



Türkçe Tweetler Üzerinde Otomatik Soru Tespiti

İslam MAYDA*

Yıldız Teknik Üniversitesi, Elektrik-Elektronik Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul

Özet

Bu çalışmada Twitter'da paylaşılan Türkçe yazılmış tweetlerden, içinde soru geçenleri kural tabanlı yaklaşımla tespit edilmeye çalışılmıştır. Veri seti olarak rastgele örneklenen tweetler yerine belirli bir etiket ile paylaşılan tweetler kullanılmıştır. Bunun sebebi, bu çalışmada belirli bir odağa yönelik sorulan soruların tespitinin amaçlanmış olmasıdır. Deneyler için, bir televizyon kanalında canlı olarak yayınlanan programda tarihi bir konu konuşulurken izleyicilerin sorularıyla programa katkıda bulunması amacıyla açılan etiketle paylaşılan 354 tweet toplanmıştır. Bu tweetlerdeki yazım hatalarını düzeltmede Zemberek kütüphanesinden yararlanılmıştır. Daha sonra Türkçe soru cümlesi yapısına göre, kesinlik değerini veya duyarlılık değerini yüksek tutmayı amaçlayan 3 farklı basit kural tanımlanmış ve her biri ayrı yöntem olarak deneylerde uygulanmıştır. Deneyler sonucunda %100 kesinlik, %96,48 duyarlılık ve 0.929 F-skoru değerleri en başarılı performanslar olarak kaydedilmiştir.

Anahtar Kelimeler: Twitter, Türkçe Tweetler, Soru Tespiti

Makale Bilgisi

Başvuru:

29/11/2018

Kabul:

24/12/2018

Automatic Question Identification on Turkish Tweets

Abstract

In this work, it was tried to identify which are includes question from the tweets written in Turkish shared in Twitter, with a rule-based approach. As a data set, tweets that are shared with a certain hashtag are used instead of randomly sampled tweets. The reason for this, it is aimed to identify the questions asked for a specific focus in this study. For the experiments, 354 tweets were collected, shared with the hashtag that was created in order to contribute of the audience to the program by asking the questions while a historical topic was being spoken in the program broadcast live on a television channel. The Zemberek library has been used to fix typos in these tweets. Then, according to the Turkish question sentence structure, 3 simple rules are defined aiming at keeping the precision value or the sensitivity value high and each one is applied as a separate method. As a result of experiments, 100% precision, 96.48% sensitivity and 0.929 F-score values were recorded as the most successful performances.

Keywords: Twitter, Turkish Tweets, Question Identification

* İletişim e-posta: islam.mayda@stu.khas.edu.tr

1 Giriş

Dünyanın en popüler mikroblog ağı olan Twitter, 2017 yılının 3. çeyreği itibarıyla 330 milyon aktif kullanıcıya sahiptir [1]. Twitter sadece kullanıcıların bilgi, görüş, öneri veya şikâyetlerini paylaştıkları bir sosyal medya ağı olmayıp, aynı zamanda cevap almak amacıyla çeşitli konularda soru sordukları bir platform haline gelmiştir [2]. Efron ve Winget [3] tarafından yürütülen araştırmada rastgele örneklenen 2 milyon tweetin %13'ünün soru olduğu görülmüştür. Li ve diğerleri [4] tarafından yapılan çalışmada ise rastgele bir saatlik dilim içerisinde atılan 2,045 tweetten oluşan veri setinin %11'inin soru içeriği belirtilmiştir.

Soru sormak için Twitter'ın tercih edilmesinin nedeni, e-posta gönderme veya telefonla aramaya göre tweet atmanın daha kolay olması ve bazen de daha hızlı geri dönüş alınması olduğu düşünülebilir. Bu yüzden, bir elektronik firması, bir üniversite veya bir belediye gibi kamu kurumları ve özel sektör kuruluşlarının resmi Twitter hesapları sık sık diğer kullanıcıların sorularıyla muhatap olmaktadır.

Televizyon kanallarında canlı olarak yayınlanan tartışma, söyleşi ve spor programları başta olmak üzere birçok programda Twitter üzerinden izleyicilerden sorular alınmaktadır. Bu tür programlarda genelde bir etiket belirlenmekte ve bu etiket üzerinden izleyicilerden soruları ve görüşlerini iletmesi istenmektedir. Programın moderatörü, izleyiciler tarafından atılan çok sayıda tweet içinden soru tweetlerini kendisi seçerek konuya veya konuklara yöneltmektedir.

Bu çalışmada amaçlanan soru tespiti işi, bir televizyon programında belirlenen etikete veya resmi bir Twitter hesabı gibi sadece belirli hesaplara ve/veya konulara yönelik atılan tweetlerden soruların otomatik algılanmasıdır. Twitter kullanıcılarının kendi aralarındaki günlük yazışmalarında geçen "Napıyorsun, Nasıl gidiyor, Naber, Noldu" gibi soruların geçtiği tweetlerin tespit edilmesi bu çalışmanın hedefi dâhilinde değildir.

Soru tespiti üzerine yapılan çalışmalara bakıldığında kural tabanlı ve öğrenme tabanlı olarak iki farklı yaklaşımın kullanıldığı görülmektedir [4]. Li ve diğerleri [4] tarafından İngilizce tweetler üzerinde yapılan soru tespiti çalışmasında kural tabanlı yöntemin öğrenme tabanlı yöntemle göre daha başarılı sonuç verdiği tespit edilmiştir. 5N1K soru kelimeleri ve soru

işareti ile çeşitli kural kombinasyonları denenmiş ve her biri farklı denemelerden en yüksek performanslar 0.969 kesinlik, 0.973 duyarlılık ve 0.93 F-skoru oranları olarak elde edilmiştir. Li ve diğerleri bu çalışmalarında sadece soru içeren tweetleri tespit etmekle kalmayıp, çalışmanın ikinci aşaması olarak soru içeren tweetlerin arasından "qweet" adını verdikleri gerçekten cevap almak amacıyla sorulan soruları içeren tweetleri de ayırt etmeye çalışmışlardır.

Türkçe'nin gramer yapısı İngilizceden farklı olduğu için soru cümlesinin tespiti de nispeten daha zordur [5]. İngilizce metinler üzerine çalışırken [4]'te olduğu gibi; 5N1K soru kelimesi cümlenin başında geçmeli veya 5N1K soru kelimesinden sonra "is, are, were, ..." gibi yardımcı fiil gelmeli gibi kurallar kullanmak, soru sormak amacıyla kullanılan 5N1K kelimesi ile soru sormak amacıyla kullanılmayan 5N1K kelimesini birbirinden ayırt etmede başarıyı artırmaktadır. Ancak, Türkçe metinler için bu tür kurallar yazmak mümkün değildir.

Türkçe tweetler üzerinde yapılan soru tespiti çalışmalarından biri, Özger ve Diri [6] tarafından yapılan öğrenme tabanlı yaklaşımın uygulandığı çalışmadır. Bu çalışmada, Twitter'da sık kullanılan soru kalıpları çıkarılmış ve bunlardan 73 özellik tanımlanmıştır. Veri seti olarak bir aylık bir sürede atılan 1 milyon tweet içerisinde el ile etiketlenen 136,449 tanesi kullanılmıştır. Makine öğrenmesi yöntemleri ile yapılan sınıflandırmada Destek Vektör Makinesi, Rasgele Orman ve 1-NN algoritmaları ile 0.857 F-skoru değeri elde edilmiştir.

Cengiz ve Diri [2] tarafından yapılan kural tabanlı yöntemin uygulandığı çalışma ise Türkçe tweetler üzerinde yapılan bir diğer çalışmadır. 1045 tweet üzerinde yürütülen araştırmada yazım hatalarını düzeltmek için Zemberek kütüphanesi [7] kullanılmıştır. Düzeltmeler, bu kütüphanenin öner fonksiyonunun önerdiği kelimeler arasından ilk öneri kabul edilerek yapılmıştır. Daha sonra soru kelimeleri belirlenerek, bu kelimeleri içeren tweetler "sözde soru tweeti" olarak etiketlenmiştir. Çalışmanın ikinci aşamasında Türkçeye özgü çeşitli dilbilgisi kuralları oluşturularak sözde soru tweetlerinin arasından gerçek soru tweetleri tespit edilmeye çalışılmıştır. Sonuç olarak, 0.68 kesinlik, 0.98 duyarlılık ve 0.80 F-skoru değerleri elde edilmiştir.

Bu çalışmanın önceki çalışmalardan en önemli farkı rastgele örneklenen tweetler üzerinde değil, belirli

bir odağa yönelik atılan tweetler üzerinde çalışılması ve bu alandaki ihtiyaca yönelik olarak oldukça basit kural tabanlı bir sistemin önerilmesidir. Ayrıca, hatalı kelimelerin Zemberek kütüphanesi yardımıyla düzeltilmesinde iyileştirme yapılmış ve bu da doğru şekilde düzeltmelerin oranında artış sağlamıştır. Bundan sonraki 2. bölümde çalışmada kullanılan verilerin toplanması ve ön işleme hakkında bilgi verilirken, 3. bölümde önerilen sistem Uygulama başlığı altında açıklanmıştır. Kesinlik veya duyarlılık değerlerinden hangisinin yüksek tutulması hedefleniyorsa ona göre üç farklı kural önerilmiştir. Son olarak 4. başlıkta da çalışmada elde edilen sonuçlar sunulmuş ve 5. başlıkta bu sonuçlar değerlendirilmiştir.

2 Veri Seti

2.1 Veri Toplama

2 Mayıs 2017 tarihinde Habertürk TV kanalında canlı olarak yayınlanan Öteki Gündem programında

"Osmanlı'nın Rönesans'ı Kabul Edilen Lale Devri" konusu ele alınmıştır¹. Cansu Canan Özgen'in moderatörlüğünü yaptığı programa Doç. Dr. Teyfur Erdoğan ve yazar Zekeriya Yıldız konuk olarak katılmıştır. Bu programa izleyicilerin sorularıyla Twitter'dan katkıda bulunabilmesi amacıyla #engüzeli etiketi belirlenmiştir ve program başlamadan önce ilan edilmiş, program boyunca da ilgili etiket ekranda sık sık hatırlatılmıştır. Programın yayınlandığı tarihte #engüzeli etiketi ile tüm gün boyunca 354 tweet atılmıştır. Bu tweetlerden 4 tanesinde sadece #engüzeli etiketi kullanıldığı için bu tweetler ihmal edilerek, geriye kalan 350 tweet bu çalışmada kullanılmak üzere toplanmıştır. Çalışmada önerilen yöntem öğrenmeye dayalı olmadığı için toplanan verinin önceden etiketlenmesi gerekmemektedir [2], ancak çalışmanın başarısını ölçebilmek amacıyla veri toplama işleminden sonra soru içeren tweetler el ile etiketlenmiştir. Etiketleme işlemi sonucunda tweetlerin 142 tanesinin (%40.57) soru içerdiği, 208 tanesinin (%59.43) ise içermediği görülmüştür.

Tablo 1. Zemberek kütüphanesinin bazı hatalı kelimeler için sunduğu öneriler

HATALI KELİME	var mıydı	ettimi	olmuş mudur?	Denilebilir mi	neden mi	nasıl	
ÖNERİLER		Eti'mi				nazil	
		etkimi				nasip	
		etlimi			nedeni	nasir	
		vargıydı	etimi		nedenli	nasıl	
		varıydı	etçimi	olmuş mudur	Denilebilirdi	nedenim	nail
		varlıydı	ettim	olmuş mudur	denilebilirim	nedenimi	nakil
		varcıydı	etsimi	olmuş mudur	denilebilirmi	nedenci	Nacil
		varsıydı	eltimi		denilebilir mi	nedendi	nasal
		var mıydı	ettiği			nedensi	nesil
			et timi			nedenmiş	asil
		etti mi			neden mi	basil	
		ettim i				gasil	

2.2 Veri Önleme

Twitter resmi bir mecra olmadığından, kullanıcıların farkında olarak veya olmayarak sıkça

yazım hataları yaptığı bilinmektedir [8]. Günümüzde her bir tweet için geçerli olan karakter limiti 2017'nin son aylarında 280'e çıkarılmıştır², ancak bu çalışmada kullanılan veriler 140 karakter limiti geçerli iken paylaşılan tweetlerdir.

¹ <https://www.youtube.com/watch?v=6p3VZlgBRt4>

² <https://www.independent.co.uk/life-style/gadgets-and-tech/news/twitter-280-characters-tweets-start-when-get-latest-a8042716.html>

Twitter'daki bu karakter kısıtlamasından dolayı da soru eki gibi ayrı yazılması gereken bazı ekler bilerek bitişik yazılabilmekte veya bazı kullanıcılar karakterden tasarruf etmek amacıyla bazı sesli harfleri yazmamayı tercih etmektedir. Bu yüzden, toplanan tweetlerdeki yazım hatalarını gidermek amacıyla Zemberek kütüphanesinden faydalanılmış, noktalama işaretlerinden ayrıştırılarak kelime bazlı yazım kontrolü yapılmış ve sözlükte bulunmayan kelimeler kütüphane tarafından öneriye göre düzeltilmiştir.

Yapılan gözlemlerde, soru ekiyle bitişik yazılan hatalı kelimeleri Zemberek kütüphanesinin çoğu kez ilk öneride değil de sonraki önerilerde doğru şekilde ayırdığı görülmüştür. Düzeltme işleminde başarıyı artırmak amacıyla; eğer hatalı kelimedede $m[iuü][^s]$ düzenli ifade deseni geçiyorsa bu hatalı kelimenin düzeltilmesi için sunulan önerilerde, varsa ikinci kelimesi "mı, mi, mu, mü" ile başlayan iki kelimelik ilk öneri tercih edilirken, eğer hatalı kelimedede ilgili desen geçmiyorsa doğrudan ilk öneri kullanılmıştır. Desende görüldüğü gibi "mı, mi, mu, mü" ekini ş harfinin takip etmemesi istenmiştir, bunun amacı da bitişik yazılan soru ekleri ile miş'li geçmiş zaman kalıbıyla yazılan hatalı kelimelerin birbirine karıştırılmasının önüne geçmektir.

Ayrıca "nasıl" soru kelimesi Türkçe karakter içerdiği için, bazı kullanıcılar tarafından sıkça "nasil" olarak yazılmaktadır. Ancak, Zemberek kütüphanesi bu hatalı yazımın önerisini ilk sırada değil, daha alt sıralarda vermektedir. Bu kelime çalışmamız için önemli olduğundan dolayı, "nasil" şeklinde hatalı yazılan kelimeler özel olarak "nasıl" önerisi ile düzeltilmiştir.

Önişlemede kullanılan kurallarla yapılan düzeltmelere Tablo 1'de örnekler verilmiştir. Düzeltmelerde kullanılan öneri kalın yazı tipi ile belirtilmiştir. Hatalı yazımların düzeltilmesinin ardından, yine Zemberek kütüphanesinin kelime çözümleme fonksiyonu kullanılarak tweetlerde geçen kelimeler çözümlenmiştir.

3 Uygulama

Bu çalışmada geleneksel kural tabanlı yaklaşım kullanılmıştır. Öğrenme söz konusu olmadığı için de önceden etiketli veriye gerek duyulmamıştır.

Soruları genel olarak aşağıdaki gibi iki gruba ayırabiliriz;

- 5N1K kelimelerini içeren sorular
- Evet/hayır soruları

Buna göre dilbilgisi hatası yapılmadan yazılmış bir soru cümlesi ya 5N1K kelimesi ya da "mı, mi, mu, mü" soru eki içerir ve her iki durumda da cümle soru işareti ile bitirilir. Ancak, tweette soru geçmesine rağmen, kullanıcı soru işareti kullanmamış da olabilir. 5N1K kelimeleri olarak seçilen kelimeler şunlardır: "ne, neden, niçin, nasıl, nerede, kim, kaç".

Otomatik soru tespiti yapılmak istenen kimi platformlarda kesinlik değerinin yüksek tutulması, kimi platformlarda ise duyarlılık değerinin yüksek tutulması amaçlanabilir. Buna göre, çalışmada üç farklı kural kullanılmıştır. İlk kuralda kesinlik değerinin yüksek tutulması amaçlanmışken, ikinci ve üçüncü kurallarda duyarlılık değerinin yüksek tutulması hedeflenmiştir. Çalışmada kullanılan üç kural aşağıdaki gibi tasarlanmıştır:

1. kural: [(5N1K kelimesi) VE (Soru işareti)] VEYA

[(Soru eki) VE (Soru işareti)]

2. kural: [(5N1K kelimesi) VEYA (Soru eki)]

3. kural: [(5N1K kelimesi) VEYA (Soru eki) VEYA

(Soru işareti)]

Soru işaretleri web sitesi sayfalarının bağlantılarında da geçebildiği için, bu tür bağlantılar içeren tweetlerin hatalı sınıflandırılmasına engel olmak için düzenli ifade ile gerekli kontrol yapılmıştır.

4 Sonuçlar

Test sonuçlarını Tablo 2'deki gibi bir karışıklık matrisi ile ifade edecek olursak, bu matristen yola çıkarak doğruluk, kesinlik, duyarlılık ve F-skoru değerlerini elde edebiliriz [9].

Tablo 2. Karışıklık matrisi.

		Tahmin	
		Soru içeren	Soru içermeyen
Gerçek	Soru içeren	Doğru Pozitif	Yanlış Negatif (FN)
	Soru içermeyen	Yanlış Pozitif (FP)	Doğru Negatif (TN)

$$\text{Doğruluk} = (TP + TN) / (TP + FP + FN + TN)$$

$$\text{Kesinlik} = TP / (TP + FP)$$

$$\text{Duyarlılık} = TP / (TP + FN)$$

$$\text{F-skoru} = (2 * \text{Duyarlılık} * \text{Kesinlik}) / (\text{Duyarlılık} + \text{Kesinlik})$$

Kullanılan üç farklı kuralın ayrı ayrı uygulanmasıyla elde edilen doğruluk, kesinlik, duyarlılık ve F ölçümü değerleri Tablo 3'te sunulmuştur.

Tablo 3. Deneysel sonuçlar.

	1. Kural	2. Kural	3. Kural
Doğruluk	%82.57	%94.00	%94.00
Kesinlik	%100.00	%91.16	%89.54
Duyarlılık	%57.04	%94.37	%96.48
F ölçümü	0.726	0.927	0.929

1. kuralda kesinlik değerinin yüksek olması, 2. ve 3. kuralda ise duyarlılık değerinin yüksek olması amaçlandığından, sonuçların bu hedeflere uygun olduğu görülmektedir. Üç ayrı deneyde ortaya çıkan karmaşıklık matrisleri Tablo 4, Tablo 5 ve Tablo 6'da verilmiştir.

Tablo 4. 1. Kural için karışıklık matrisi.

		Tahmin	
		Soru içeren	Soru içermeyen
Gerçek	Soru içeren	81	61
	Soru içermeyen	0	208

Tablo 5. 2. Kural için karışıklık matrisi.

		Tahmin	
		Soru içeren	Soru içermeyen
Gerçek	Soru içeren	134	8
	Soru içermeyen	13	195

Tablo 6. 3. Kural için karışıklık matrisi.

		Tahmin	
		Soru içeren	Soru içermeyen
Gerçek	Soru içeren	137	5
	Soru içermeyen	16	192

Bir soru tweetini tespit etmek için soru kelimesi veya soru eki yeterli olmalıdır. Bu nedenle 2. kuralda sadece 5N1K kelimesi veya soru eki içeren tweetler soru tweeti olarak etiketlenmiştir. Ancak, Twitter'da sıkça yazım hatası yapıldığı ve yazım hatalarını düzeltmek için kullanılan kütüphanenin

bütün hataları düzeltilmediği görülmüştür. Yanlış yazılan ve kütüphane tarafından düzeltilmeyen 5N1K kelimeleri ve soru eki içeren soru tweetlerini de soru tweeti olarak etiketleyebilmek için 3. kuralda sadece soru işareti içeren tweetler de soru tweeti olarak sınıflandırılmıştır. 2. kuralda soru tweeti olarak etiketlenmeyen ancak 3. kuralda yakalanan tweetlere örnekler aşağıda verilmiştir.

- *peki bu kürkler, yüzükler, kemerler **nerde?** #engüzeli*
- *#engüzeli @ht_otekigundem çiçek aşısının bulunma süreci nasıl işledi aşının bulunmasını tetikleyen olay neydi,o sıralar bu hastalık salgın **mydi?***
- *#engüzeli halkın ayaklandıktan sonra damat ibrahim paşa'nın kellesini istemişti olayı **doğrumu?***

F ölçümü değeri daha yüksek olan 3. kuralın kullanıldığı test sonucunda 21 adet tweet hatalı sınıflandırılmıştır. Hatalı sınıflandırılan tweetler incelendiğinde, bunların 4 tanesinin Zemberek kütüphanesinin düzeltilmediği yazım hatası içermesi nedeniyle yanlış sınıflandırıldığı görülmüştür. Düzeltilmeyen yazım hataları nedeniyle 3. kuralla yapılan deneyde yanlış sınıflandırılan tweetler aşağıda sunulmuştur.

- *#engüzeli hollandilo resamin ismini tekrarlıya **bilirmisini.***
- *#engüzeli hollanda'li resamin kitabına tekrar **gosterebilirmisiniz.***
- *#engüzeli avrupada lale tulipa die geciyor bunun sebebi osmanlida lale tülbentlere sarılıp satılmasından geliyor die biliyorum bu **doğrumu..***
- *#engüzeli nevhşhrlı ibrahim acaba şavaş dönemi veziri olmamasının sebebi yeniçeriocağından **gelmemesi idir acaba***

3. kuralla yapılan deneyde, 17 tweetin 5N1K soru kelimeleri veya soru ekleriyle ilgili bir yazım hatası içermemesine rağmen yanlış sınıflandırdığı görülmüştür. Bu tweetlerden bazıları aşağıda listelenmiştir.

- *#engüzeli **ne** üzücü baltacı gibi bir kahraman bir paşaya atılan bir iftira*
- *yeniliğe karşı olmak zaten kaybetmektir. asıl mesele bu zaten yenileşmenin başarısızlığı çöküşün en büyük **nedenlerinden** biri oldu. #engüzeli*
- *#engüzeli tarih dersni **nasıl** sıkıçı bulunur anlamış değilin 🐱🐱*

- #engüzeli @cansucananozgen twitter hesabınız @cansucananozgenn yazıyor?
- #engüzeli her hafta osmanlı tarihi sıkıyor biraz da diğer imparatorluklarımıza dönsek? mesela türk çocuğu emir timur ve babür şah' bilmiyor.

5 Tartışma

Çalışma sonucunda Türkçe metinlerde hata düzeltme için daha başarılı çalışan bir araca ihtiyaç duyulduğu görülmüştür. Hata düzeltmede kullanılan Zemberek Kütüphanesi her hatalı kelimeyi tekil olarak değerlendirdiği için, deney sonuçlarında görüldüğü gibi basit bazı yazım hatalarını bile doğru şekilde düzeltememektedir. Hata düzeltmede başarının artırılması için hatalı kelimeler tekil olarak değil, bağlama bağımlı şekilde değerlendirilmelidir. Hata düzeltmelerdeki başarının artması soru tespiti işindeki başarının artmasına da katkı sağlayacaktır. Hatalı yazımların düzeltilmemesi dışında, soru sormak amacıyla kullanılan 5N1K kelimeleri ile soru sormak amacıyla kullanılmayan 5N1K kelimelerinin birbirinden ayırt edilememesi de hatalı sınıflandırmaya sebep olmaktadır. Bu ayırımın yapılabilmesi için Türkçe'ye özgü birtakım kurallar da sisteme dâhil edilebilir. 5N1K kelimesinin öncesindeki ve sonrasındaki kelimeler incelenerek, bu kelimenin gerçekten soru sormak amacıyla mı kullanıldığına dair bir tahminde bulunulabilir. Bu çalışmada temel olarak soru içeren tweetler tespit edilmeye çalışılmış ve basit kurallar uygulanarak, kesinlik oranı yüksek bir sonuç elde edilmek isteniyorsa 1. kuralın, duyarlılık oranı yüksek bir sonuç alınmak isteniyorsa 2. veya 3. kuralın tercih edilmesi önerilmiştir.

Teşekkür

Bu çalışmaya verdiği destekten ötürü çok değerli hocam Prof. Dr. Banu DİRİ'ye teşekkürlerimi sunarım.

Kaynaklar

- [1] Statista. "Number of monthly active Twitter users worldwide from 1st quarter 2010 to 3rd quarter 2017 (in millions)". <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/> (10.10.2017).
- [2] Cengiz C, Diri B. "Türkçe Tweetlerden Soru İfadelerini Bulmak". *Sinyal İşleme ve Uygulamaları Kurultayı (SİU)*, Haspolat, Türkiye, 24-26 Nisan 2013.
- [3] Efron M, Winget M. "Questions are content: a taxonomy of questions in a microblogging environment". *Proceedings of the American Society*

for Information Science and Technology (ASIST), Pennsylvania, USA, 22-27 October 2010.

- [4] Li B, Si X, Lyu MR, King I, Chang EY. "Question Identification on Twitter", *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, Glasgow, Scotland, UK, 24-28 October 2011.
- [5] Oflazer K. "Türkçe ve Doğal Dil İşleme". *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5(2), 2012.
- [6] Özger ZB, Diri B. "Sınıflandırma Tabanlı Türkçe Soru Algılama", *Akıllı Sistemlerde Yenilikler ve Uygulamaları (ASYU)*, İzmir, Türkiye, 9-10 Ekim 2014.
- [7] Google Code Archive. "Zemberek". <https://code.google.com/archive/p/zemberek/> (10.10.2017).
- [8] Özger ZB. Türkçe Twitter'da Soru Algılama. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, İstanbul, Türkiye, 2014
- [9] Lavrac N, Flach P, Zupan B, "Rule evaluation measures: a unifying view", *International Conference on Inductive Logic Programming (ILP)*, Bled, Slovenia, 24-27 June 1999.