about a food which they tried and said that it is too salty, then it shows as a common theme and you immediately have a better idea of why consumers aren't happy.

2.1.2.2. Problems with Sentiment Analysis

Sentiment analysis has problems in its own and it actually deals with the evaluation type of opinions or opinions which imply positive or negative sentiments but opinions itself is a broad concept and not all researchers agree with a set of emotions (Liu 2012). Moreover, there are common formulations of classification problems in sentiment analysis and opinion mining and also ranking problems (Pang and Lee 2008, pp. 26). For example, in this study ranking problem is very popular because there is not a ready database for all the Turkish words that the software could use. However, the SentiWordNet has a database of English words which has been created for a long time period. Although there is not a ready Turkish sentiment words' database, it is being created with the one that is coded with this master thesis study.

Furthermore, emerging new ideas for an existing opinion is also a problem in sentiment analysis. When this happens, the new information is not aggregated to an existing opinion for the given entity (Tkalčič and De Carolis and De Gemmis and Odić and Košir 2016). According to the researchers mentioned, there is an experiment with Twitter data and the example is related with Nexus4 smartphone. Opinions about this device is formed by a set of posts expressing sentiments. However, opinions about this device may change when there is a new bug or technical problem. Sentiment analysis's techniques should therefore spot opinion changes for these entities. Moreover, the main of goal of sentiment analysis especially in product reviews on the Internet is to rank the entire review or separate into different categories or strength polarities like strongly-positive, weakly-positive, fair, weakly-negative and strongly-negative (Management Association 2013). However, reviews are generally a combination of affirmative and negative opinions.

2.1.2.3. Areas to Use Sentiment Analysis

There are many areas to use sentiment analysis. If there is a text related area, then probably sentiment analysis can be applied to it. As Singh and Husain explained in their study, when there is a decision making moment, both for an individual or organizational level, then search of other's opinion becomes important (2014). For example, nowadays blogs are a big part of our life and according to Balijepalli, the texts in these blogs generally represents the sentiments of people. There are many topics there including sports, politics, business, entertainment and etc. He also emphasizes that a good sentiment detection tool is necessary for us in order to make better analysis (2008).

Moreover, according to a research, the subject areas are also retail, call centers, financial institutions and telecommunications (Ishikiriyama and Miro and Gomes 2015). In addition to that, many of the text analysis techniques can be applied to diverse subject areas like marketing, politics and medicine. In order to obtain the best possible results, a certain degree of domain-subject knowledge is required (Kagan and Rossini and Sapounas 2013). That is why in this master thesis, firstly a Turkish sentiment database for the words is created and the language's properties are studied in order to determine which part of them can be applied to the software.

Furthermore, sentiment analysis can also be used to analyze organizations in social networks. When this analysis is completed, then you can understand that which areas in your organization is performing well and what areas may need some more work (Pierson 2015). In addition to that, there is a very related research and study area that local governments use Twitter data for sentiment analysis and they try to understand their people's opinions by doing that. One example about this is that this kinds of analysis focused on the capital city local governments and a sample of six Sydney suburban local governments. These governments were chosen because their social media use was significant enough to produce results (Sobaci 2015).

2.1.2.4. Why Does Sentiment Analysis Become Popular?

There are lots of reasons to answer this question but the first answer which is more imaginable is that people want to know what the other people think. Besides that, the below factors are determined by Pang and Lee (2008, pp. 4 - 5);

- i. The rise of machine learning methods in natural language processing and information retrieval
- ii. The availability of datasets for machine learning algorithms
- iii. Realization of the fascinating intellectual challenges and commercial and intelligence applications that the area offers.

2.1.3. Identifying The Scope of Work

The topic which I have chosen is very difficult and that is why my work will be experimental. It is my hope that I can make some contributions on this field.

It is important to describe what I am going to try to solve in my thesis work. That is why I have to draw some lines and set some rules for inputs which in this case will be Turkish debates. Every text will show a basic Turkish debate that basically two parties' arguments will be there. At the end, the output will be the sentence sentiment points of

the both parties. The following can be a good template example as a result of conflict analysis;

Table 2.1: Illustration of Result Screen

Topic: Here written the topic (discussion)	
Causes of Party A (Stand: Positive)	Causes of Party B (Stand: Negative)
Sentence Sentiment Point	Sentence Sentiment Point

2.1.4. Research Questions to Be Investigated

- i. Can a software program find arguments or reasons of a specific thought by specifying some key words?
- ii. Can a software program detect the familiarities between two different arguments by using words' sentiment points?
- iii. Can we understand one word's different meanings or sentiments in one sentence by using words' sentiment points?

2.1.5. Expected Inputs and Outputs

For the sake of my experimental research work in this thesis, I will keep it simple and give the topic and stand statuses of both parties before expecting an analysis. By doing so, it will be a bit easier to specify the causes during getting the related texts as inputs. To make it clear, let's give a simple example;

- i. **Given Topic:** "Bazı suç vakalarında ölüm cezasının uygulanması" (Execution of death punishment for some criminal cases)
- ii. **Given the Stand Status of Party A:** ['Merve', 'İdam cezası olmalı.', 1.0] (First column: Name; Second column: Sentence; Third column: Sentence Sentiment Point)
- iii. **Given the Stand Status of Party B:** ['Ali', 'İdam cezası olmamalı.', -1.0] (First column: Name; Second column: Sentence; Third column: Sentence Sentiment Point)

In the light of these information, it is easier to identify the sentiment points of A and B's sentences. Of course, there will be more complicated examples which contains more complex sentences. In this way, I can prove that until what extend my work can be useful in order to make a conflict analysis on Turkish debates.

2.2. WHAT IS THE ANTICIPATED SIGNIFICANCE AND/OR IMPACT?

There are mainly four types of areas which I detected that this research could be helpful.

2.2.1. Business and Organizations

Product and service benchmarking. Market intelligence. Conflict Analysis and Resolution.

- i. Business spends a huge amount of money to find consumer sentiment and opinions.
- ii. Consultants, surveys and focused groups, etc.
 - a. Governments and international political organizations want to solve conflicts among them and spend lots of money for it. For example, one research says that because sentiment analysis needs more opinions to put under examine, they focus on the political blogs (Balijepalli 2007).

2.2.2. Individuals

Interested in other's opinions when

- i. Purchasing a product or using service,
- ii. Finding opinions on political topics.

2.2.3. Ads Placements

Placing ads in the user-generated content

- i. Place an ad when one praises a product.
- ii. Place an ad from a competitor if one criticizes a product.

2.2.4. Opinion Retrieval/Search

Providing general search for opinions.



3. DATA AND METHOD

3.1. WHICH TECHNIQUES WILL BE USED DURING CONDUCTING THE RESEARCH?

There are some techniques which are used for this master thesis study and they are explained in this section.

3.1.1. Word & Sentence Limitations

I have decided to put some limitations for my Turkish debate examples that Turkish language like every other language is very flexible in the context of meaning of its words. For instance, some sentences can have a meaning of irony. In this case, it is more difficult to understand for a software program whether that sentence has a positive or a negative meaning. That is why my debate examples will have very simple sentences with straight meanings. The following is an example of this;

“Sinemaya gitmek istemiyorum çünkü başım ağrıyor.” (I do not want to go to the cinema because I have a headache.)

3.1.2. Preparation of Database and Python Functions

I found a prepared English sentiment word dictionary in the website of SentiWordNet. In this dictionary, there are English words with their positive or negative sentiment points. I will convert it to Turkish and use those points as they are because the meanings of sentiments are not changeable that much. For example, in this dictionary if a word “happy” has positive point of 1, then we can use this point as it is with its Turkish equivalent word “mutlu”. Moreover, I will extend my dictionary by adding grammatical rules of Turkish, so that my Python functions can get benefit from it.

3.1.3. Baseline Algorithm

The following is the list of techniques that can be used for this type of work;

- i. Tokenization
- ii. Feature Extraction

3.1.3.1. Tokenization

Tokenization is a way that sentences are divided into symbols, words, phrases and other elements in lexical analysis and these are called tokens. In parsing or text mining, the tokens are used as inputs. This method is very helpful in linguistics and in computer science.

Furthermore, while applying this method, keyword spotting technique can be also used because all the words in the database and when the software takes a sentence, it splits that into words and searches them one by one. This keyword spotting is seen as the most naïve approach because its accessibility and economy for coding (Cambria and Schuller and Xia and Havasi 2013, pp. 18).

3.1.3.2. Feature Extraction

Its main duty is to lower necessary needed resources in order to explain a huge set of data. One of the main problems is that when analyzing a data set, the number of variables can increase. The analysis which contains many variables need more memory space and processor power. Moreover, this situation decreases prediction performance. Feature extraction overcome these problems by forming different combinations of the variables.

3.2. WHAT RESEARCH METHODS WILL YOU USE TO UNDERTAKE THE PROPOSED RESEARCH?

There are some research methods that are needed to be undertaken for this study. In this section, they are examined in detail.

3.2.1. Research Methodologies

The most important elements of research methodology expected to be covered in my master thesis include research philosophy, research approach, research design, methods of data collection, sampling and ethical considerations. Let's look briefly at each of them separately.

3.2.1.1. Research Philosophy

Research philosophy is a vast topic and here we will not be discussing this topic in great details. In this thesis, I am not expected to discuss research philosophy in a great level of depth, and about one page in methodology chapter devoted to research philosophy usually suffices.

The research philosophy will reflect my important assumptions and these assumptions are base for the research strategy. Generally, research philosophy has many branches related to a wide range of disciplines. Within the scope of studies in particular there are two main contrasting research philosophies – positivism and phenomenology. The following table best illustrates the differences between these two approaches:

Table 3.1: Differences between Positivist and Phenomenology Paradigm

	Positivist Paradigm	Phenomenology Paradigm
Basic notions	The world is perceived as external and objective Independency of the observer Value-free approach to science	The world is perceived to be socially constructed and subjective Observer is considered a part of the object of observation Human interests drives science
Responsibilities of researcher	Focusing on facts Causalities and fundamental laws are searched Phenomenon are reduced to the simplest elements Hypotheses formulation and testing them	To be focusing on meanings Aiming to understand the meaning of events Exploring the totality of each individual case Ideas are developed by induction from data

Most suitable research methods	Concepts have to be operationalized	Using several methods in order to different aspects of phenomena
	Samples have to be large	Small samples are analyzed in a greater depth or over longer period of time

Resource: Positivism Research Philosophy

My research depends on positivism. However, my study also contains human interests and subjective ideas from the dialogs / debates as they are a part of phenomenology paradigm.

3.2.1.2. Research Approach

Within the methodology chapter of my thesis, I need to explain the main differences between inductive and deductive approaches and specify the approach I have adopted for my research by breaking down my arguments into several points.

Table 3.2: Comparison between Deductive and Inductive Methods

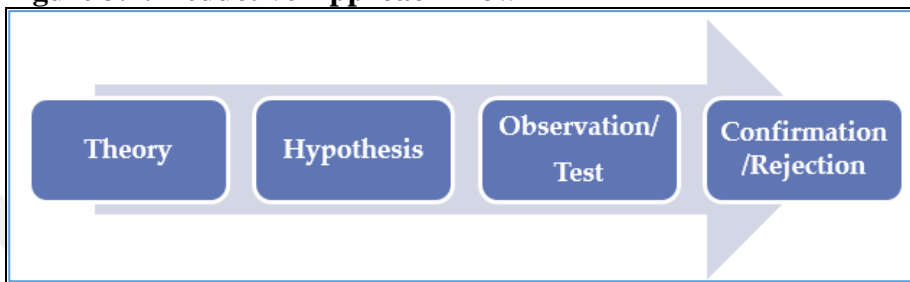
Deductive methods	Inductive methods
<ul style="list-style-type: none"> • Principles based on science • Movement is done from theory to data • Casual relationships between variables need to be explained • Quantitative type of data is mainly collected • Measures of control are applied in order to ensure the validity of data • Concepts are operationalized in order to ensure the clarity of definitions • The approach is highly structured • Researcher is independent from the research process • Samples need to be selected of a sufficient size in order to be able to generalize research conclusions 	<ul style="list-style-type: none"> • The meaning of human attachment to events are aimed to be explored • Research context is understood in a deeper manner • Qualitative type of data is collected • More flexible approach to research structure to ensure provisions for changes during the research • Researcher is perceived to be a part of the research process • Research findings do not have to be generalized

3.2.1.2.1. *Deductive Approach*

If you have formulated a set of hypotheses for your thesis that need to be confirmed or rejected during the research process you would be following deductive approach.

Dissertations (thesis) with deductive approach follow the following path:

Figure 3.1: Deductive Approach Flow



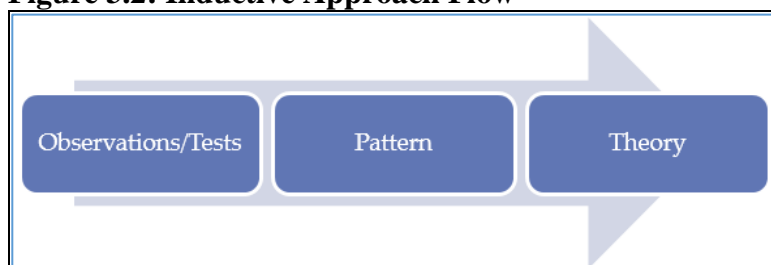
Resource: Deductive Approach

3.2.1.2.2. *Inductive Approach*

Alternatively, inductive approach does not involve formulation of hypotheses. It starts with research questions and aims and objectives that need to be achieved during the research process.

Inductive studies follow the route below:

Figure 3.2: Inductive Approach Flow



Resource: Inductive Approach

In my understanding, my study will use the inductive approach as I first get data and run my test cases to get an expected result. This pattern may eventually lead my study to create a theory that some cases can be generalized after some observations.

3.2.1.3. Research Design

The following table illustrates the main differences between exploratory and conclusive research in relation to important components of the thesis.

Table 3.3: Differences between Exploratory and Conclusive Research

Research project components	Exploratory research	Conclusive research
<i>Research purpose</i>	<ul style="list-style-type: none"> • General: to generate insights about a situation 	<ul style="list-style-type: none"> • Specific: to verify insights and aid in selecting a course of action
<i>Data needs</i>	<ul style="list-style-type: none"> • Vague 	<ul style="list-style-type: none"> • Clear
<i>Data sources</i>	<ul style="list-style-type: none"> • Ill defined 	<ul style="list-style-type: none"> • Well defined
<i>Data collection form</i>	<ul style="list-style-type: none"> • Open-ended, rough 	<ul style="list-style-type: none"> • Usually structured
<i>Sample</i>	<ul style="list-style-type: none"> • Relatively small; subjectively selected to maximize generalization of insights 	<ul style="list-style-type: none"> • Relatively large; objectively selected to permit generalization of findings
<i>Data collection</i>	<ul style="list-style-type: none"> • Flexible; no set procedure 	<ul style="list-style-type: none"> • Rigid; well-laid-out procedure
<i>Data analysis</i>	<ul style="list-style-type: none"> • Informal; typically non-quantitative 	<ul style="list-style-type: none"> • Formal; typically quantitative
<i>Inferences/recommendations</i>	<ul style="list-style-type: none"> • More tentative than final 	<ul style="list-style-type: none"> • More final than tentative

Resource: Research Design

As can be seen from the above table, my thesis study will mainly be a part of an exploratory research. That is because opinion mining studies are still in progress to have clearer results and there are only some methods to be certain to have some results such as words' sentiment points. As my study continues and I get results from my tests, the conclusion of the study may change depending on these.

3.2.1.4.Data Collection Methods

The following table presents the main differences between qualitative and quantitative research methods:

Table 3.4: Differences between Qualitative and Quantitative Research Methods

Characteristic	Quantitative research	Qualitative research
<i>Type of data</i>	<ul style="list-style-type: none">• Phenomena are described numerically	<ul style="list-style-type: none">• Phenomena are described in a narrative fashion
<i>Analysis</i>	<ul style="list-style-type: none">• Descriptive and inferential statistics	<ul style="list-style-type: none">• Identification of major schemes
<i>Scope of inquiry</i>	<ul style="list-style-type: none">• Specific questions or hypotheses	<ul style="list-style-type: none">• Broad, thematic concerns
<i>Primary advantage</i>	<ul style="list-style-type: none">• Large sample, statistical validity, accurately reflects the population	<ul style="list-style-type: none">• Rich, in-depth, narrative description of sample
<i>Primary disadvantage</i>	<ul style="list-style-type: none">• Superficial understanding of participants' thoughts and feelings	<ul style="list-style-type: none">• Small sample, not generalizable to the population at large

Resource: Research Methods

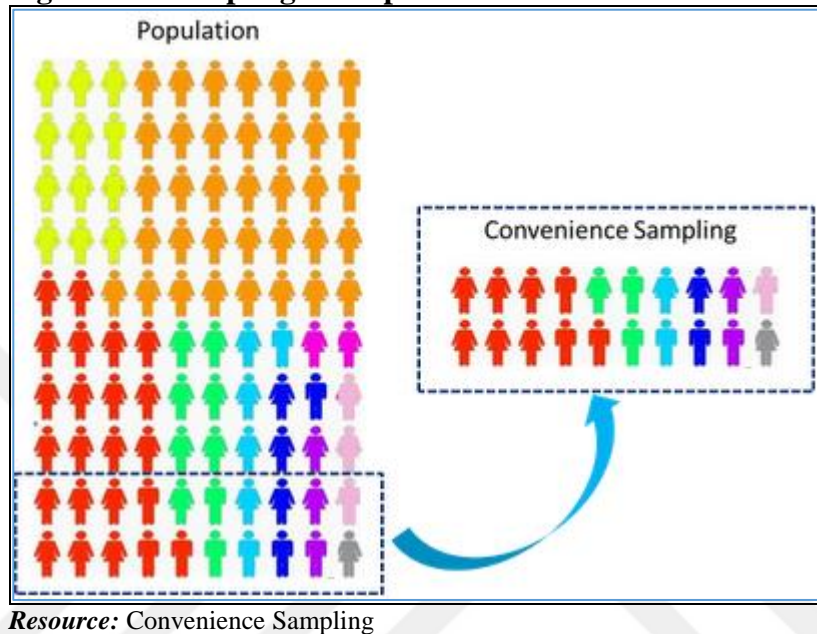
My thesis study will cover both quantitative and qualitative research approaches as I will use data from other sources and my own generated sources. My research results can be generalized because of words' sentiment points and also be specific to the language that I am working on.

Moreover, as Uçan indicated on his study, SentiWordNet English sentiment database is converted by considering Google Translate's function (2014). In this these, those words are converted into Turkish not by Google Translate's API but manually. However, it would be more convenient to use the API in order to make the software ready for future studies and works.

3.2.1.5. Sampling

There are two categories for this type of sampling that first one is random sampling and second one is non-random sampling.

Figure 3.3: Sampling Example



In my thesis, I will use non-probability sampling technique. The reasons of this are listed as below;

- i. This study is experimental that is why this technique is also useful.
- ii. If the data is huge, then this sampling technique can be used especially when the randomization is not possible.

To be more specific, I will use two methods of non-probability sampling technique is that they are convenience sampling and judgmental sampling.

3.2.1.5.1. Convenience Sampling

Convenience sampling is probably the most common of all sampling techniques. With convenience sampling, the samples are selected because they are accessible to the

researcher. Subjects are chosen simply because they are easy to recruit. This technique is considered easiest, cheapest and least time consuming.

3.2.1.5.2. *Judgmental Sampling*

Judgmental sampling is more commonly known as purposive sampling. In this type of sampling, subjects are chosen to be part of the sample with a specific purpose in mind. With judgmental sampling, the researcher believes that some subjects are more fit for the research compared to other individuals. This is the reason why they are purposively chosen as subjects.

Furthermore, in this study Hidden Markov Model is not applied but it can be very useful if this is also included in such software. As Blunsom indicated that the Hidden Markov Model is a most supported statistical tool for modelling a wide range of time series data. It has been exercised with big success to issues as part-of-speech tagging and noun-phrase chunking in the situation of natural language processing (2004).

3.2.1.6. Sentiment Classification Techniques

There are two main classification techniques that one is machine learning approach and the other is lexicon-based approach. Lexicon-based approach has three sub-approaches and they are dictionary-based approach, corpus-based approach and lexicon-based and natural language processing techniques. In this thesis, dictionary based-approach is applied (Medhat and Hassan and Korashy 2014).

3.2.1.6.1. *Dictionary-based approach*

According to the Mehdat, Hassan and Korashy's explanations, dictionary-based approach has a small set of opinion words which is collected manually. The newly found words are added to the seed list then next iteration starts. The iterative process stops when no new words are found. After the process is completed, manual inspection can be carried out to remove or correct errors (Medhat and Hassan and Korashy 2014, pp. 1102).

3.2.1.7. Ethical Considerations

Ethical issues are very important in researches. There are lots of subjects which every researcher should pay attention. The following subjects are the main and general principles and they show a guideline for people when they are working on their studies. These principles should be remembered all the time.

3.2.1.7.1. Honesty

Honesty is one of the most important rules in ethical considerations. Publication status, procedures and methods, results and data reports should be completed honestly. If researches play with data in order to show the results in the way of desire, then no one can be sure about that research paper's result. This can damage the trustworthiness against the researcher.

3.2.1.7.2. Objectivity

Objectivity is very important in order to show the results in a clear way because if researcher cannot be objective with the result, then researcher can mislead readers about the facts. Interpretation in data, analysis in data, in reviews, in decisions and etc. objectivity is needed for the sake of the research and future works of other people. Researches should pay attention in this subject in order not to make wrong effect to other researchers' both interests and quality of works.

3.2.1.7.3. Integrity

Integrity means that when a researcher says something in the paper that an example will be made, then that example should be given as promised before and readers and other researchers should not have any question marks on their minds like suspecting. This subject is also very related with honesty. Once this trust chain is broken between the researcher and the reader, then it is very difficult to repair it back.

3.2.1.7.4. Carefulness

This term can be explained with being clear in your work. A researcher should care the research paper and criticize the work when it is necessary. A researcher should also take notes about research activities like data collection and its design. This subject can also be taken as a habit for researchers' future work. It is very useful when this is made a habit.

3.2.1.7.5. Openness

Openness is like a door because when you use that for opening to others, then another door is opened for you by someone else. That is why a researcher should know what to share, why to share and when to share. Sharing resources, data, ideas, results are very important. Moreover, when a researcher opens all of these, it is normal to get criticism for what it is shared. This is also good because researchers can get benefit from it with new ideas.

3.2.1.7.6. Respect for Intellectual Property

Plagiarism is one of the bad ideas in researches. It is both dishonest and disrespectful against the people who actually own that work. That is why researchers should always pay attention to all forms of intellectual property. This can be patents, copyrights and etc. It is also not useful to use the data, method or result which are unpublished because all of these should be used with permission. Giving credit and acknowledgment are the best way to respect the intellectual property.

3.2.1.7.7. Confidentiality

Confidentiality means that if there is any personal or military secret data or records like communication records and etc., then these should be protected and should not be published in researches. This confidential data can belong to a patient or can be trade related. These records should also be protected and not published via the papers.

3.2.1.7.8. Responsible Publication

Responsible publication is another important subject in ethical considerations because when a researcher makes all the effort for career, then this is a responsible publication. The reason is that paper does not serve for future studies or scholarships.

3.2.1.7.9. Responsible Mentoring

Responsible mentoring means being a leader. What can be understood from this explanation is that when you make your research, it is good to mentor, advice and educate other students. This can also lead to a brainstorming and everybody can get benefit from it.

3.2.1.7.10. Respect for Colleagues

This subject reminds that every researcher should respect to other researcher because every study contains useful information and this information are used to do the World a better place. Respecting is the key for starting understanding each other's' works and efforts.

3.2.1.7.11. Social Responsibility

Researches can contain social messages for the good of public. These messages can be taken from another researchers and applied to our world. That is why when a study is undertaken, social responsibilities should not be forgotten.

3.2.1.7.12. Non-Discrimination

Definitely, discrimination is one of the most offensive behavior in our century and it is totally unacceptable. Because of this fact, a researcher should pay attention more when mentioning about any race, sex or ethnicity.

3.2.1.7.13. Competence

A researcher should take the study seriously. If the subject is his or her interest, then it should be remembered that competence is only possible with a lifelong education. This specialty can lead his or her to work with more professionals and this also leads to more knowledge and more success.

3.2.1.7.14. Legality

Researchers need to follow the governmental rules and laws in the country where they currently live in. It is their responsibility to know these rules and policies when they are conducting the researches.

3.2.1.7.15. Animal Care

If a researcher needs an animal in order to use it for the related study, then he or she must be very careful for not hurting it. Every researcher should show respect to animals and should not use them if the experimental tools are proper enough.

3.2.1.7.16. Human Subjects Protection

If a researcher wants to use humans for their research and experiments, then he or she needs to pay attention for not harming them. People's dignity, autonomy and privacy are always important. That is why a researcher must do his or her experiments or complete the study fairly.

3.3. WHAT ARE THE INTENDED TIMELINES TO UNDERTAKE THE PROPOSED RESEARCH?

Below figure represents the timeline of master thesis;

Figure 3.4: Master Thesis Timeline



3.4. WHAT RESOURCES, FACILITIES, EQUIPMENT, FIELDWORK WILL BE REQUIRED TO UNDERTAKE THE PROPOSED RESEARCH?

In this part, all the needs for this study are listed. They are as following; resources, facilities, equipment and fieldwork.

3.4.1. Resources

- i. Internet for research
- ii. Sample codes and functions
- iii. Python libraries
- iv. SentiWordNet - English Sentiment Dictionary
- v. Turkish Grammar Books and Dictionaries
- vi. Free or paid APIs such as Google Translate

3.4.2. Facilities

- i. There is none.

3.4.3. Equipment

- i. A hosting server to serve
- ii. A computer to code and implement
- iii. An IDE such as PyCharm

3.4.4. Fieldwork

- i. Study Turkish grammar



4. RESULTS

4.1. HOW WILL THE PROPOSED RESEARCH CONTRIBUTE TO THE RESEARCH OUTPUTS AND RESEARCH CULTURE IN THE RESEARCH AREA?

My thesis study aims to show that we as humans can solve all our conflicts among us. This can be done by formulating our languages. All languages have already their rules and we call them as grammars. These grammars can be formulized and calculated to understand what people say and mean in the context of a dialog. Most of the time, people have difficulties for explaining their opinions during a discussion. It is because when there is a dispute or conflict between two people or parties, not only the thoughts people try to explain but also the feelings they should suppress to be more clear against each other. Because of these feelings' interruptions, machines can help us to understand and transmit the ideas of people.

4.2. STRUCTURE OF THE SOFTWARE

The software is written in Python language. The Python version is 3.5 and MySQL is used for database purposes. Besides that, PyMySQL connector is used in order to connect the Python and MySQL together.

There are seven py files which are as following;

- i. analysis.py
- ii. db_operations.py
- iii. description.py
- iv. file_operations.py
- v. greeting.py
- vi. io_operations.py
- vii. main.py

There are also thirty-six txt files for testing purposes. Each name of it describes the structure of every first sentence in those text files. These files are categorized into 4 groups and inside them there are sub-groups.

First of all, there are 4 main categories and their sub-categories for these test texts and below is the list;

- i. Anlam İlişkilerine Göre Cümleler (Sentences in comparison with their meaning relationships)
 - a. Neden-Sonuç Cümleleri (Cause-Result Sentences)
 - b. Amaç-Sonuç Cümleleri (Aim-Result Sentences)
 - c. Koşul Cümlesi (Condition Sentence)
- ii. Anlamına Göre Cümle Çeşitleri (Sentence types in comparison with meanings)
 - a. Olumlu Cümle (Positive Sentence)
 - b. Olumsuz Cümle (Negative Sentence)
 - c. Soru Cümlesi (Question Sentence)
 - d. Ünlem Cümlesi (Exclamation Sentence)
- iii. Cümle Yorumlama (Sentence interpretation)
 - a. Eş veya Yakın Anamlı Cümleler (Synonym or Adjacent Sentences)
 - b. Karşıtlık Bildirenler (Opposition Sentences)
 - c. Birbiriyle Çelişenler (Conflicting Sentences)
 - d. Açıklama Bildirenler (Commented Sentences)
 - e. Olasılık-Tahmin Cümlesi (Possibility Sentences)
 - f. Öneri Cümleleri (Suggestion Sentences)
 - g. Tasarı Anlamı Taşıyanlar (Intention Sentences)
 - h. Varsayım Cümleleri (Assumption Sentences)
 - i. Ön Yargı Bildirenler (Biased Sentences)
 - j. Eleştiri Cümleleri (Criticism Sentences)
 - k. Aşamalı Durum Bildirenler (Gradual Case Sentences)
 - l. Gerçekleşmemiş Beklenti Bildirenler (Unrealized Expectation Sentences)
 - m. Yakınma-Sitem Anlamı (Complaint Sentences)
 - n. Hayıflanma (Regret Sentences)

- iv. Anlatımına Göre Cümleler (Sentences in comparison with expression)
 - a. İçerik Cümlesi (Contentful Sentences)
 - b. Üslup Cümlesi (Style Sentences)
 - c. Nesnel ve Öznel Anlatım (Objective and Subjective Expression)
 - d. Doğrudan ve Dolaylı Anlatım (Direct and Indirect Expression)
 - e. Tanımlama Cümlesi (Description Sentence)

There are two database tables which are always in interaction with Python. These table names are;

- i. allwords
- ii. turkce_kelimeler

In the next sections, more information about the above points are given.

4.3. MAIN FEATURES

There are two main features of the software and these two main features also have some sub-features in order to make the user experience high and more logical and useful.

Two main features:

- i. Search for the sentiment points of a Turkish word
 - a. View all the sentiment points of negative and positive
 - b. Add a new word to database if the word cannot be found
- ii. Analyze a given text's each sentence to see their sentence sentiment points
 - a. List the attendees' names which are got involved in the given text, discussion
 - b. Find the ngram points of each sentences
 - c. Return as a list for the given text that includes attendees'names, their sentences and sentences' sentiment points

- d. Add a new word to database if the word in the text cannot be found in the database

4.4. ANALYZING THE PY FILES

As mentioned above earlier, there are seven py files and all of their functions are explained below respectively.

There are as following;

- i. analysis.py
- ii. db_operations.py
- iii. description.py
- iv. file_operations.py
- v. greeting.py
- vi. io_operations.py
- vii. main.py

4.4.1. Analysis.py

This python executable is one of the important one which includes crucial functions. Most of the operations go through this executable file and reach other executable files to make the job done.

There are ten functions inside this py file and the below is a list of them;

- i. attendee_number()
- ii. list_sentences_of_moment()
- iii. get_newword_information_and_update()
- iv. check_the_word()
- v. print_sentiword_score()
- vi. return_sentiword_score()

- vii. `synset_score()`
- viii. `split_into_words()`
- ix. `give_score_to_words()`
- x. `sentisentence_score()`

4.4.1.1. Function 1: attendee_number

This function gets the file name as the parameter and prints the number of attendees in the given text and also returns the names of attendees. In the main.py file, there is a list which is defined in order to get the names of attendees. This list gets the names after this function is run in main.py by calling it from analysis.py module.

4.4.1.2. Function 2: list_sentences_of_moment

This is the function where the user gets the expected result as the attendee names, their sentences and sentences' sentiment points. The function gets the attendee list and the file name as the parameters. Attendee list comes from the above function which is the attendee_number. Attendee_number returns the attendee list in main.py and again in the main.py when list_sentences_of_moment is called; it takes the attendee list as a parameter along with the file name.

4.4.1.3. Function 3: get_newword_information_and_update

This function's size in the analysis.py is one the highest. It is normal because it has an important duty for the software. It takes the word from the user as its parameter. When a word cannot be found in the database, then this function comes up and ask the user to give necessary information to save it to the database. This information is the following;

- i. Is this word positive, negative or neutral to you?
- ii. What is the type of this word? Noun, pronoun, adjective, adverb, preposition, conjunction, interjection, verb.
- iii. Please give the meaning of this word.
- iv. Give a sentiment score for this word between 1 to 1000.

When the above information is provided, then this new word is saved into the database for both current and future usage. This function also returns the score of the word in the case of need for other functions.

4.4.1.4. Function 4: check_the_word

This is basic function that it checks whether the taken word is in the database or not. If the word cannot found in the database, then it returns “NOK” string as a result. If the word can be found in the database, then it returns “OK” string for the calling module.

4.4.1.5. Function 5: print_sentiword_score

This function prints the sentiment positive and negative scores of a word along with some other information like word id, word itself and word’s meaning. It takes the given word from the user input as a parameter and nothing returns as a result.

4.4.1.6. Function 6: return_sentiword_score

This function returns the sentiment positive and negative scores of a word along with some other information like word id, word itself and word’s meaning as a list. It takes the word from the user input as its parameter. The actual job is handled in db_operations.py file. This function only calls another function in db_operations.py and get the needed result.

4.4.1.7. Function 7: synset_score

When a word’s sentiment point is calculated, its positive and negative scores should be summed up. That is called as synset score. This function basically does this job. When a sentence’s sentiment score is found, the software goes through its words one by one and finds their positive and negative points. After that, the below mathematical operation is applied to find it:

$$\text{Synset Score} = \text{Positive Score} - \text{Negative Score}$$

This function gets two parameters and these are positive score and negative score. It returns the synset score.

4.4.1.8. Function 8: split_into_words

This function gets a sentence as a parameter and splits this sentence into their words. It writes these words into a list and returns this list as a result.

4.4.1.9. Function 9: give_score_to_words

This function also takes a sentence as a parameter. After that it calls `split_into_words` function in order to get it word by word list. It needs it like that because it calculates word's points. After it takes this word by word list, it searches these words in a loop that if a word cannot be found in the database by calling the function `check_the_word`, then it calls the `get_newword_information_and_update` function to create this new word in the database. If the word exists in the database, then it calls `print_sentiword_score` function which lists a word's id, itself, positive score, negative score and meaning. That is important because if there is more than one meaning for a word, then the software allows the user to pick the most related one for its case in order to create most accurate result.

After it completes all the necessary steps, it returns a list of words with their scores.

4.4.1.10. Function 10: sentisentence_score

This function takes a sentence as a parameter and calls the `split_into_words` function in order to get the word one by one in a returned list. After getting that list, it calculates the ngram score for that sentence by performing the below mathematical operation;

$$\text{Ngram} = \text{Number of Words} - (\text{Number of Grouping Words} - 1)$$

To make the above formula more clear, here is an example that this software uses in it;

Example sentence: Bu düşünceye katılmıyorum. (I don't agree with that.)

The software uses bigram structure which means that it takes the words two by two but in a different way. Here how it works when it is a bigram structure;

Bu düşünceye
düşünceye katılmıyorum

Another example sentence: Bu düşünceye kesinlikle katılmıyorum. (I absolutely don't agree with that.)

Bu düşünceye
düşünceye kesinlikle
kesinlikle katılmıyorum

For the first example sentence, formula works like this;

$$\text{Ngram} = 3 - (2 - 1)$$

$$\text{Ngram} = 2$$

As a result, for the first sentence there are 2 ngrams in the sentence.

For the second example;

$$\text{Ngram} = 4 - (2 - 1)$$

$$\text{Ngram} = 3$$

These results can also be seen when you count the ngram results one by one above.

After all, this function prints sentences' sentiment points and returns sum sentiment scores of the sentences as a result.

4.4.2. DB_operations.py

This python executable is another important module which includes vital functions. All of the operations are database related.

There are six functions inside this py file and the below is a list of them;

- i. insert_new_word()
- ii. find_word()
- iii. list_words_and_meanings_with_scores()
- iv. find_score()
- v. return_score()
- vi. db_close()

4.4.2.1. Function 1: insert_new_word

As it can be seen from its name, this function inserts new words into the database. This function is called from analysis.py module. It takes five parameters and these are; word type, word, positive / negative / neutral sign for the word, score and meaning. All of these parameter values are asked to the user during the program flow and when the program gets all of them it passes them starting from main.py to analysis.py and then to db_operations.py.

This function also holds the word type information in a list that when the user selects the type of the word, this function picks it from that list and saves in the related database tables.

Saving to database tables are operated at the same time. When a new word comes, the insert operation are done. One of the table is allwords which is the main table for searching the words. If a word cannot be found in this table, then new words are first inserted into turkce_kelimeler table, then into allwords table. The reason to insert into

two database tables is that the another aim of this software is to create our own Turkish sentiment dictionary based on people's scores for the words.

This function does not return any value but prints an error message if something goes wrong during saving the new record to the database.

4.4.2.2. Function 2: find_word

This function is called from analysis.py module and takes one parameter and that is a word. It returns two values in different two cases that one is "OK" string for successful operations and the other is "NOK" string for not successful operations. Successful operations are the ones that the searched words could be found in the allwords table. Not successful operations are the ones that the searched words could not be found in the allwords table.

4.4.2.3. Function 3: list_words_and_meanings_with_scores

This function takes three parameters and these are sql statement, word with a '#' character at the end of it and word itself.

Sql statement is passed because if there is a change need in the sql statement, then it will easy to change that from its own function. Fowrd is thw word with '#' character at the end of it. Program searches words with this character because one word can have many meanings and '#' character is the separator for that. For example; "hayır#1" has the meaning of "reddetme" and also "hayır#2" has the meaning of "hayır işi". That is why '#' is important. Lastly, word is itself and is used for mostly in print function. This function returns the words along with their meanings and scores.

4.4.2.4. Function 4: find_score

This function takes one parameter and that is a word. After that, it prints the following information; word id, word itself, positive score, negative score, word's meaning.

4.4.2.5. Function 5: return_score

This function takes one parameter and that is a word. After that, it returns a list which includes words, their meanings and scores.

4.4.2.6. Function 6: db_close

This function only closes the database connection after all the necessary operations are performed.

4.4.3. Description.py

This module has only one function in it and that's name is `what_this_software_does()`.

4.4.3.1. Function 1: what_this_software_does

As it can be understood from the function's name, this function describes what this program does for the users. When the `main.py` is run, one of the first modules that is get started is this `description.py` along with its this only function. This function has no parameters and only prints the main features of the program and their sub-features in short.

4.4.4. File_operations.py

This module has two functions in it and its job is to do file operations like viewing the content of a file and etc. The functions are as following;

- i. `list_files()`
- ii. `view_content()`

4.4.4.1. Function 1: list_files

This function has no parameters and only lists the files which are located on the project folder's "texts" directory.

4.4.4.2. Function 2: view_content

This function takes one parameter and that is a file name. When it gets a file name, then it prints all the content of that file.

4.4.5. Greeting.py

This function has only one function and that's name is say_hi().

4.4.5.1. Function 1: say_hi

This is the greeting function of this software. When the main.py is run, then this function gets enabled and prints its greeting message. It takes no parameters and does not return a value.

4.4.6. IO_operations.py

This function has only one function and that's name is print_to_screen().

4.4.6.1. Function 1: Print_to_screen

Actually it does the same functionality with Python's print function. It takes a string parameter and prints that string for the user. It does not return any value at the end.

4.4.7. Main.py

This is the first module that is run when the software is up and running. This module keeps working until the software is closed. It keeps asking for input in a while loop and there are two options for the user in this cycle; search sentiment score for a word or make a given text analyzed by calculation of sentence sentiment scores.

4.5. ANALYZING THE DATABASE STRUCTURE

MySQL database program is used for this project and there are two tables in the project's database.

These tables are as the following;

- i. allwords
- ii. turkce_kelimeler

4.5.1. Database Table 1: allwords

This table is created by translating the SentiWordNet's English sentiment dictionary.

This table eight columns and here is the list of them and their functions;

- i. **Wn:** This column's name stands for "word number". It's data type is integer. It has this following specifications; Primary key, Not null, Unique, Auto increment
- ii. **Pos:** This column stands for prefix. It's data type is varchar and keeps words' prefixes like if a word is an adjective, then this keeps 'a' character in it. This column has this following specifications; Not null
- iii. **Id:** This column is migrated from the SentiWordNet database. It's data type is varchar. It keeps the id for the words. In this project's database it is here only for reference. This column has this following specifications; Not null
- iv. **Pos_score:** This column is migrated from the SentiWordNet database. It's data type is double. It keeps positive scores of the words. This column has this following specifications; Not null
- v. **Neg_score:** This column is migrated from the SentiWordNet database. It's data type is double. It keeps negative scores of the words. This column has this following specifications; Not null
- vi. **Synset_terms_en:** This column is migrated from the SentiWordNet database. It's data type is varchar. It keeps the English words. This column has this following specifications; Not null

- vii. **Synset_terms_tr:** This column is created for this project's purpose and it keeps both the Turkish translations of above English words and newly created Turkish words. It's data type is varchar. This column has this following specifications; Not null
- viii. **Meaning:** This column is created for this project's purpose and it keeps the meanings of the Turkish words. It's data type is varchar. This column has this following specifications; Not null

4.5.2. Database Table 2: turkce_kelimeler

This table is created because there are lots of words that do not exist in "allwords" table which includes the translated words of SentiWordNet English words. So, this table will keep newly created Turkish words and this has also lots of advantages that the one of the main purposes of this table is that there will be a unique Turkish sentiment dictionary in the future.

This table six columns and here is the list of them and their functions;

- i. **Id:** This is the id column and its data type is integer. This column has this flowing specifications; Private key, Not null, Unique, Auto increment.
- ii. **Tur:** This is the type column. This keeps the types of words like adjective or word. The coumns's data type is varchar. This column has no specifications and can be null.
- iii. **Kelime:** This column keeps the Turkish words. Its data type is varchar. This column has this following specification; Not null, Unique.
- iv. **Pozitif_puan:** This column keeps the positive score of a word. Its data type is double. This column has this following specifications; Not null.
- v. **Negatif_puan:** This column keeps the negative score of a word. Its data type is double. This column has this following specifications; Not null.
- vi. **Anlami:** This column keeps the meaning of a word. Its data type is varchar. This column has this following specifications; Not null.

The software returns to its main menu after it completes the text analyze. User can continue with searching word's sentiment points or text analyze or quit the software.



5. FINDINGS

The main purpose of this study is to make contribution to sentiment analysis for Turkish language. When it comes to code the software, only the document analyzing was thought to complete. However, in the process of time it was understood that word searching functionality should be added to this software. In any way, word searching function was a part of document analysis but additionally separating them created better user experience.

Moreover, it is interesting that Marcus asked on his study that the way the sentiment analysis is performed and why a statement is characterized as positive, negative or neutral (2014). This characterization is important because otherwise we cannot specify the sentences' states and what people think about.

5.1. MEASUREMENT

In this thesis's findings, sensitivity and specificity measuring techniques are used (Solutions to Diagnostic Testing Practice Exercises 2016). In addition to that, ngram technique which was mentioned in the section of results is used in order to increase the accuracy of the calculations. Lastly, as the research approach, deductive approach was followed because it was the thought that the lack of applying all the grammar rules could cause some faulty results and that is proved at the end the tests.

There are some terms which are used in the below table. That is why, firstly those terms are explained;

- i. **Sensitivity:** Correctly receive positive test result
- ii. **Specificity:** Correctly receive negative test result
- iii. **False positive rate:** Incorrectly receive a positive test result
- iv. **False negative rate:** Incorrectly receive a negative test result
- v. **Positive predictive value:** Correctly receive positive test result and truly the sentence state is positive

- vi. **Negative predictive value:** Correctly receive negative test result and truly the sentence state is negative

Below table shows the calculated scores of sentences and those sentences' true states. These true states are decided by reading the sentences, so it is completed manually. Those calculated scores are the scores that the software gives.

The abbreviations' meanings in the table are as the following;

- i. **FSR:** First Sentence Result
- ii. **SSR:** Second Sentence Result
- iii. **FSTS:** First Sentence True State
- iv. **SSTS:** Second Sentence True State

Table 5.1: Calculated scores and true states of sentences in the files

ID	File Name	FSR	SSR	FSTS	SSTS
1	AnlamIliskilerineGoreCumleler-AmacSonucCumleleri-Ornek1.txt	0,0	0,0	Positive	Negative
2	AnlamIliskilerineGoreCumleler-KosulCumleleri-Ornek1.txt	-1,3	0,5	Positive	Negative
3	AnlamIliskilerineGoreCumleler-KosulCumleleri-Ornek2.txt	-0,3	-0,4	Positive	Negative
4	AnlamIliskilerineGoreCumleler-KosulCumleleri-Ornek3.txt	0,4	-0,2	Negative	Negative
5	AnlamIliskilerineGoreCumleler-NedenSonucCumleleri-Ornek1.txt	2,2	0,5	Positive	Negative
6	AnlaminaGoreCumleCesitleri-OlumluCumle-Ornek1.txt	4,2	0,0	Positive	Negative
7	AnlaminaGoreCumleCesitleri-OlumsuzCumle-Ornek1.txt	2,8	0,0	Positive	Negative
8	AnlaminaGoreCumleCesitleri-OlumsuzCumle-Ornek2.txt	2,6	0,7	Positive	Negative
9	AnlaminaGoreCumleCesitleri-OlumsuzCumle-Ornek3.txt	0,9	0,9	Negative	Positive
10	AnlaminaGoreCumleCesitleri-SoruCumlesi-Ornek1.txt	2,4	0,7	Negative	Negative

11	AnlamınaGoreCumleCesitleri-SoruCumlesi-Ornek2.txt	0,4	0,4	Positive	Positive
12	AnlamınaGoreCumleCesitleri-SoruCumlesi-Ornek3.txt	-0,6	1,0	Negative	Positive
13	AnlamınaGoreCumleCesitleri-SoruCumlesi-Ornek4.txt	-0,2	0,0	Negative	Positive
14	AnlamınaGoreCumleCesitleri-SoruCumlesi-Ornek5.txt	1,0	-0,04	Negative	Positive
15	AnlamınaGoreCumleCesitleri-SoruCumlesi-Ornek6.txt	1,3	1,02	Negative	Positive
16	AnlamınaGoreCumleCesitleri-UnlemCumlesi-Ornek1.txt	3,0	-0,07	Negative	Negative
17	AnlatımınaGoreCumleler-DogrudanAnlatımVeDolaylıAnlatım-Ornek1.txt	4,9	0,4	Positive	Negative
18	AnlatımınaGoreCumleler-IcerikCumlesi-Ornek1.txt	2,5	-1,0	Positive	Negative
19	AnlatımınaGoreCumleler-NesnelveOznelAnlatım-Nesnel-Ornek1.txt	2,8	0,6	Positive	Negative
20	AnlatımınaGoreCumleler-NesnelveOznelAnlatım-Oznel-Ornek1.txt	2,5	1,7	Positive	Negative
21	AnlatımınaGoreCumleler-TanımlamaCumlesi-Ornek1.txt	4,4	1,0	Positive	Negative
22	AnlatımınaGoreCumleler-UslupCumlesi-Ornek1.txt	3,2	-0,7	Negative	Positive
23	CumleYorumlama-AcıklamaBildirenCumleler-Ornek1.txt	2,8	-0,1	Positive	Negative
24	CumleYorumlama-AsamalıDurumBildirenler-Ornek1.txt	3,3	0,3	Negative	Negative
25	CumleYorumlama-BirbiriyleCelisenCumleler-Ornek1.txt	4,4	1,2	Negative	Positive
26	CumleYorumlama-ElestiriCumleleri-Ornek1.txt	5,3	0,6	Negative	Positive
27	CumleYorumlama-EsVeyaYakınAnlamlıCumleler-Ornek1.txt	-0,8	-0,3	Negative	Positive
28	CumleYorumlama-GercekleşmemişBeklentiBildirenler-Ornek1.txt	1,7	0,0	Negative	Negative

29	CumleYorumlama-Hayiflanma-Ornek1.txt	1,8	1,1	Negative	Positive
30	CumleYorumlama-KarsitlikBildirenCumleler-Ornek1.txt	1,8	-1,6	Positive	Negative
31	CumleYorumlama-OlasilikTahminCumlesi-Ornek1.txt	0,5	0,3	Positive	Negative
32	CumleYorumlama-OneriCumleleri-Ornek1.txt	0,9	0,9	Positive	Negative
33	CumleYorumlama-OnYargiBildirenler-Ornek1.txt	1,3	0,0	Negative	Positive
34	CumleYorumlama-TasariAnlamiTasiyanCumleler-Ornek1.txt	1,4	1,2	Positive	Negative
35	CumleYorumlama-VarsayimCumleleri-Ornek1.txt	1,4	1,2	Positive	Positive
36	CumleYorumlama-YakinmaSitemCumlesi-Ornek1.txt	1,3	2,2	Negative	Positive

Below table shows the comparison between test results which the software generates the statuses and true sentence states which are decided manually.

Table 5.2: Comparison between test results and true sentence states

True Sentence State	Test Result		Total
	Positive (+)	Negative (-)	
Positive (+)	33	8	41
Negative (-)	30	39	69
Total	63	47	110

Given the numbers on above table, results are shown below;

- i. **Sensitivity** is $33 / 41 = 0,8$ or 80 %
- ii. **Specificity** is $39 / 69 = 0,56$ or 56 %
- iii. **False positive rate** is $30 / 69 = 0,43$ or 43 %
- iv. **False negative rate** is $8 / 41 = 0,19$ or 19 %
- v. **Positive predictive value** is $33 / 63 = 0,52$ or 52 %
- vi. **Negative predictive value** is $39 / 47 = 0,82$ or 82 %

5.2. HIGHLIGHTS

In this part, some of the critical points are highlighted for readers. It is important to know them in order to have a better understanding about this study.

5.2.1. Comparison Between the Initial Hypothesis and the Current Results

Considering the initial hypothesis, the current results show that getting the most accurate and correct results are only possible when all grammatical rules are applied to the software. The code in the software acts by the scores of the words and if only the scores are correct, then sentences' sentiment score also become correct. That being said, it can be said that the initial hypothesis and the current results overlap because at the beginning possible outcomes were calculated.

As it is indicated in the study of Agarwal and Mittal, dictionary-based approaches utilize the pre-developed polarity lexicons like SentiWordNet, WordNeti, SenticNet, etc (2015). In this study, SentiWordNet is used and that is why it is a ready sentiment dictionary which was prepared in English. This dictionary is translated into Turkish and after this process, most of the words' sentiment points were known. Our dictionary-based approach was made ready after these processes.

6. DISCUSSION

6.1. THEORETICAL IMPLICATIONS

Text mining and sentiment analysis are huge subjects and very demanding nowadays. In this study, a part of them as a technique was tried to adapt to Turkish language's sentiment analysis. One of the main theoretical implication could be that especially for the Turkish websites which have comment areas for user to make their comments can be taken under examination and profile based results can be generated. According to Coleman and Freelon, classification studies do not go quite far enough to qualify as social science; their goal is typically to optimize algorithmic performance rather than to contribute to theory (2015). In their statement, they may not agree with the idea of optimization of this process but it is crucial to have this optimization in order to process more data. If this optimization process takes less than now it takes, then the second theoretical implication would be that all the interactions between both people to people and people to other devices can be taken and made different assumptions in order to find different needs of people.

6.2. PRACTICAL IMPLICATIONS

First of all, it should be admitted that this software has lots of missing features but the main purpose of creating such a master thesis is that there are very rare studies about Turkish sentiment analysis and there are not much open resources for curious students to examine them. This project should be accepted as an experimental project because without adding Turkish grammar rules to this software, nobody can be sure about the results, these sentences' sentiment scores. They can mean something but it works for the very limited version of sentences. When it comes to the complicated ones, the results could be very inconsistent. There are too many Turkish grammar rules. In this project, Turkish sentence grammar rule names are grouped in the list of given texts. This can be seen when the user selects "t" option in order to make a text analyze. That list is also a roadmap for the future work of this study. Turkish sentences' grammar rule names can

be applied to this project. If it can be, then more accurate and consistent result the software will generate.

Moreover, as Bhattacharjee and Fitzgerald specified on their study, sentiment analysis can be used for consumer products and gauging public reactions to political campaigns and etc. (2012). Therefore, like their explanations, in this master study's result also leads to many other researches that the software which was implemented for this study can be used and improved in that direction. Furthermore, there is another important practical implication that when users are asked to save a new word into the database, the software asks some questions to users. Those questions are necessary in order to from the base of the related word. For example, one of the questions there is the meaning of the word. This question is important because then users really starts to think about what they are looking for. When this happens, users start being more careful and types the word's meaning more carefully. This causes more accurate results in next iterations of the software because another user will look at those meanings and decide accordingly. As Stuckenschmidt and Jannach indicate on their study, reviews affect other people's thoughts and help them to decide correctly like in our situation (2015).

7. CONCLUSION

This study's aim is to show that text mining and sentiment analysis techniques can be applied to Turkish language and by doing that positive results can be generated. It is expected that some positive results were obtained. In the software which is coded for this study purpose there are some pre-defined documents. Those documents have Turkish sentences with Turkish sentence rules. Every document has its own specific rule in it and every first sentence in those documents carries that rules. Every second sentences there are the answers for the first sentences.

Results of the study is given in the part of results in this paper. Moreover, in the part of findings, measurement techniques and some highlights are shared. Furthermore, in the part of discussion, both theoretical and practical implications are given to readers. All the results show that when word's positive and negative scores are correctly types, then the sentences' sentiment points are calculated correctly. The important point here is that all of the words' sentiment points should be found beforehand and made ready for giving sentences. The words' sentiment points should not be based on one or two person's judgements. Therefore, the number of participants who will complete the missing words and their sentiment points should be increased as much as possible.

Lastly, text analysis is one of the most popular subject in computer science nowadays. In some languages like English, there are very good software programs with very accurate, consistent and correct results. However, this field could not be applied to Turkish language much enough. There is not much study like there are in English. One of the main purposes of this project is to increase the momentum of current studies of related topics in this area and to attract attention for this area from curious students. It is the hope that this study can be helpful or inspirational for other students.

REFERENCES

Books

- Ahmad, K., (2011). *Affective computing and sentiment analysis: emotions, metaphor and terminology*. Springer Science & Business Media. 45, pp. viii.
- Ertav, B., (2014). *Türkçe dilbilgisi notları: ygs – lys'ler için ders notları*. pp 20.
- Balijepalli, S., (2008). *Information retrieval based on sentiment analysis*. VDM Publishing.
- Agarwal, B., Mittal, N., (2015). *Prominent feature extraction for sentiment analysis*. Springer, pp. 13.
- Gökalp, M., (1976). *Türk dili bilgisi*. Şark Press, pp 157.
- Kagan, V., and Rossini, E., and Sapounas, D., (2013). *Sentiment analysis for ptsd signals*. Springer Science & Business Media, pp. 23.
- Akdeniz, S., (1973). *Türkçede yeni gelişmeler*. Özek Press, pp 56.
- Tkalčić, M., and De Carolis, B., and De Gemmis, M., and Odić, A., and Košir, A., (2016). *Emotions and personality in personalized services: models, evaluation and applications*. Springer, pp. 134.
- Gülsevin, G., and Boz E., (2006). *Türkçenin çağdaş sorunları*. Divan Press, pp 78.
- Liu, B., (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers, pp 22.
- Stuckenschmidt, H., and Jannach, D., (2015). *E-commerce and web technologies: 16th international conference on electronic commerce and web Technologies, ec-web 2015*. Valencia: Springer. 239, pp. 109.
- Liu, B., (2015). *Sentiment analysis: mining opinions, sentiments, and emotions*. Cambridge University Press.
- Pang, B., and Lee, L., (2008). *Opinion mining and sentiment analysis*. Now Publishers Inc, pp 26.
- Coleman, S., and Freelon, D., (2015). *Handbook of digital politics*. Edward Elgar Publishing, pp. 465.
- Özaydın, Ü., (2015). *Temiz türkçe kılavuzu: iletişim güzeldir*. Nubuto Limited Press, pp 29.

- Mandal, D., and Kar, R., and Das, S., and Panigrahi, B. K., (2015). *Intelligent computing and applications: proceedings of the international conference on ica*. Springer. 343, pp. 368.
- Bhattacharjee, A., and Fitzgerald, B., (2012). *Shaping the future of ict research: methods and approaches: ifip wg 8.2 working conference*. Tampa: Springer. 13-14, pp. 3.
- Management Association, Information Resources. 2013. *Enterprise resource planning: concepts, methodologies, tools, and applications: concepts, methodologies, tools, and applications*. IGI Global, pp. 1505.
- Göker, O., (1997). *Uygulamalı türkçe bilgileri vol3*. Ankara: The Ministry of Education, pp 95.
- Pierson, L., (2015). *Data science for dummies*. John Wiley & Sons, pp. 277.
- Sobaci, M. Z., (2015). *Social media and local governments: theory and practice*. Springer, pp. 121.
- Kutlu, M., (1977). *Türk dili ve edebiyatı ansiklopedisi: devirler, isimler, eserler, terimler, vol7*. Dergah Press, pp 119.
- Marcus, A., (2014). *Design, User Experience, and Usability: User Experience Design for Diverse Interaction Platforms and Environments: Third International Conference, DUXU 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceeding*. Springer. 2, pp. 487.

Periodicals

- Ishikiriya, C. S., and Miro, D., and Gomes, C. F. S., (2015). *Text Mining Business Intelligence: a small sample of what words can say*. *Procedia Computer Science*. 55, pp. 261 – 267.
- Medhat, W., and Hassan, A., and Korashy, H., (2014). *Sentiment analysis algorithms and applications: A survey*. *Ain Shams Engineering Journal*. 5, pp. 1093 – 1113.
- Singh, P. K., and Husain, M. S., (2014). *Methodological Study of Opinion Mining and Sentiment Analysis Techniques*. *International Journal on Soft Computing (IJSC)*. 5 (1).
- Cambria, E., and Schuller, B., and Xia, Y., and Havasi, C., (2013). *New Avenues in Opinion Mining and Sentiment Analysis*. *IEEE Computer Society*. March / April.
- Pang, B., and Lee, L., (2008). *Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval*. 2 (1-2), pp. 4 - 5.
- Istanbul University Faculty of Literature. 2006. *Türk dili ve edebiyatı dergisi*. Istanbul: Istanbul University Faculty of Literature Press. 34, pp. 44.
- Türk Dil Kurumu. 2006. *Türk dili araştırmaları yillığı 1. bölüm*. Turkish Language Association, pp. 80.
- Balijepalli, S., (2007). *Blogvox2: a modular domain independent sentiment analysis system*. Thesis for M.S. Degree. Maryland: University of Maryland.

Other Sources

- Demirel, Ö., (2009). *Yapıca Olumlu Anlamca Olumsuz Cümleler Hakkında Bir Değerlendirme*. A research about positive and negative sentences. Istanbul: Istanbul University.
- Demirel, Ö., (2002). *Türkiye Türkçesinde Olumsuz Cümle Yapıları*. Thesis for M.A. Degree. Istanbul: Inonu University.
- İlhan, N., (2007). *Türkçede Olumsuzluk*. A research about the negative sentences in Turkish. Elazığ: Fırat University.
- Uçan, A., (2014). *Automatic Sentiment Dictionary Translation and Using in Sentiment Analysis*. Thesis for M.A. Degree. Ankara: Hacettepe University.
- Blunsom, P. (2004). *Hidden Markov Models*. Melbourne: The University of Melbourne.
- Steels, L., and Kaplan, F., and McIntyre, A., and Van Looveren, J., (2000). *Crucial factors in the origins of word-meaning*. Paris: Sony Computer Science Laboratory.
- Learn Turkish with Manisa Turkish. 2016. *Basic Grammar*. [internet]. Istanbul: SIGN (Published 2016) <http://www.turkishlanguage.co.uk/> [accessed 27 August 2016].
- SentiWordNet database. <http://sentiwordnet.isti.cnr.it/> . [23 July 2016].
- Solutions to Diagnostic Testing Practice Exercises. 2016. <http://sites.stat.psu.edu/~lsimon/stat250/sp00/solutions/misc/diagttests.htm> [28 August 2016]
- Positivism Research Philosophy. 2016. <http://research-methodology.net/research-philosophy/positivism/> [31 August 2016]
- Deductive Approach. 2016. <http://research-methodology.net/research-methodology/research-approach/deductive-approach-2/> [31 August 2016]
- Inductive Approach. 2016. <http://research-methodology.net/research-methodology/research-approach/inductive-approach-2/> [31 August 2016]
- Research Design. 2016. <http://research-methodology.net/research-methodology/research-design/> [31 August 2016]
- Research Methods. 2016. <http://research-methodology.net/research-methods/> [31 August 2016]

Convenience Sampling. 2016. <http://research-methodology.net/sampling-in-primary-data-collection/convenience-sampling/> [31 August 2016]

