**CANKAYA UNIVERSITY**

**FACULTY OF ENGINEERING**

**COMPUTER ENGINEERING DEPARMENT**

**Sentiment Analysis on Turkish Sentences**

**LITERATURE REPORT**

**CENG 407**

**Prepared by**

Naz BAYRAKTAR

201111502

MERVE TEMÜR

201111050

**Table of Content**

Abstract …....................................................................................1

Introduction..................................................................................2

Social Media….............................................................................2

What is Twitter? ….......................................................................2

What is sentiment analysis? ….....................................................2

Sentiment Analysis in Turkish.......................................................3

Conclusion.....................................................................................5

References….................................................................................6

**ABSTRACT**

Sentiment analysis has been an active research area for a much time. But unfortunately, the most of works are specific to the English language. In this paper, we focus on Turkish language which is an agglutinative languages that makes the sentiment analaysis a little more complicated. In this work we will work on Turkish texts especially on informal text that written by the Twitter users as a tweets on Twitter, we will get tweets and location information throughout Twitter. İn this project we will trying to determine the attitude of the Twitter user with respect to a specific topic. The attitude may be like their evaluation, judgment, their feelings and the intended emotional communication. English sources will be used for the project because Turkish sources are limited.The vital aim of this project is to develop a new algorithm to do sentiment analysis on Turkish sentences

and present it in a Windows Form Application to sentiment analaysis methods on Turkish social media (Twitter).

Keywords: Sentiment Analysis, Turkish.

**ÖZ**

Sentiment analysis (düşünce çözümlemesi) uzun zaman önce aktif bir araştırma alanı olmuştur, ama mallesef çalışmaların çoğu inglizce diline özeldir. Bu çalışmada, Türkce dili üzerinde çalışma yapacağız. Bu dil eklemeli (bitişik) bir dil olduğundan dolayı, sentiment analysis biraz daha zor olacaktır. Burada ilgilendiğimiz kısım daha çok Twitter kulanıcıların resmi olarak yazılmayan tweetler olacaktır, tweetleri ve tweet konumlarını Twitter üzerinden erişeceğiz. Türkçede duygu analizi yapabimek için sosyal medyada kullanılan kısaltmalar ve hatalı ifadeler düzeltilmeli ve doğrusu oluşturulmalıdır.Mesajı anlayabilmek için cümleler en küçük birimlerine ayrılmalı, kelime türleri bulunmalıdır.Kelimenin ekleri, kökleri tespit edilmelidir.Çıkartılmaya çalışılan görüş¸ twitter kullanıcısının konu hakkındaki kararı ya da değerlendirmesi, yazarken hissettiği ruh hali, ya da belirtmek istediği etki de olabilir. Türkçe metinler için duygu analizi yapabilmek şu an üzerinde çalışılan bir konudur.Bizim projemizin amacı da İngilizce kaynaklardan yararlanarak duygu analizi için Türkçe bir algoritma tasarlamaktır.Aynı zamanda düşünce çözümlemesi için Türkce olarak yeni bir algoritma oluşturmaktır.

Anahtar kelimeler: Düşünce çözümleme, Türkçe.

**INTRODUCTION**

Social medya has becomed essential part of our lives, we use it in various situations like communication and transmission. There more than millions of users all over the world that uses social medya every single day. For instance Twitter, thousands of tweets are tweeted evey hour. We are planing to code a program and design a new feature that can be used in Twitter. Liu [12], this program analyses the tweets by using customized utilities and classifies them into positive, negative and neutral tweets by using some helper modules which can separate the Turkish sentenses according to the root, negativity and tenses of the word.

**What is Social Medya?**

Social media is the collective of online communications channels dedicated to community. Websites and applications dedicated to forums, microblogging, social networking, social bookmarking, and wikis are among the different types of social media. There are several types of social medya like Facebook, Google+, Wikipedia, Linkedln and Twitter. Social media is becoming an vital part of people lifes, specially Twitter which is a free microblogging service that allows registered members to broadcast short posts called tweets.

**What is Twitter?**

Twitter is a famous social network in the world.It is an application for people to share their thoughts and feelings online.People instantly,can be found in sharing, what they do, what they feel like and where they are.They share their thoughts and feelings with 140-character messages.These messages that are 140-character called tweet.Thanks to twitter,can access a lot of people's tweets with follow button.Additionally,picture and video can be added into tweet and can share the other followers.Users can use the hashtag in the message.Thanks to most used hashtag is created trend topic.Trending topics are those topics which discussed between peoples.So,people can follow the agenda.Through twitter,many people can communicate easily with each other who recognize or not recognize each other.

**What is sentiment analysis?**

Sentiment analysis is the process of detecting the content of text. In other words, it determines whether a piece of writing is positive, negative or neutral.Yi et al.[1] says sentiment about a subject is the polarity of the idea.Polarity means positive(+) or negative(-) of idea. Pang[2] claims sentiment polarity classification uses for reviews in the context of text.If there is like in this context polarity is “positive”,unlike this dislike “negative” opinions.For a lot institutions or trademarks important a text is positive or negative .For instance,a trademark can detect customers' thoughts through thrown tweets using a hashtag. In the past, quality-based studies that were done using methods as wish and complaint boxes are inadequate for institutions and trademarks nowadays.With evolving technology people's feelings and thoughts can be accessed via social media like twitter. Bruns et al.[3] tells in article twitter offers two application programming interfaces (APIs) for collecting tweets.These applications are the search API and the streaming API.Thanks to search API ,is received past tweets of specific user or can searched for the tweets according to a hashtag or by a word. In the other applications, researchers don’t need to define any specific criteria to receive data from the streaming API. So, to understand these thoughts and feelings an automatic social media analyzer is needed. Sentiment analysis is a method that will meet these needs. There are many studies about sentiment analysis. This studies aims to explore the technological opportunities (information extraction tools) of social media. Burnap et al.[4], in study entitled detecting tension in online communities with computational analysis twitter includes the following information. Content of tweets is important for data mining tools used in sentiment analysis.SentiStrength is a sentiment analysis tool.After receiving the text content from social media it test and evaluation this text.This tool can use for academic research. Mostafa [5] consider the biggest obstacle in front of the studies is related to the language. For example,to make social media analysis abbreviations and improper expression used should be corrected. Go et al.[6] consider misspellings because always text of a tweet are not correct.Sometimes can be seen in the following tweet: @BATMANNN :( i love chutney.... If is not the emoticon, tweet to be positive. Thus, processing of language successfully is very important. A natural language processing (NLP) application uses to identify text sentiment, typically as positive, neutral or negative. This technique is called text mining literature as emotional. Text mining literature that polarity analysis (EPA) have three levels.These levels are opinion mining, review mining, or appraisal extraction (Zagal, Tomuro, & Shepitsen, 2012).To calculate a sentiment score, can do words dictionary and this words have certain points to indicate positive, negative or neutral. Thelwall et al [7] claims to calculate score, is analyzed individually words in a sentence.Then,is determined a positive and a negative sentiment score for each word in the sentences.If words is positive scores range from +1 (neutral) to +5 (extremely positive).If it is negative scores range from -1 (neutral) to -5 (extremely negative). Gelenbe et al.[8] explain different word lists were created to give the word to the score.These lists are sentimental word list,booster word list,idiom list,negation word list and emoticon list.Human editors are compiled the sentimental words and their sentiment scores.Booster word list contains words that strengthen or weaken the sentiment as adjectives or adverbs.For example "good" have +2 points but "extremely good" have +3 points.There are some common phrases in idiom list . For instance, "how are you" has a sentiment score of +2, instead of a neutral score of +1.Negation word list contains negative words. Emoticon list includes common emoticons, like ":)" has a score of +2.

**Sentiment Analysis In Turkish**

The motivation behind creating a sentiment analysis framework specic to Turkish,

rather than using an existing framework for English, is because of certain differences

between Turkish and English languages. Eroğul [10]handled the sentimental analysis problem as a supervised machine learning classifcation problem and applied different ML techniques with different features like unigrams, bigrams, POS tags and combination of them. In English, negation is handled with the adverb ”not” but Turkish is an agglutinative language, there are a lot of combinations using the same root with different suffixes in Turkish also, long words can be created by adding many suffixes to a root word. Vural [11] in Turkish, most of the negation is done by the negation suffix in the predicate, which usually affects the whole sentence The added suffixes may change the polarity of words, the negation occurs in different ways in Turkish, Türkmenoğlu and Tantuğ [14] first, negation adding (“-ma -me”) to the word will change the polarity of that word, second, using the negation word (“Yok”, “değil”). Moreover, Turkish has several letters that are missing in English (“ç”, “ı”, ”ğ”, “ş”, “ü”,”ö”) and, in informal writing on the Web especially on Twitter, people tend to substitute these Turkish letters with the closest ASCII English letters (“c”, “i”, “g”, “s”, “u”, “o”). This creates complication in identifying the words. We decided to creat a sentiment analysis framework taking into account the above mentioned differences. Dehkharghani et al.[9], Our framework consists of a pipeline of several software modules, each providing some input to the succeeding module in the pipeline. The input to the framework is a piece of text written in Turkish and the output is a prediction about the polarity of the sentiments in the text, all a positive, a negative and a neutral class prediction. There are some modules that helps to correct the sentinces of the input text.

Sentence extractor: This is a simple module which splits the input text into sentences based on certain sentence separators (“.!?”). Each sentence is then passed to the next module as a separate input.

ASCII character converter: Each word in the input sentence is looked up in a dictionary and checked for spelling errors. If a corresponding term is not found in the dictionary or there is a spelling error, the term is passed as input to an ASCII-tolerant parser to see if the word is written using ASCII character substitution. At this step, the parser may rewrite the term by substituting certain characters for example “uzum” becomes “üzüm”.

Morphological analyzer: This performs morphological analysis on the words in the sentence. By the help of some customized dictionary, it basically finds all possible root forms and suffixes of a given word. After the morphological analysis,certain suffixes are seperated from the original word form. This is because some suffxes tense and person suffixes are not valuable for sentiment analysis.

Negation handler: The negation takes places in Turkish most often in two forms, either in the form of a separate word negating one of the preceding words for example, “guzel degil” or in the form of the (“-ma, -me”) suffix, which is a part of the negated word (“Yapmayacak”). To handle the negations of the first form, To handle the second form of negations, we modify the sentence and introduce an artificial keyword before the negated word. This artificial word is added to the negation word list.

Polarity predictor: Dehkharghani et al.,This module takes as input the sentiment scores associated with each word in the initial input text as well as the information about sentence splitting. The polarity of the input piece of text is determined according to the sentiment score assigned to the text.[9]

As a result, turning the ASCII conversion module on seems to help more in case of positive reviews.

**CONCLUSION**

In this work, we proposed some strategy that analyzes the emotions of the sentences which were written in Turkish language whether its happy, sad or it does not show anything. The system distinguished the orals by using customized analyze tools with the help of modules as we explained in the earlier beginnings. We hope to achieve success and see this actually works in Twitter and we are planning to expand this for all the other social media webs like Instagram in the future.

**REFERENCE**

[1]. J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack. Sentiment analyzer: extracting

sentiments about a given topic using natural language processing techniques. In

Proc. 3rd IEEE Int'l Conf. Data Mining, pages 427-434, 2003.

[2]. B. Pang and L. Lee. Opinion mining and sentiment analysis. Found. Trends Inf.

Retr., 2:1-135, 2008.

[3]. A. Bruns and Y. Liang, 'Tools and methods for capturing Twitter data during natural disasters', *First Monday*, vol. 17, no. 4, 2012.

[4].P. Burnap, O. Rana, N. Avis, M. Williams, W. Housley, A. Edwards, J. Morgan and L. Sloan, 'Detecting tension in online communities with computational Twitter analysis', *Technological Forecasting and Social Change*, vol. 95, pp. 96-108, 2015.

[5]. M. Mostafa, 'More than words: Social networks’ text mining for consumer brand sentiments', *Expert Systems with Applications*, vol. 40, no. 10, pp. 4241-4251, 2013.

[6]. Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision.Technical report, Stanford.

[7]. M. Thelwall, K. Buckley, G. Paltoglou, D. Cai and A. Kappas, 'Sentiment strength detection in short informal text', *J. Am. Soc. Inf. Sci.*, vol. 61, no. 12, pp. 2544-2558, 2010.

[8]E. Gelenbe and R. Lent, *Computer and information sciences III*. London: Springer, 2013.

[9]R. Dehkharghani, Y. Saygin, B. Yanikoglu and K. Oflazer, 'SentiTurkNet: a Turkish polarity lexicon for sentiment analysis', Lang Resources & Evaluation, 2015.

[10] U. Eroğul. ‘Sentiment analysis in turkish’. Middle East Technical University,

Ms Thesis, Computer Engineering, 2009.

[11] A. Vural, B. Cambazoglu, P. Senkul, and Z.Tokgoz. ‘A framework

for sentiment analysis in turkish: Application to polarity detection of movie reviews in turkish’. In Computer and Information Sciences III. Springer, 2013.

[12] B. Liu, Sentiment analysis and opinion mining. San Rafael, Calif.: Morgan & Claypool, 2012.

[13] E. Cambria, B. Schuller, Y. Xia and C. Havasi, 'New Avenues in Opinion Mining and Sentiment Analysis', IEEE Intell. Syst., vol. 28, no. 2, pp. 15-21, 2013.

[14] C. Türkmenoğlu and A. Tantuğ. Sentiment Analysis in Turkish Media. *Istanbul Technical Universty,* 2014.

[15] S. Seker and K. Al-NAAMI. Sentimental Analysis on Turkish Blogs via Ensemble Classifier. *Istanbul Medeniyet University.* 2013