AUTOMATIC DETECTION OF CYBER SECURITY EVENTS FROM TURKISH TWITTER STREAM AND TURKISH NEWSPAPER DATA

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF INFORMATICS

OF

THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

ÖZGÜR URAL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR

THE DEGREE OF MASTER OF SCIENCE

IN

CYBER SECURITY

MAY 2019

Approval of the thesis:

**AUTOMATIC DETECTION OF CYBER SECURITY EVENTS FROM TURKISH TWITTER STREAM AND TURKISH NEWSPAPER DATA**

Submitted by **ÖZGÜR URAL** in partial fulfillment of the requirements for the degree of **Master of Science in Cyber Security Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin

Dean, **Graduate School of Informatics**

Prof. Dr. Aysu Betin Can

Head of Department, **Cyber Security**

Assist. Prof. Dr. Cengiz Acartürk

Supervisor, **Cognitive Science Dept., METU**

Assoc. Prof. Dr. XXX

Co-Supervisor, **Computer Engineering Dept., METU**

**Examining Committee Members:**

Prof. Dr. XXX (\*)

Computer Engineering Dept., METU

Prof. Dr. XXX (\*\*)

Medical Informatics Dept., METU

Prof. Dr. XXX

Cognitive Science Dept., METU

Prof. Dr. XXX

Computer Engineering Dept., Boğaziçi University

Prof. Dr. XXX

Computer Engineering Dept., Bilkent University

**Date: \_***Write your defense date!*

*\*Write the name of the head of the examining committee in the first row.*

*\*\*Write ten name of the supervisor in the second row.*

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

**Name, Last name : ÖZGÜR URAL**

**Signature :**

# ABSTRACT

AUTOMATIC DETECTION OF CYBER SECURITY EVENTS FROM TURKISH TWITTER STREAM AND TURKISH NEWSPAPER DATA

Ural, Özgür

MSc., Department of Cyber Security

Supervisor: Assist. Prof. Dr. Cengiz Acartürk

May 2019, XX pages

Internet has many different kind of timely information. Every day, security experts scan the internet and face with lots of security events that effect people, institutions and governments. An information analyst constantly scan many sources to staying up to date on security events require, which leads to information overload to them. For example, an information analyst might want to be aware of cyber security incidents such as a DDOS attack on a government agency website. The earlier they are able to detect and understand the threats, the longer time remaining to alleviate the obstacle and to investigate them forensic. Therefore, they need to have a situation awareness of the existing security events and their possible effects. However, due to the large number of events, it can be difficult for security analysts and researchers to handle this flow of information in an adequate manner. Moreover, internet uses different languages to share informations, like the humans do. As expected, security events related informations which effect Turkish people, Turkish institutions and Turkish government also share in internet mostly in Turkish language. In this thesis, I investigate detection of existing security incidents with using Turkish language, ITU NLP Api, Twitter social network and Hürriyet Turkish newspaper. I propose an automatic, Turkish language specific software system that can detect cyber security events in real time over the Twitter stream in Turkish language and Hurriyet newspaper stream.

Keywords: Cyber Security, Event Detection, Turkish, Twitter, Hurriyet Newspaper.

# ÖZ\_\_

TÜRKÇE TWITTER AKIŞI VE TÜRKÇE GAZETE VERİLERDEN SİBER GÜVENLİK OLAYLARININ OTOMATİK TESPİT EDİLMESİ

Ural, Özgür

Yüksek Lisans, Siber Güvenlik Bölümü

Tez Yöneticisi: Yrd. Doç. Dr. Cengiz Acartürk

Mayıs 2019, XX sayfa

İnternet birçok farkı zamana bağlı bilgi içerir. Güvenlik uzmanları her gün interneti tarıyor ve insanları, kurumları ve hükümetleri etkileyen birçok güvenlik olayıyla karşı karşıya kalıyorlar. Bir bilgi analisti, gerekli olan güvenlik olayları hakkında güncel bilgi sahibi olmak için sürekli olarak birçok kaynak taramakta ve bu da onlarda aşırı bilgi yüklenmesine neden olmaktadır. Örneğin, bir bilgi analisti, bir devlet kurumu web sitesine yapılan DDOS saldırısı gibi siber güvenlik olaylarının farkında olmak isteyebilir. Tehditleri ne kadar erken saptarsa ​​ve anlarlarsa, problemleri hafifletmek ve adli olarak soruşturmak için o kadar uzun süreye sahip olurlar. Bu nedenle, mevcut güvenlik olayları ve olası etkileri hakkında durum bilgisine sahip olmaları gerekir. Ancak, çok sayıda olay nedeniyle, güvenlik analistlerinin ve araştırmacıların bu bilgi akışını yeterli şekilde ele alması zor olabilir. Ayrıca internet, insanlar gibi bilgileri paylaşmak için farklı dilleri kullanır. Beklenildiği gibi, Türk halkını, Türk kurumlarını ve Türk hükümetini etkileyen güvenlik olayları ile ilgili bilgiler de çoğunlukla Türkçe olarak internette paylaşılmaktadır. Bu tezde, mevcut güvenlik olaylarının tespitini Türk dilini, İTÜ NLP Api, Twitter sosyal ağı ve Hürriyet gazetesini kullanarak araştırıyorum. Türk dilindeki Twitter akışında ve Hürriyet gazetesi akışında siber güvenlik olaylarını gerçek zamanlı olarak algılayabilen otomatik, Türkçe’ye özgü bir yazılım sistemi öneriyorum.

Anahtar Sözcükler: Siber Güvenlik, Olay Tespiti, Türkçe, Twitter, Hürriyet Gazetesi.

# DEDICATION

To My Family

# ACKNOWLEDGMENTS

First of all, I would like to express …..

Besides my supervisor, I would like to thank …..

I would also like to thank all of colleagues from …..

To my wife, …..

# TABLE OF CONTENTS

[ABSTRACT iv](#_Toc6532464)

[ÖZ\_\_ vi](#_Toc6532465)

[DEDICATION viii](#_Toc6532466)

[ACKNOWLEDGMENTS ix](#_Toc6532467)

[TABLE OF CONTENTS x](#_Toc6532468)

[LIST OF TABLES xii](#_Toc6532469)

[LIST OF FIGURES xiii](#_Toc6532470)

[LIST OF ABBREVIATIONS xiv](#_Toc6532471)

[1 INTRODUCTION 15](#_Toc6532472)

[1.1 Motivation 15](#_Toc6532473)

[1.2 Objectives 19](#_Toc6532474)

[1.3 Use Cases 20](#_Toc6532475)

[1.4 Outline 20](#_Toc6532476)

[2 BACKGROUND INFORMATION 22](#_Toc6532477)

[2.1 What is an Information Security Analyst? 22](#_Toc6532478)

[2.2 What is Natural Language Processing? 23](#_Toc6532479)

[2.3 What is Text Mining? 24](#_Toc6532480)

[2.4 Twitter Social Network 25](#_Toc6532481)

[2.4.1 What is Twitter? 25](#_Toc6532482)

[2.4.2 Twitter API 26](#_Toc6532483)

[2.5 Hürriyet Turkish Newspaper as a Data Source 27](#_Toc6532484)

[2.6 Python Programming Language and why I choose Python for the implementation 29](#_Toc6532485)

[2.7 Istanbul Technical University NLP API 29](#_Toc6532486)

[2.8 SQLite Database 29](#_Toc6532487)

[3 LITERATURE REVIEW 29](#_Toc6532488)

[4 SYSTEM ARCHITECTURE, DESIGN AND IMPLEMENTATION 30](#_Toc6532489)

[4.1 Approach 30](#_Toc6532490)

[4.2 Taxonamy 30](#_Toc6532491)

[4.3 Data Collection 30](#_Toc6532492)

[4.4 Data Preprocessing 30](#_Toc6532493)

[4.5 Event Detection 30](#_Toc6532494)

[5 IMPLEMANTATION AND EVALUATION 31](#_Toc6532495)

[5.1 Implemantation 31](#_Toc6532496)

[5.2 Events Evaluation 31](#_Toc6532497)

[5.2.1 Yahoo Data Breach ie 31](#_Toc6532498)

[6 RESULTS 31](#_Toc6532499)

[7 CONCLUSION AND FUTURE WORK 31](#_Toc6532500)

[7.1 Conclusion 31](#_Toc6532501)

[7.2 Future Work 31](#_Toc6532502)

[REFERENCES 32](#_Toc6532503)

[APPENDICES 34](#_Toc6532504)

[APPENDIX A 34](#_Toc6532505)

[APPENDIX B 35](#_Toc6532506)

# LIST OF TABLES

[Table 1: Decision Making Terminology 1](#_Toc444263431)

[Table 2: xxxx 7](#_Toc444263432)

# LIST OF FIGURES

[Figure 1: Tweets in Turkish After the TurkTrust Vulnerability Announcement on 3 January 2013 17](file:///C:\Users\Ozgur\source\repos\MSThesis\tez\tez.docx#_Toc6532356)

[Figure 2: Hürriyet newspaper news after the Turktrust SSL vulnerability detected 18](file:///C:\Users\Ozgur\source\repos\MSThesis\tez\tez.docx#_Toc6532357)

[Figure 3: Research results of IBM Security Lab about Cyber Security Analysts 23](file:///C:\Users\Ozgur\source\repos\MSThesis\tez\tez.docx#_Toc6532358)

[Figure 4: A Simple diagram to explain what NLP does 24](file:///C:\Users\Ozgur\source\repos\MSThesis\tez\tez.docx#_Toc6532359)

[Figure 5: Sample Turkish Tweets Related with a Security Incident 26](file:///C:\Users\Ozgur\source\repos\MSThesis\tez\tez.docx#_Toc6532360)

# LIST OF ABBREVIATIONS

|  |  |
| --- | --- |
| **DDOS** | Distributed Denial of Service |
| **DOS** | Denial of Service |
| **REST** | Representational State Transfer |
| **API** | Application Programming Interface |
| **HTTP** | Hyper-Text Transfer Protocol |
| **OData** | Open Data Protocol |

**CHAPTER 1**CHAPTER

# INTRODUCTION

## Motivation

Security awareness tools helps security analysts to protect a company's sensitive and mission-critical data from being stolen, damaged or compromised by hackers. The delay between the disclosure and the moment when the security analyst is aware for each newly discovered vulnerability is very important.

On 3 January 2013, Google announced a security vulnerability which could allow spoofing using fraudulent digital certificates issued by TURKTRUST Inc.[1] Other companies like Microsoft and Mozilla which may affect from this vulnerability followed Google and announced the vulnerability, shared their affected software and devices and suggested actions. After these announcements, Twitter and Turkish newspapers showed quick reaction. As you can see in the Figure 1, Twitter users had shared the information on the same day immediately after the announcement on 3 January 2013. I notice that one of the profile who share information about the vulnerability was Asst. Prof. Dr. Süleyman Özarslan, one of the past instructors of METU Cyber Security Department. This is an example of even university professors share timely and valuable informations at Twitter. Because of TurkTrust directly related with Turkey, there were so many Tweets in Turkish related with that vulnerability. According to Statistia website, Twitter has 321 million monthly active users at worldwide[2] and with nearly 9 million active users Turkey is fifth country in the list of leading countries based on number of Twitter users as of January 2019.[3] Twitter users can tweet in any languages they wish. As predictable, 9 million Turkish Twitter users share their tweets generally in Turkish language.

During my literature review and commercial research and products reviews, I realized that even there are a few researches and products to focus on automatic security Event Detection, they mainly focus data mining in English language. There is no research focused on real time security event detection focused Turkish language streams. With emerging internet adoption in Turkey, there are many timely information shared within Turkish language. These Turkish informations not much useful without automated event detection systems Turkish tweets can be also used to detect cyber security events if Turkish language specific steps and requirements shall be added to the detection systems and algorithms.

Figure 1: Tweets in Turkish After the TurkTrust Vulnerability Announcement on 3 January 2013



Of course, social media is not only option to learn such information. The security analyst has a wide range of sources available like specialized press, blogs, forums, news agencies, newspapers and so on to gather cyber threat informations. However, their first reference for detect such security events is generally social networks. After a trending event occurred, people increasingly share posts about it on social media. To illustrate, A DDOS attack to a service or a website generally firstly recognized and reported by users and they share that on online platforms like tweeting "X website is unreachable".

An alternative way to learn such information is newspapers. After the Türktrust SSL vulnerability in 2013, the newspapers also share that information as fast as possible. In the Figure 2 can be seen a news published in Hurriyet newspaper related with the vulnerability.[4]

Figure 2: Hürriyet newspaper news after the Turktrust SSL vulnerability detected



An autonomous system which can use different kind of data source for security event detection would be very useful for a security analyst. I designed and developed a software system capable of detecting and monitoring cyber security related events over the Twitter Stream in Turkish. It can technically process millions of documents per day and detect security events. To gain more accurate results, I added Hürriyet Turkish newspaper stream to analyze and detect security events. The software solution’s infrastructure supports adding new data resources easily. For example, it can be added LinkedIn, Facebook, Eksisozluk website streams to gain more accurate results.

## Objectives

The objectives of this thesis are that;

* Researching and comparing “state of the art” studies and software systems in real time event detection with literature review.
* Researching possible data sources to determine the omst suitable ones to use it for real time event detection with Turkish language.
* Research useful studies and Api’s related with NLP to use it for normalization of Turkish texts.
* Designing and developing the software system to use it real time cyber security event detection with using Turkish texts.
* Design the system as a framework in order to be used it for other researches. Turkish datasets is very useful for various research areas like text classification, author detection, automatic question answering and so on. However finding datasets in Turkish language is extremely hard because there is not enough shared datasets online. With the thesis software framework researchers easily get datasets in Turkish language. Moreover they can select and modify their queries with changing keyword vectors.
* Validating the proposed approach using several detected events within shared in Turkish language.

## Use Cases

Cyber security is an emerging topic in Turkey just like the rest of the world. There are few researches about automated security event detection system in the world. But these are mainly focus on data mining in English language. Even if these systems can be very useful for detecting global level cyber security events such systems cannot be used with other languages like Turkish, because data mining is a language specific area. Security analysts who work in Turkey or just interested in Turkey specific security events can be use data in Turkish language to detect such events. With observing these, a security analyst gets situation awareness in cyberspace and take actions against new threads. For example, assume that you are a security analyst work for a Turkish company and your company uses local websites api like ekşisozluk api e-devlet api or libraries/frameworks developed for focused Turkish people. If these api’s, libraries or frameworks has vulnerabilities and they are newly discovered, they are probably discussed and announced within social media like Twitter in Turkish. Turkish newspapers are published it as breaking news and so on. To detect such events automatically, the software system shall listen Turkish data sources and process the data Turkish language specific. My research meets these requirements to develop such software system and framework.

## Outline

* Chapter 2 is related with background information. Technologies, terminologies and common terms discussed and explained in this section.
* In Chapter 3, It is shared literature researches related with my thesis.
* In Chapter 4, the software system is explained architectural and design perspective.
* In Chapter 5, the software system is explained implementation and evelutation perspective.
* In Chapter 6, thesis results are discussed.
* And finally in chapter 7, the conclusion is stated and possible future works are explained.

**CHAPTER 2**

# BACKGROUND INFORMATION

## What is an Information Security Analyst?

According to careerexplorer website[5] an information security analysts main responsibility is taking measures to protect his company’s mission critical and sensitive information and be prepared for cyber attack. In order to be prepared for cyber attack they use different tools and systems. One of their responsibilty is analyzing data and recommending changes to higher ups. However they are not authorize and implement changes. Their main job is keeping cyber attacks out.

Security analysts spend nearly one hour with getting cought up on the latest security news through bulletins, forums, news, social networks an so on to identify new threads. Their two or three hours are spent with repeatedly investigatation of potential security incidents with using online resources. Rest of their daily time is spent with manually copying and pasting information from disparate and siloed tools to correlate data. They generally face with ten to twenty challenge daily. Their investigation time gives cyber attackers advantages and it is very hard for them to keep up with threads.

In the figure Figure 3[6] below, there is a result of a research about security analysts to explain why automated systems are needed for a security analyst.

It is not sustainable without automated systems.

Figure 3: Research results of IBM Security Lab about Cyber Security Analysts



## What is Natural Language Processing?

NLP is “ability of machines to understand and interpret human language the way it is written or spoken”[7]. In the figure 4[8], can be seen a simple explanation of What NLP does. In this thesis, I used a few NLP techniques and Istanbul Techincal University’s NLP Api[9] for normalization of the texts purpurse.

Figure 4: A Simple diagram to explain what NLP does



## What is Text Mining?

The Oxford English Dictionary defines text mining as “the process or practice of examining large collections of written resources in order to generate new information, typically using specialized computer software.”[10] Text mining consists of a broad variety of methods and technologies. In my thesis, I used Keyword-based technologies and statistics technologies. According to expertsystem website Keyword-based technologies definition is “The input is based on a selection of keywords in text that are filtered as a series of character strings, not words nor concepts.”[11] and statistics technologies definition is “Refers to systems based on machine learning. Statistics technologies leverage a training set of documents used as a model to manage and categorize text.”[11]

## Twitter Social Network

### What is Twitter?

Twitter is an online social networking service, which was created in October, 2006 by Jack Dorsey, Even Williams and Biz Stone. Twitter can be used for various purposes[12].

First of all, It can be used as a social messaging service. Users can interact with the other users, communicate with their friends and family and share details of their lives. Secondly, It can be use as a micro blogging service for sharing details of a person’s life. Thirdly, Twitter can be used as a marketing tool for public relations. Many celebrities and politicians use Twitter for interacting with their audience. Lastly, Twitter is an information platform on which users can get news via broadcasting agents’ or journalists’ accounts fast and easily. Moreover, there are lot of Twitter bots created by developers for a precise function like Bitcoin ticker bot will tweet every hour the price of Bitcoin in Turkish Lira.

According to the first quantitative study on Twitter “What is Twitter, a Social Network or a News Media?” which is published in 2010[13], Twitter is more an information sharing network than a social network. They found that result while working on Twitter follower graph. They decided that because of the low rate of reciprocated ties. People tend to use Twitter as a news feed by following multiple online news media but other Twitter users will only follow “real” users.

Twitter users can post a short message called tweet which is limited with 280 characters, or retweet another user tweet.Photos, videos or URLs can be added to the tweets. Users can follow other accounts and creates their own networks. They can mention about each others or reply each other within their tweets. In order to identify what the tweet is about, users use word preceded by a hash sign (#). Twitter use these hastags to define trending topics both locally and globally. Users use the trending topic lists to identify popular subjects at that time on Twitter. In default settings, all Twitter accunts are public. Users can interact with each others like replying another users tweets, sending private direct message and so on.

Figure 5: Sample Turkish Tweets Related with a Security Incident



### Twitter API

The Twitter API is basically set of URLs. The URLs cant take parameters and let users access lots of Twitter features like finding tweents which contain a set of specific words and so on.

Twitter provides several Application Programming Interfaces (APIs) to get tweets:

Twitter’s REST API allows users to get tweets which or search terms which includes specific paramaters. It is useful for analytics on historal data because this api does not give users live data. Moreover, this api cannot retrive tweets older than 7 days.

Twitter’s Streaming API gives users access to live data in Twitter and keeps sending it until asked it to stop. Developers can access only 1% sample of all the tweets. Its ideal usage is when doing analytics over live campaigns on Twitter and so on.

## Hürriyet Turkish Newspaper as a Data Source

* + 1. What is Hurriyet Newspaper?

Hürriyet is one of the major Turkish newspapers, founded in 1948. As of January 2018, it had the highest circulation of any newspaper in Turkey at around 319,000.[14]

* + 1. Hurriyet Newspaper API

Hürriyet API is an interface which enables the usage of Hürriyet data programmatically in web / mobile / desktop applications. It is a free service. With Hürriyet API, it can be reached news, columns, writers, photo galleries and pages. Hürriyet API was established with a RESTful-based, resource-oriented architecture. Hürriyet data can be accessed via standard HTTP requests. The resultant set of results is in JSON format. Requests via the API are limited to 5 per second and 500 per hour to prevent abuse.[15]

* + 1. What is OData

OData is REST-based data source using the HTTP protocol is a global protocol for querying services. With OData standards, you can quickly, without wasting much time on basic standards such as request and response headers, status codes, HTTP methods (GET, POST and so on), and query options. you can only create RESTful APIs by building business logic.

It is easy to consume OData services. OData metadata is easily rendered by the client - interpretable. Therefore, integration into powerful and expandable client applications can be done quickly and easily.

* + 1. Hürriyet API - OData Usage

The OData structure has a unique query structure. Below are some of the most basic query keywords and their functionality briefly outlined:

$ select : Limits the columns / properties in the response set from the query. Example use;

* https://api.hurriyet.com.tr/v1/articles?$select=Title

To limit relational properties such as Files, RelatedNews; it is necessary to use $ select filter with $ expand. Example use;

* https://api.hurriyet.com.tr/v1/articles?$select=Files&$expand=Files

$ filter : By adding a filter to the query, the answer set can be limited. Example use;

* https://api.hurriyet.com.tr/v1/articles?$filter=Path eq '/gundem/'

You can also use these keywords together to increase the number of filters in the result set and make it easier to reach the desired result set.

Using OData protocol on Hürriyet API service, these can be queried and used in applications.

* Articles in the system
* Columns in the system
* In-system photo galleries
* The pages in the system and the pages assigned to the pages
* Folders in the system
* Writers

Requests via the API are restricted for the purpose of blocking abuse. These limits are 5 requests per second and 500 requests per hour.

## Python Programming Language and Why I Choose Python for the Implementation

“Python is an interpreted, object-oriented, high-level programming language with dynamic semantics.”[16] It is a multi-paradigm programming language and supports so many paradigm like object oriented programming, structured programming, functional programming and so on. It has enough frameworks and Apis to work on cognitive science, text mining, NLP like areas. It is fast enough, and learning it is easy. Most of big companies use Python in data mining projects. To illustrate, according to a 2014 article in Fast Company magazine, Facebook choose to use Python for data analysis because it was already used so widely in other parts of the company.[17] In my thesis, I use Python version 3.6.6.

## Istanbul Technical University NLP API

## SQLite Database

**CHAPTER 3**

# LITERATURE REVIEW

**CHAPTER 4**

# SYSTEM ARCHITECTURE, DESIGN AND IMPLEMENTATION

## Approach

## Taxonamy

## Data Collection

## Data Preprocessing

## Event Detection

**CHAPTER 5**

# IMPLEMANTATION AND EVALUATION

## Implemantation

## Events Evaluation

### Yahoo Data Breach ie

**CHAPTER 6**

# RESULTS

**CHAPTER 7**

# CONCLUSION AND FUTURE WORK

## Conclusion

## Future Work

# REFERENCES

[1] A. Langley, “Enhancing digital certificate security,” 2013. [Online]. Available: https://security.googleblog.com/2013/01/enhancing-digital-certificate-security.html. [Accessed: 15-Apr-2019].

[2] Twitter, “Twitter: Number of active users 2010-2017,” *statista.com*, 2019. [Online]. Available: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/. [Accessed: 15-Apr-2019].

[3] “Countries with most Twitter users 2019 | Statistic,” *Statista*, 2019. [Online]. Available: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/. [Accessed: 15-Apr-2019].

[4] “Yanlış sertifika Google’dan döndü - Teknoloji Haberleri.” [Online]. Available: http://www.hurriyet.com.tr/teknoloji/yanlis-sertifika-googledan-dondu-22290509. [Accessed: 16-Apr-2019].

[5] “What does an information security analyst do? ‐ CareerExplorer.” [Online]. Available: https://www.careerexplorer.com/careers/information-security-analyst/. [Accessed: 18-Apr-2019].

[6] M. Borrett, “Security in the Cognitive Era BRINGING THE POWER OF COGNITIVE SECURITY TO THE SECURITY ANALYST Motivations for Change,” 2017.

[7] “International Journals of Management, IT et Engineering IJMIE.” International Journals of Multidisciplinary Research Academy, IJMRA.

[8] “Overview of Artificial Intelligence &amp; Role of NLP in Big Data - XenonStack Blog.” [Online]. Available: https://www.xenonstack.com/blog/ai-nlp-big-deep-learning/. [Accessed: 18-Apr-2019].

[9] “ITU Turkish Natural Language Processing Web Interface.” [Online]. Available: http://tools.nlp.itu.edu.tr/index.jsp. [Accessed: 18-Apr-2019].

[10] Stephanie Prato, “What is Text Mining? - Information Space,” 2013. [Online]. Available: https://ischool.syr.edu/infospace/2013/04/23/what-is-text-mining/. [Accessed: 18-Apr-2019].

[11] “Text mining vs data mining: discover the differences -.” [Online]. Available: https://www.expertsystem.com/text-mining-vs-data-mining-differences/. [Accessed: 18-Apr-2019].

[12] B. A. Huberman, D. M. Romero, and F. Wu, “Social Networks that Matter: Twitter Under the Microscope,” *SSRN Electron. J.*, 2008.

[13] H. Kwak, C. Lee, H. Park, and S. Moon, *What is Twitter, a Social Network or a News Media?* .

[14] “Tiraj | MedyaTava - Yazmadıysa Doğru Değildir.” [Online]. Available: http://www.medyatava.com/tiraj/2018-01-08. [Accessed: 11-Apr-2019].

[15] “Hurriyet Developers API v1.0 Docs — Hürriyet Public API.” [Online]. Available: https://developers.hurriyet.com.tr/docs/versions/1.0. [Accessed: 11-Apr-2019].

[16] Python, “What is Python? Executive Summary | Python.org,” *Python Software Foundation*. 2017.

[17] “Businesses Can Now Use The Same Stats Language As Universities, Thanks.” [Online]. Available: https://www.fastcompany.com/3030877/businesses-can-now-use-the-same-stats-language-as-universities-thanks-to-pandas. [Accessed: 19-Apr-2019].

# APPENDICES

# APPENDIX A

**TITLE**

xxx

# APPENDIX B

**TITLE**

xxx