



**METİN MADENCİLİĞİNDE ALAYCI VE İRONİ İFADELERİN
TANIMLANMASI**

Tümer ŞİBİK

Burak BİNGÖL

Danışman: Prof. Dr. M. Ali AKÇAYOL

GAZİ ÜNİVERSİTESİ

MÜHENDİSLİK FAKÜLTESİ

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

BM - 496 BİLGİSAYAR PROJESİ II

HAZİRAN 2017

ANKARA

ÖZET

Son yıllarda popüler olan ve günümüz dünyasında her geçen gün önemi daha da artan veri madenciliği ve bununla doğrudan ilişkili olan metin madenciliği, ulusal güvenlik ve istihbarat gibi hayati alanların yanı sıra kurumsal iş ve rekabet zekâsı, otomatik reklam yerleştirme gibi uygulamaları ile iş dünyasında ve buna benzer birçok alanda karşımıza çıkmaktadır. Metin madenciliği uygulama alanlarından olan ironi ve alaycı ifadelerin tespiti sorununa, özellikle yapay zekâ ve makine öğrenmesi gibi günümüzün popüler teknolojileri ile birlikte çözümler aranmaktadır.

Bu çalışmada, yazılı metinler içerisindeki ironi ve alaycı ifadelerin bilgisayarlar tarafından tespit edilebilmesi problemine sözlük tabanlı yöntemler kullanılarak çözüm getirilmesi amaçlanmıştır.

İÇİNDEKİLER

Sayfa No

1. GİRİŞ	1
2. PROJE İŞ PLANI	3
2.1. Yazılım Süreç Modeli	3
2.3 Görev Dağılımı	5
2.4 Araç ve Teknikler	6
3. ALAN BİLGİSİ	7
3.1. Daha Önce Yapılan Duygu Durum Analizi Çalışmalarında İzlenen Çözüm Yolları ...	7
3.2. Daha Önce Yapılan İroni Tespiti Çalışmalarında İzlenen Çözüm Yolları.....	9
3.2.1. Parantez içinde soru işareti / ünlem işareti.....	9
3.2.2. Ünlem işareti.....	10
3.2.3. Soru işareti	10
3.2.4. Tırnak işareti	11
3.2.5. Cümle sonundaki nokta (tam durak).....	11
3.2.6. Yüz ifadeleri.....	12
3.2.7. Cümle içindeki duygusal kutuplaşma	14
3.2.8. İsimlerde küçültme.....	15
3.2.9. Ünlem kelimelerinin kullanımı	15
3.2.10. Büyük harf kullanımı	16
4. GELİŞTİRİLEN UYGULAMA	17
4.1. Veri Toplama Aşaması	17
4.2. Sözlük Oluşturma Aşaması.....	18
4.3. Analiz Algoritmasının Geliştirilmesi	18
4.4. Uygulamanın Test Aşaması	20
4.5. Test Sonuçlarının Değerlendirilmesi	21
5. SONUÇ	23
6. KAYNAKÇA.....	25

1. GİRİŞ

İroni, Söylenen şeyin aslında tam tersini kast ederek karşıdaki kişi ya da ilgili olayla dolaylı yoldan alay etmek anlamındadır. Mizah ya da doğrudan alay etmeye göre duruma daha eleştirel bir bakış sunar. İroni mimik, jest ve tonlama ile söylenmek istenenin altını, dolaylı olarak çizer.

Günümüzde literatüre geçmiş üç çeşit ironi varlığı kabul edilmektedir. [1]

- **Verbal İroni (Sözlü İroni):** Burada kişi söylediği şeyin tam tersini kastetmektedir. Bu ironi türü günlük hayatta en çok karşılaşılan ironi türüdür. Dilbilimcilere göre alay ve iğneleme (sarcasm) bu sınıftandır.
- **Situational irony (Durumsal İroni):** Durumsal ironi istenen ya da beklenen sonuca aykırı bir eylemin doğrudan bir sonucudur. Bu tür ironide genellikle ard arda 2 cümle ve bu cümleler arasında tutarsızlık vardır.
- **Dramatic irony (Dramatik İroni):** Daha çok tiyatro, öykü anlatımı ya da filmlerde kullanılır. Örneğin bir karakterin oldukça iyi bir baba ve aynı zamanda çok iyi bir cinayet dedektifi olarak rol aldığı fakat filmin-oyunun sonunda aslında bir katil olduğunun anlaşılması gibi.

Bu çalışmaya konu olan ironi türü, verbal (sözlü) ironinin metinsel (yazılı) boyutudur. Bu çalışmada bilgisayara girdi olarak verilen cümlelerin ironi içerip içermemesi bakımından incelenmesi hedeflenmiştir.

Giderek kullanım alanları ve popülerliği artan Metin ve web madenciliği son yıllarda oldukça fazla çalışılan birbiri ile ilişkili alanlardır. Metin madenciliği, çok büyük belgelerin analizi ve metin tabanlı verinin içerisindeki gizli kalıpların elde edilmesidir. Web madenciliği ise, web içerikleri, sayfa yapıları ve web bağlantı istatistiklerinin de içinde olduğu web ile ilişkili olan verinin analizini içermektedir [3].

İroni ve alaycı ifadelerin tespiti açısından bakıldığında ironinin anlaşılması, yazılı iletişimin mimik, ses tonu, jestler gibi beden dili desteğinden yoksun olması ve hatta bahsi geçen konudan haberdar olma gerekliliği taşıması gibi kişiden kişiye farklılık gösterebilecek özellikler taşıması nedeniyle, bu günkü genel kabul görmüş kanı; günümüz metin madenciliği teknikleri ile ironi ve alaycı ifadelerin bilgisayarlar tarafından algılanması konusuna başarısız olarak bakılmaktadır. Ancak çalışmalar sürmekte ve parçalar halinde umut verici olumlu sonuçlar alınmaktadır.

Konu ile ilgili daha önce yapılan çalışmalar incelendiğinde; İroni içeren ifadelerin öncelikle duygu Durum analizine tabii tutulduğu, ironik olarak tespit edilen cümlelerin büyük bir çoğunluğu olumlu (yüksek olumluluk skoruna sahip) olmasına karşın neredeyse tamamının anlam bakımından olumsuz cümleler olduğu görülmüştür. Bu alanda yapılan çalışmalar oldukça sınırlıdır. Ancak, bu çalışmalarda kullanılan diller çok çeşitlidir. (İngilizce, Portekizce ve Hollanda da kullanılan flemenkçe) ironi tespit etmek için kullanılan bazı teknikler tüm diller için ortak iken, bazı diller için sadece ilgili dile ait olan özelliklerden de faydalanılmıştır [1].

Türkçe, kelime köklerine ekleri ekleyerek yeni kelime ve anlamlar türetilen bir dil olduğundan morfolojik analizi en zor dillerden bir tanesi olarak kabul edilmektedir. (Morfoloji: sözcüklerin yapısının tanımlanmasını, çözümlenmesini ve kimliklendirilmesini inceleyen bilim dalı) Buna örnek olarak Türkçedeki “Gelmemeliydim” sözcüğü İngilizceye çevrildiğinde: “I should have not come” şeklinde tam bir cümle ifade eder. Bu nedenle Türkçede ironi ya da duygu durum tespiti yapabilmek için öncelikle morfolojik analize ihtiyaç duyulmaktadır. Yine İngilizce ile karşılaştırmalı düşünersek İngilizcede bir cümlenin olumsuz olup olmadığını “not” içerip içermediğine bakarak kolayca anlarız. Fakat Türkçe için *-me* , *-ma* gibi eklerin yanında “değil” gibi cümleyi doğrudan olumsuz yapabilecek kelimelerin (ör: Bu gün hava güzel değil) yanı sıra cümle içerisindeki sıfatlara da teker teker bakılmalıdır. Duygu durum analizini İngilizceye uygularken kullanılan en yaygın yaklaşım eklerden kurtulmaktır. Fakat bu yaklaşım Türkçenin morfolojik analiz zorluğu nedeni ile Türkçeye çok uygun görünmemektedir [1].

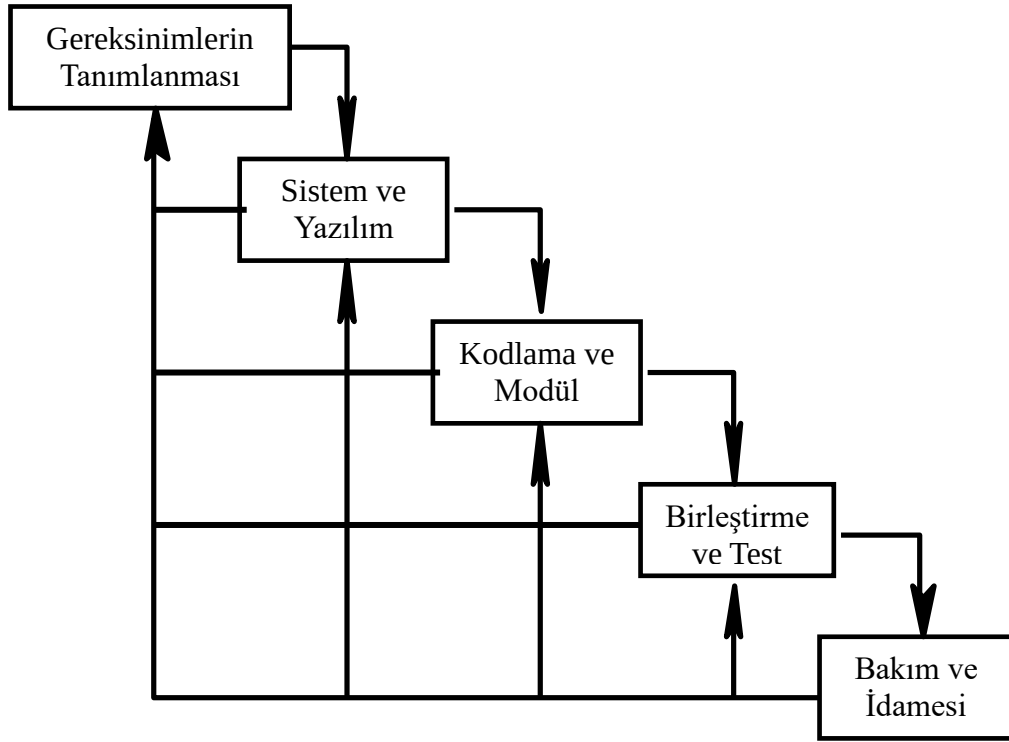
2. PROJE İŞ PLANI

2.1. Yazılım Süreç Modeli

Gerçekleştirilecek olan projede, yapacağımız işe en uygun yazılım süreç modeli olarak şelale yöntemi belirlenmiştir. Bunun nedeni şelale modelinin; yazılım geliştirme sürecinde analiz, tasarım, kodlama, test, sürüm ve bakım gibi safhalardan oluşmasıdır. Bu durum aşağıda verilen, projemize ait iş planı incelendiğinde ve şelale yönteminin söz edilen özellikleri göz önüne alındığında okuyucu tarafından daha iyi anlaşılacaktır.

Şelale metoduna ait özelliklerin tarafımızca önemli görülen birkaçından bahsedecek olursak;

- Şelalenin her basamağında yer alan aktiviteler eksiksiz olarak yerine getirilir. Bu bir sonraki basamağa geçmenin şartıdır [4].
- Yazılım süreci doğrusaldır, yani bir sonraki safhaya geçebilmek için bir önceki safhada yer alan aktivitelerin tamamlanmış olması gerekir [4].
- Kullanıcı katılımı başlangıç safhasında mümkündür. Kullanıcı gereksinimleri bu safhada tespit edilir ve detaylandırılır. Daha sonra gelen tasarım ve kodlama safhalarında müşteri ve kullanıcılar ile diyaloga girilmez [4].



Şekil 2.1 (Şelale Modeli)

Aşağıdaki şekilde proje aktivite planı (2. Derece başlıklar seviyesinde) ve zaman çizelgesi verilmiştir.

İş Kalemi Adı	Gün	Tarih
1. PLANLAMA SAFHASI	5	06.03.2017 - 11.03.2017
1.1. Proje Amacının Belirlenmesi		
1.2. Zaman Analizi		
2. GEREKSİNİM BELİRLEME (ANALİZ) SAFHASI	10	11.03.2017 - 21.03.2017
2.1. Proje Gereksinimlerinin Toplanması		
2.2. Kullanılacak Teknolojilerin Belirlenmesi		
3. TASARIM SAFHASI	15	21.03.2017 - 05.04.2017
3.1. Kullanıcı Arayüzü Tasarımı		
3.2. Veri Tabanı Oluşturulması		
4. YAZILIMIN GERÇEKLEŞTİRİLMESİ	35	05.04.2017 - 10.05.2017
4.1. Duygu Durum Analizi Yapılacak Verinin Elde Edilmesi		
4.2. Sözlük Oluşturulması		
4.3. Düzenlenmiş Veriye Uygulanacak Analiz Algoritmasının Kodlanması		
5. TEST SAFHASI	15	10.05.2017 - 25.05.2017
5.1. Modüllerin Test Edilmesi		
5.2. Kullanılan Analiz Algoritmasının Başarı Oranının Ölçülmesi		
5.3. Sistemin Stabilitate Testi		
5.4. Sistemin Hızının ve Sistemdeki Olası Gecikmelerin Ölçülmesi		
6. DESTEK SAFHASI	10	25.05.2017 - 04.06.2017
6.1. Test Sonuçlarının Değerlendirilmesi		
6.2. Önemli Göritlen Eksikliklerin ve Hataların Giderilmesi		
TOPLAM : 90 Gün		

Resim 2.2 Proje Aktivite ve Zaman Planı

Burada şelale modeline ait şekil ile bizim projemize ait proje aktivite planı birlikte incelendiğinde bahsi geçen çalışma için en uygun yazılım süreç modelinin şelale modeli olduğu anlaşılmaktadır.

2.3 Görev Dağılımı

İlgili Çalışmada, projenin detaylarına karar verme, proje geliştiriminde kullanılacak olan teknolojilerin seçimi, proje dokümantasyonu ve belgelendirmelerin hazırlanışı, veri tabanı ve ara yüzlerin tasarımı, projenin kodlanması ve proje tamamlandığında testinin yapılmasında gruptaki her birey etkin olarak rol oynayacaktır. Projenin teslim tarihine kadar projenin her aşaması, gruptaki her birey tarafından kontrol

edilecek ve uygulanacak olan yeniliklere karar verebilme yetkisine sahip olacaklardır.

2.4 Araç ve Teknikler

Projede kullanılacak olan araçlar ve teknikler tamamen Open Source (Açık Kaynak) olacaktır. Gerçekleştirilen projede Netbeans IDE 8.1 platformu üzerinden Java yazılımı kullanılması tarafımızca uygun görülmüştür. Veri tabanı olarak ise Microsoft Office uygulamalarından (Ms Excell, Access vb..) ve metin editörü olan Notepad uygulamasından faydalanılacaktır.

3. ALAN BİLGİSİ

Bilgisayar tarafından bir ifadenin ironik olup olmadığını (ironi içerip içermediğini) tespit etmek için öncelikle ilgili ifade duygu durum analizine tabii tutulmaktadır. Bu noktada duygu durum analizinin ne olduğunu kısaca açıklarsak;

Duygu durum analizi (Sentimental Analysis); metinlerde geçen duygusal ifadelerin çıkarılmasını amaçlar. En sık kullanılanı duygusal kutupsallıktır (sentimental polarity). Buna göre bir konu hakkında geçen mesajların veya yazıların olumlu veya olumsuz olmasına göre iki sınıfa ayrılması hedeflenir. Ancak duygu analizi bunun dışında, metinlerdeki ruh hali, kanaat ve daha karmaşık duyguların çıkarılması üzerinde de çalışmaktadır [1].

Türkçenin sondan eklemeli bir dil olması sonucu, eylem belirten her fiil sonunda *-me*, *-ma* gibi ekler alarak ya da almayarak cümleyi anlam olarak olumlu ya da olumsuz yapabilmektedir. Ayrıca Türkçede bir cümlenin anlam olarak olumlu, olumsuz şeklinde sınıflandırılabilmesi için cümle içinde geçen sıfatlara da (iyi, kötü, güzel, çirkin vs..) mutlaka bakılmalıdır. Cümlenin anlamını doğrudan değiştirebilecek bazı özel kelimeler aranmalı (örnek: “değil”) ve bu kelimelerde karar verme sürecine dahil edilmelidir. Ayrıca yapılan bazı çalışmalarda, ilgili dile ait bazı özel ikili, üçlü kelime gruplarının yan yana geldiklerinde cümleyi anlam olarak değiştirdikleri görülmüş ve bunlarda duygu durum analizi sırasında cümle içerisinde ayrıca aranmış ve bulunanlar değerlendirmeye alınmıştır.

3.1. Daha Önce Yapılan Duygu Durum Analizi Çalışmalarında İzlenen Çözüm Yolları

Proje konumuzun ilk kısmı olan duygu durum analizi ile ilgili yaptığımız literatür araştırmalarında problemin parçalar halinde ele alınarak aşama, aşama çözülmeye çalışıldığını görmüş bulunmaktayız. Yapılan duygu durum analizi çalışmalarında gördüğümüz kadarı ile kullanılan teknikler genel olarak sözlük tabanlı yaklaşım, makine öğrenmesi yaklaşımı ve melez yaklaşım olarak üç ana kısma ayrılmaktadır.[5]

Kullanılan teknik farklı olsa da problem çözümü için verinin toplanması, düzenlenmesi analiz edilebilir hale getirilmesi gibi bazı ortak adımları içermektedir. Makine öğrenmesi yaklaşımında, dilin özellikleri ve bilinen Makine öğrenmesi algoritmaları kullanır. Sözlük tabanlı yaklaşım, önceden hazırlanmış duygusalılık terimlerinden oluşan duygu durum sözlüğüne dayanır. Melez yaklaşımda ise her iki metodun bir araya getirilmesiyle duygu durum analizi yapılmaya çalışılır. Makine öğrenmesi yaklaşımı, duygu analizini söz dizimsel ve/veya dilsel özellikleri kullanan olağan bir metin sınıflandırma problemi olarak görür ve çözmek için Naive Bayes ve Support Vector Machine gibi sınıflandırma algoritmalarını kullanır. Sözlük tabanlı yaklaşımda ise önceden oluşturulmuş sözlükler analiz için kullanılır. Görüş sözcükleri birçok düşünce analizinde uygulanmaktadır. Pozitif görüş sözcükleri beğenilen durumlarda, negatif görüş sözcükleri ise beğenilmeyen durumlarda kullanılır.

Duygu analizi alanında yapılan çalışmalar karşılaştırıldığında, görülmüştür ki sözlük tabanlı duygu analizinde ulaşılan doğruluk oranı, makine öğrenmesi tabanlı duygu analizinde ulaşılan doğruluk oranına göre daha yüksektir. Ayrıca, duygu analizi çalışmaları sırasında doğal dil işleme (Natural Language Processing) tekniklerinin kullanımı ile belirgin biçimde iyileşme sağlandığı görülmüştür. Doğal dil işleme (NLP) dilin kurallı yapısının çözümlenerek anlaşılması veya yeniden üretilmesi amacını taşır.

Aşağıdaki tabloda duygu durum analizi ile ilgili daha önce yapılan çalışmalara ait izlenen yöntemler ve başarı oranları verilmiştir.

Kullanılan Yöntem	Veri Seti	Başarı Oranı	Sonuç
N-Gram ve Sözlük	Twitter Verisi	Belirtilmemiş	Sözlük Yöntemi Daha Başarılı
Gözetimli ML & Gözetimsiz ML (SentiStrenght)	Twitter Verisi	Belirtilmemiş	Gözetimsiz ML Daha Başarılı
Hadoop Üzerinde Naive Bayes Sınıflandırıcı	15 Günlük Gazete Manşetlerinin Altına Yazılan yorumlar	% 80	Karşılaştırma Yapılmamış
Naive Bayes, Maksimum Çaba ve SVM	Politik Haberler Altına Yazılan yorumlar	% 66 - % 77	Karşılaştırma Yapılmamış
Naive Bayes ve Karar Ağacı	6000 Adet Tweet	% 63	Karşılaştırma Yapılmamış
Support Vector Machine	1021 Adet Sinema ve Eğitim Konulu Makale	Belirtilmemiş	Duygu Durum Makale Başlığına Bağlıdır

3.1 Daha Önce Yapılan Çalışmalar ve Başarı Oranları

3.2. Daha Önce Yapılan İroni Tespiti Çalışmalarında İzlenen Çözüm Yolları

Duyarlılık sınıflandırması sadece üç farklı kategoride sonuçlanır; pozitif, negatif ve nötr. Bununla birlikte, bazı cümleler pozitif olarak sınıflandırılmış olsa da, ironiktirler ve esasen olumsuz bir anlam taşımaktadırlar. Bunun tam tersi durumlar da benzer şekilde ortaya çıkabilmektedir (Negatif olarak sınıflandırılan, ancak gerçekte olumlu bir anlam taşıyan cümleler).

Bu tür durumların tespiti için bazı yöntemler uygulanır;

Cümledeki ironi varlığının tespiti için, bir ipucu olabilecek özellikler çıkarılır. Bu özelliklerden bazıları Türkçeye özgüdür, bazıları ise tüm dilleri kapsayacak şekilde ortak kullanılan noktalama işaretleri gibidirler. Bu özelliklerin tanımları aşağıda detaylı bir şekilde açıklanmaktadır.

3.2.1. Parantez içinde soru işareti / ünlem işareti

Parantez içlerinde kullanılan soru işaretleri ve ünlem işaretleri, ironiyi yazılı bir söylemle belirtmenin en yaygın yoludur. Basılı yayımların yanı sıra (yazar ve

gazeteciler gibi...), sosyal medya kullanıcıları da kendi cümlelerinde alay konusu yapmak için bu kalıbı sıklıkla kullanmaktadırlar.

Buna Türkçe için bir örnek olarak

“Evimin önüne ne kadar da güzel (!) bir çukur kazmışlar.”

Cümlesini verebiliriz.

Bu cümle, "güzel" sıfatıyla pozitif etiketine sahiptir. Bununla birlikte, parantez içindeki ünlem desenine sahip olması, anlamı tersine çevirir. Yazarın, etkinlik hakkında zevk almadığını, aksine, yazarın öfkeli olduğunu anlarız.

Bu durum Marcellin Jobard ve Fransız şair Alcanter de Brahm tarafından 19. yüzyılda önerilmiştir. Ortaya atılan öneride, desenin varlığı araştırılır ve desenler toplamı bir özellik olarak kullanır. [18]

3.2.2. Ünlem işareti

Aslına bakarsak Cümlede bir ünlem işareti bulunması cümlemin ironik olduğuna işaret etmemektedir. Normalde, ünlem işareti cümledeki bir haykırış ya da nida sonrasında cümledeki güçlü duyguları belirtmek altını çizmek amacı ile kullanılır. Bununla birlikte, tekrarlanan ünlem işaretlerine sahip olmak da önemli ipuçlarından biridir.

Örneğin: Evimin önüne ne kadar da güzel bir çukur kazmışlar !!!!

Benzer şekilde, bu cümle "güzel" sıfatı nedeniyle pozitif etikete sahiptir; Tekrarlanan ünlem işaretleri cümlemin anlamını değiştirir.

3.2.3. Soru işareti

Burada da aslında cümle içerisinde bir soru işareti olması ilgili cümlemin ironik olduğuna işaret etmemektedir. Normalde, soru işaretleri, sorgulayıcı bir cümle, madde veya ifade belirtmek için kullanılır. Ancak, bizim için tekrar eden soru işaretlerine sahip olmak ironiyi tespit etmek için bir ipucudur.

Örneğin: Evimin önüne ne kadar da güzel bir çukur kazmışlar ????

Bu cümleyi yazan kişinin niyeti, görünüşte bir soru sormak değildir. Cümleye uygulanan duygu durum analizi sonucu pozitif etiket almasına neden olacak olan "güzel" sıfatından dolayı, tekrarlanan soru işaretleri Cümlenin anlamını değiştirecektir.

Ayrıca, Soru ve ünlem işaretlerine sahip olmak da (!?!?) alay konusu (ironi) için bir işarettir. Eğer ki bu durum değerlendirilecek ise bu noktalama işaretlerinin sayısı ayrı ayrı hesaplanmalı ve bu desenin yeni bir özellik olarak kullanılmamasına dikkat edilmelidir.

3.2.4. Tırnak işareti

Tırnak işaretleri genellikle doğrudan bir konuşmaya veya başlığa işaret ederek bir isim cümlesi içerisinde yer almaktadır. Ayrıca, bir kelimenin veya cümlenin farklı bir anlamını belirtmek için kullanılırlar ve genellikle tırnak işaretleri arasındaki kelime veya cümle pozitif kutuplu olduğunda ironi ifade etmek için kullanılmaktadır.

Örnek: Evimin önüne “ne kadar da güzel” bir çukur kazmışlar.

Tırnak işaretleri, bu cümleye ironi enjekte etmek için, yani tırnak işaretleri arasında kalan kısmın cümlenin anlamını ortadan kaldırması için kullanılır. Bu tür kullanım "Scare Quotes" olarak bilinir. [19] Geliştirilen uygulama, tırnak işareti var mı diye bakar ve cümledeki tırnak işareti sayısını bir özellik olarak kullanır.

3.2.5. Cümle sonundaki nokta (tam durak)

Burada full stop ile anlatılmak istenen cümlenin sonuna gelindiğini belirtmek amacı ile kullanılan nokta (‘.’) İşaretidir. Cümle sonlarını belirtmek amacı ile kullanılan düzenli nokta (‘.’) işaretleri ilgili cümlenin ironi içermesi bakımında herhangi bir ipucu vermez. Burada belirtmeliyiz ki; Üç nokta olarak adlandırılan tam durakların

özel bir kullanımı vardır. Bunlar “three periods of full stop” olarak adlandırılır. Normal kullanımının yanı sıra, tamamlanmamış bir düşünce, sinir veya garip bir sessizlik olduğunu gösterir. Dolayısıyla, tekrarlanan tam duraklara sahip olmak bazen alaycılığı ifade etmek için kullanılabilir.

Örnek: Bu inanılmaz destek için gerçekten çok teşekkür ederiz!...

Yukarıdaki örneğe ilk bakışta, cümlelerin pozitif kutupları var gibi görünmektedir. Yazar sanki zevk ve keyif hissediyor gibi düşündürmektedir. Bununla birlikte, cümlelerin sonunda tam durak bulundurulması, yazarın daha çok sitem duygusu içerisinde olduğunu belirtmektedir.

Sonuç olarak geliştirilen uygulama, cümle içerisinde tam durak olup olmadığını araştırarak cümledeki tam durak sayısını bir özellik olarak kullanabilir.

3.2.6. Yüz ifadeleri

Yüz ifadelerine örnek olarak mutlu yüzlerin noktalama işaretleriyle resimsel temsilini verebiliriz. Yaygın kullanılan mutlu yüzler şunlardır:

“ :), :-), :)) “

Mutlu yüzlerin varlığı daima Pozitif anlam ifade eder. Bazen gülünçlük yapmak için kullanılırlar.

Örnek: Bu hashtag de ne kadar çok tweet almış :)) :)) :))

Bu cümlede, yazar cümledeki ironiyi belirtmek için suratlar (gülen yüzler) dizisini kullanmıştır. Aslında yazar durumdan mutlu değildir ve durumla alay etmektedir.

Sonuç olarak geliştirilen uygulama, cümle içerisindeki yüz ifadelerini bulup cümledeki yüz ifadeleri sayısını bir özellik olarak kullanabilir.

Ayrıca yüz ifadelerine ait bir liste aşağıdaki tabloda daha ayrıntılı olarak görülmektedir.

%-)	:-D
(-:	:-P
(:	:-}
(^ ^)	:3
(^_ ^)	:9
(^.^)	:d
(^_ ^)	:D
(o:	:P
\o/	:P
0:)	:X
8)	:l
:)	:b)
:))	:o)
:-)	:p
:))	:^
:-*	;^)
:-*	<3
:-d	=)

3.2 Mutlu yüz ifadeleri

Ayrıca diğer bazı yüz ifadeleri de üzgün ve öfkeli duyguların belirtilmesi amacıyla kullanılmaktadır. En yaygın üzgün emoji şunlardır; (, :-(ve tek başlarına ironi için bir ipucu değildirler. Pozitif kutuplu bir cümle ile kullanıldıklarında, ironiyi göstermek için bir ipucu haline gelirler.

Ör. Okullar açılıyor, mükemmel ya :(

Cümledeki emoji ifadeleri kaldırıldığında olumlu bir anlam taşır. Yazar, cümledeki ironiyi ifade etmek için üzgün yüz ifadesi kullanıyor.

Üzgün ve öfkeli duygu durumu belirtmek amacıyla yaygın kullanılan yüz ifadeleri aşağıdaki tabloda görülmektedir.

%-(	:-\
):-	:-
):	:/
)o:	:E
:#	:F
:'(	:O
:'-('	:S
:(	:[
:((	:[
:,(	:_(
:-&	:o(
:-('	:s
:-('o)	:
:-/	:*
:-S	=(
>:(	=[
>[	>/

Tablo 3.3 üzgün ve öfkeli yüz ifadeleri

3.2.7. Cümle içindeki duygusal kutuplaşma

Duyarlılık analizi bu çalışmada iki açıdan kullanılmaktadır. Öncelikle, cümlelerin kutupsallığı diğer özellikler için tamamlayıcı bir bilgi olarak kullanılır. Olumlu veya olumsuz olmak bazı özellikler için önemlidir. İkincisi, pozitif ve negatif duyarlılık puanları arasındaki boşluk, dengesizliği göstermektedir. Karşıt kutupluğa sahip kelimeleri birlikte kullanmak ironi için bir ipucudur.

Ör. Kötü oynuyor diye istenmeyen futbolcu UEFA tarafından yılın en iyi oyuncusu seçilmiş.

Bu örnekte görüldüğü üzere cümlelerin ilk bölümünde negatif yönde kelimeler varken cümlelerin sonunda pozitif kelimeler yer almakta. Bu duygusal kutuplaşma ironi tespiti için oldukça geçerli bir ipucudur.

3.2.8. İsimlerde küçültme

Genellikle küçültme ekleri nesnenin küçüklüğünü göstermek, samimiyeti ifade etmek ve iltifat etmek için kullanılır. Küçümsemeler, ironik bir perspektife baktıklarında, temsil ettikleri varlığa yönelik bir hakaret veya amortisman(yıpratma) belirtmek için kullanılabilir.

Ör. Çok acıktım çünkü bugün sadece bir kilocuk et yedim.

-*cuk* eki kilo kelimesini hafifletmek ve ironi yapmak için kullanılmıştır.

3.2.9. Ünlem kelimelerinin kullanımı

Açıklamalar, bir duygu veya sürprizi, tiksinti, sevinç, heyecan veya coşku gibi bir hissi iletme için kullanılır. Bir cümlede çıkarım yapmak, ironiyi tespit etmek için potansiyel bir ipucudur. Çoğu zaman, bir ünlem sözcüğü kullanıldığında cümlelerin sonunda bir ünlem işareti kullanılır. Ancak içinde ünlem kelimelerinden herhangi birinin olmadığı cümlelerin sonunda ünlem işareti kullanılması bize ironi tespiti için bir ipucu olabilir.

Ör. Kapımın önüne güzel bir çukur kazmışlar!

Bu cümle duygu durum analizi açısından pozitif bir cümle gibi durmakta ve içinde herhangi bir ünlem kelimesi geçmemesine rağmen sonunda ünlem işareti ile bittiği için bize ironi olduğunu gösteriyor.

3.2.10. Büyük harf kullanımı

Türkçe'de, özel isimlerin ilk harfi ve bir cümlenin ilk kelimesi büyük harflerle yazılır. Birisi bir cümle içinde bir kelimeyi veya cümleyi vurgulamaya çalışırken, o kelime veya cümlenin tamamını büyük harflerle yazabilir. Bu tür dilbilgisinin kötüye kullanımını bir avantaja dönüştürürsek, bu bize ironi hakkında bir ipucu verebilir.

Ör. O kadar ZEKİ ki bütün sınıfları çift dikiş gidiyor.

Yazılı söylemde vurguyu ifade etmenin yollarından biri de budur.

Bu cümledeki ironi bu özellikten anlaşılabilir

4. GELİŞTİRİLEN UYGULAMA

Yaptığımız literatür taraması ve araştırmalar sonucu sözlük tabanlı bir uygulama ile probleme çözüm getirmeyi hedefledik. Alan bilgisi başlığı altında bahsi geçen tüm bilgiler ışığında düşünerek sahip olduğumuz problemi daha önce yapılan çalışmalara benzer şekilde alt parçalar halinde ele aldık. Her bir aşamada probleme yaklaşım ve uygulama bazında getirilen çözümler ana başlıklar halinde aşağıdaki gibidir.

- Veri Toplama
- Sözlük Oluşturma
- Analiz Algoritmasının Geliştirilmesi
- Uygulamanın Testi
- Test Sonuçlarının Değerlendirilmesi

Uygulama gerçekleştirilirken bu adımların her birinde neler yapıldığı aşağıda ilgili başlıklar altında detaylı olarak açıklanmaktadır.

4.1. Veri Toplama Aşaması

Yaptığımız araştırmalar bize göstermiştir ki; metin madenciliği uygulamalarında analiz edilecek olan verinin, twitter gibi sosyal medya uygulamaları üzerinden toplanması oldukça yaygın kullanılan bir veri toplama yöntemidir. Özellikle duygu durum analizi çalışmalarında (güz dönemi itibarı ile projemizin birinci kısmında bizimde tercih ettiğimiz gibi) oldukça sık rastlanan ve geçerliliği olan bir durumdur. Fakat ironi tespiti için düşünüldüğünde özellikle Türkçe için sosyal medya paylaşımlarının veri olarak alınması bu çalışmanın giriş bölümünde de belirttiğimiz üzere; bahsi geçen konudan haberdar olmayı gerektirecek ve dolayısı ile daha çok durumsal ironi sınıfına girecektir.

Çalışmanın bu kısmında, veri toplama aşaması, web kaynaklarından ve daha çok YÖK Ulusal Tez Merkezi aracılığı ile Türkiye'deki edebiyat ve dil fakültelerinde yapılan ironi içerikli çalışmalar incelenerek, ironik cümleler elde edilmesi esasına dayanmaktadır. [20]

4.2. Sözlük Oluşturma Aşaması

İlgili uygulama için bir cümleyi olumlu ya da olumsuz anlamlı yapabilecek kriterler, daha önce yapılan çalışmalar incelenerek, Türkçe dilbilgisi (yapım ve çekim eklerinin anlam bakımından cümleye etkileri, anlamına göre yüklemelerin olumluluk ve olumsuzluk durumları vb...) araştırmaları yapılarak Türkçe cümlelerin anlam bakımından incelenmesinde bize fayda sağlayacağına inandığımız Türkçe yüklem, zarf ve sıfatların geniş bir veri tabanı oluşturularak kaydedilmesine karar verilmiştir. Bunun sonucunda 7000 adet Türkçe fiil içeren bir Türkçe fiiller kümesi oluşturulmuştur. Ayrıca yaptığımız araştırmalar sırasında rastladığımız, Yıldız Teknik Üniversitesi Doğal dil işleme Çalışma Grubu (KEMİK) tarafından yapılan bir çalışmada, toplam iki milyon Türkçe tweet incelenerek n-gram yöntemi ile frekans sıklığına bakılmış, en çok geçen dörtyüz kelime belirlenmiştir. Veri tabanımızın sözlük, ontoloji kısmı, bahsi geçen 400 kelimenin yanı sıra yaklaşık 800 adet olumlu ve olumsuz sıfat içermektedir.[7] Bunlara ek olarak alan bilgisi bölümünde (Bölüm 3) bahsettiğimiz, yine yıldız teknik üniversitesinde yapılan bir çalışmada rastladığımız, 100 adet ikili kelime grubu da proje veri tabanına eklenmiştir.

4.3. Analiz Algoritmasının Geliştirilmesi

Bu aşamada uygulamaya girdi olarak gönderilen verinin ironi içerip içermemesi bakımından incelenerek ironik ya da ironik değil şeklinde bir karara varılması amaçlanmıştır.

Bir cümlede yüklem olumlu ya da olumsuz olduğunu belirlemek amacı ile yüklemdeki çekim ekleri atılır. Olumsuzluk eki (-me, -ma) kökün hemen ardına geldiğinden elimizde kalan son fiil, fiilin olumsuz hali ya da fiilin kökü olmaktadır. Bunun sonucunda fiil olumlu ya da olumsuz olarak (+ ya da -) işaretlenir. Burada gereklilik kipi olan -meli, -malı için özel bir durum oluşturma ihtiyacı doğmuştur. Fiil köke indirgenerek yapılan taramada fiile kökten sonra olumsuzluk eki (-me, -ma) gelmiş ise bir sonraki hece incelenerek -lı, -li ekleri aranmıştır. Bu ekler tespit

edildiğinde bunun gereklilik kipi olduğu anlaşıp, olumsuz değerlendirmesi yapılmamıştır.

Fiiller incelenerek olumlu ve olumsuz şeklinde 2 kategoriye ayrıldıktan sonra, içerisinden fiiller atılan cümle, yaklaşık 1300 kelimeden oluşan sözlük içeriği ile karşılaştırılmış ve eşleşen kelimeler ilgili cümle için sahip oldukları etikete bakılarak olumlu, olumsuz (+ , -) şeklinde 2 farklı kategoriye ayrılmıştır.

Son aşamada cümle içindeki fiillerden gelen olumlu, olumsuz (+ , -) değeri ile, cümlede bulunan diğer kelimelerden elimizdeki sözlüğümüz ile eşleşenlerden gelen olumlu, olumsuz değerleri çarpılarak cümlenin duygu durum analiz sonucuna karar verilmiştir.

Buradaki çarpma işlemini; kabaca fiilden gelen ve diğer kelimelerden (sıfatlar ve zarflar) gelen değerlerin pozitif ve negatif olarak matematiksel çarpımı gibi düşünebiliriz.

Esas kısım olan cümledeki ironi varlığının tespiti için yapılan işlem ise; ilgili cümlelerin parçalara (alt cümleciklere) ayrılarak parçalanması ile elde edilen her bir cümlecik için yukarıda anlatılan algoritmanın çalıştırılmasıdır.

İroni tespiti gerçekleştirilirken, alan bilgisi bölümünde açıkladığımız tüm özellik çıkarımlarından, tarafımızca önemli görülen iki temel özellik olan “cümle içindeki duygusal kutuplaşma” ve “ünlem kelimelerinin kullanımı” özellikleri cümle içerisinde aranarak ironi tespiti yapılmaya çalışılmıştır.

Burada geliştirilen program kendisine girdi verisi olarak gelen cümleyi alır. En başından başlayarak cümle içerisindeki fiil ve fiilimsilere kadar okur. Bulduğu her fiilin ya da fiilimsinin öncesindeki fiile kadarki kısmı, bir alt cümle olarak kabul eder ve elde ettiği her bir alt cümle için olumlu - olumsuz ayrımı yapar. Ve daha sonra öncelikle alt cümleciklerden gelen duygu geçişlerini göz önünde bulundurarak (ör: olumlu, olumsuz, olumlu gibi) duygusal kutuplaşma olup olmadığına bakar. Sonrasında cümlede duygusal kutuplaşma tespit edilemez ise ünlem cümlelerinin kullanımı özelliğini tarar. Tüm bu işlemler sonrasında ilk başta verilen tam cümlelerin ironi içerip içermediğine karar verir.

Bu son kısmı bir örnek ile açıklayacak olursak;

Örnek cümlemiz ;

“Kötü oynuyor diye istenmeyen futbolcu, UEFA tarafından yılın en iyi oyuncusu seçilmiş.” Olsun

Burada geliştirilen uygulamanın ne şekilde çalıştığını anlatalım;

Öncelikle fiil ve fiilimsilere göre ilgili cümle alt cümleciklere parçalanır. Şöyleki;

1- “Kötü oynuyor”

Burada kötü (-) ve oynuyor (+) değere sahip olduğundan kötü oynuyor cümlesi tam bir cümle olarak olumsuz etikete sahip olacaktır.

2- “Diye istenmeyen futbolcu”

Burada istenmeyen fiilimsisinden dolayı bu cümle olumsuz etikete sahip olacaktır.

3- “UEFA tarafından yılın en iyi oyuncusu seçilmiş”

Burada ise program iyi sıfatını ve seçilmiş fiilini olumlu olarak algılayıp cümleye olumlu etiketi verecektir.

Daha sonra alt cümleciklere ait elde edilen etiketler (sırası ile; olumsuz, olumsuz, olumlu, “-, -, +”) değerlendirilerek ilgili cümle için ironi içerir sonucuna ulaşılabacaktır.

4.4. Uygulamanın Test Aşaması

Bu kısmın ana amacı geliştirilen uygulamanın ironi tespiti sonuçlarının doğruluk bakımından istatistiksel olarak değerlendirilmesidir. Ayrıca kullanıcı modülünün testi gibi genel sistem değerlendirmesi amaçlarını da taşımaktadır.

4.5. Test Sonuçlarının Değerlendirilmesi

Geliştirilen uygulama test aşamasında yapılan işlemler ile, sistemin hem kullanıcı modülleri (veri elde etmede ve analiz etmede kullanılan butonlar), hem de asıl önemli kısım olan analiz algoritmasının başarı oranları bakımından değerlendirmeye alınmıştır. Sonuç olarak sistemde yeniden düzenlenmesi gereken ciddi performans kayıpları ya da gecikmelere rastlanmamış, kullanıcı modüllerinin sağlıklı şekilde çalıştığı görülmüştür. Tarafımızca uzun çabalar sonucu geliştirilen ve asıl önemli kısım olduğuna inandığımız ilgili duygu durum analizi ve ironi tespiti algoritmasının başarı ve yanlışma oranları ise istatistiksel olarak sayılar ve yüzdelik oranlar ile net bir şekilde ortaya konmuştur.

Program test sonuçları istatistiksel olarak aşağıdaki gibidir:

Test edilen 50 ironi içeren cümleden; 15 tanesi ironi, 35 tane ironi değil olarak etiketlenmiştir.

Test edilen 100 ironi bulunmayan cümleden ise; 69 tanesi ironi değil, 31 tanesi ironi olarak etiketlenmiştir.

Positive (P): İroni içeren cümleler

Negative (N): İroni içermeyen cümleler

True Positive (TP): İroni içeren cümlelerin doğru bulunması

False Negative (FN): İroni içermeyen cümlelerin yanlış bulunması

True Negative (TN): İroni içermeyen cümlelerin doğru bulunması

False Positive (FP): İroni içeren cümlelerin yanlış bulunması

$$\text{Precision : } \frac{TP}{TP+FP} = \%48.38$$

$$\text{Recall: } \frac{TP}{TP+FN} = \%30$$

$$\text{F1-score : } \frac{2\text{Precision*Recall}}{\text{Precision+Recall}} = \%37$$

Overall Accuracy : %49,5

5. SONUÇ

İnsanlar tarafından internet ortamında yapılan fikir alışverişi ve paylaşımların duygu içeriklerinin anlaşılabilmesi günümüzde medya, siyasi kurumlar, büyük ölçekli marka ve firmalar gibi birçok alandan farklı kesimlerin oldukça ilgisini çekmektedir. Kişi ve kuruluşlar toplumun genel olarak kendileri hakkındaki duygu ve düşüncelerini merak etmekte, bu konuda aldıkları geribildirimler doğrultusunda hedef ve planlarını değiştirmektedir.

Metinler üzerinde duygu durum ve ironi analizi gerçekleştirmenin farklı metodolojileri, yaptığımız araştırmalar ile incelenmiş ve sonunda probleme sözlük tabanlı yaklaşım yöntemi ile çözüm aramak tarafımızca uygun görülmüştür. Gerçekleştirilen çalışmada, Türkçe dilindeki 7000 adet fiil ve 2 milyon tweet taranarak oluşturulmuş 400 kelimelik bir sözlük ile birlikte 800 kelimelik olumlu olumsuz etiketine sahip türkçe sıfatın yanı sıra 100 adet ikili kelime grubu birlikte kullanılmıştır. Türkçe dili için yaptığımız çalışmada, bu kadar büyük bir veri tabanı ile çalışmanın avantajı, % 49,5 başarı skoru ile görülmüştür.

Bugüne kadar yapılan çalışmalar ile tarafımızdan gerçekleştirilen çalışma karşılaştırmalı düşünüldüğünde, hali hazırda var olan hiçbir algoritma kullanılmamış olması çalışmanın esas kısmının tamamen kendine özgü olduğu gerçeğini ortaya koymaktadır. Sonuç olarak yaptığımız dilbilgisi araştırmalarından edindiğimiz bilgiler ışığında, çalışmanın esas kısmı olan karar algoritmasının düşünsel boyutu üzerinde zaman ve emek harcayarak deneme yanılma yöntemiyle % 49,5 gibi önemi görmezden gelinemeyecek bir başarı skoru elde etmiş bulunmaktayız.

Gerçekleştirilen çalışmanın eksik yönlerini ele alacak olursak; Veri tabanında kayıtlı olan 7000 adet Türkçe fiilin, veri tabanına binary search tree (ikili arama ağacı) olarak kayıt edilmesi, analiz edilecek veride yapılan fiil taraması işleminde uygulamaya kayda değer ölçülerde hız katacaktır. Var olan veri tabanının (filler ve sözlük) daha büyük boyutlara genişletilmesi, uygulama ile daha yüksek başarı skoru elde edilmesini sağlayacaktır. Ayrıca ironi tespiti algoritmasının üzerinde çalışılarak daha fazla geliştirilmesi, özellikle anlaşılması zor ve karmaşık cümlelerde daha başarılı sonuçlar alınmasını sağlayacaktır. Son olarak bölüm 3.2 de detaylı bir

biçimde açıklanan özelliklerden, ironi tespiti için cümle içerisinde aranacak olanların önemli görülen bir veya birkaçının daha uygulama analiz algoritmasına eklenmesi de uygulamanın başarı skorunu artıracaktır.

6. KAYNAKÇA

- [1] Taşlıoğlu, Hande (Department of Computer, Engineering Middle East Technical University, September 2014)
- [2] "SentiStrength", [internet]. Erişim: sentistrength.wlv.ac.uk/. [Ekim 2016].
- [3] A. Alphan Arslan, Türkçe Metinlerde Anlamsal Bilgi Çıkarımı İçin Bir Veri Madenciliği Uygulaması, Başkent Üniversitesi Fen Bilimleri Enstitüsü (2011)
- [4] "Şelale Yazılım Süreç Modeli", [internet]. Erişim: https://tr.wikipedia.org/wiki/Waterfall_model [Ekim 2016]
- [5] Medhat, Walaa and Hassan, Ahmed and Korashy, Hoda, Sentiment analysis algorithms and applications: A survey, Ain Shams Engineering Journal,1093--1113,2014
- [6] "twitter via api ve OAuth Kütüphanesi" [İnternet]. Erişim: <https://dev.twitter.com/oauth> [Kasım 2016]
- [7] "2 milyon tweet incelemesi sonucu oluşturulan 400 kelimelik sözlük" [internet]. Erişim :<http://www.kemik.yildiz.edu.tr/?id=28> [Kasım 2016]
- [8] Kaya, Mesut and Fidan, Guven and Toroslu, Ismail H, Sentiment analysis of turkish political news, Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01, 174--180 , 2012

- [9] Vural, A Gural and Cambazoglu, B Barla and Senkul, Pinar and Tokgoz, Z Ozge, A framework for sentiment analysis in Turkish: Application to polarity detection of movie reviews in Turkish, Computer and Information Sciences III,2013
- [10] Akgül, Eyüp Sercan and Ertano, Caner and Banu, DİRİ, Twitter verileri ile duygu analizi, Yıldız Teknik Üniversitesi Elektrik elektronik Fakültesi Bilgisayar Mühendisliği Bölümü
- [11] Salur, Mehmet Umut, Büyük araçlarından hadoop kullanarak veri madenciliği,Pamukkale Üniversitesi Fen Bilimleri Enstitüsü,2016
- [12] Amasyalı, M Fatih and others, Active learning for Turkish sentiment analysis, Innovations in Intelligent Systems and Applications (INISTA), 2013 IEEE International Symposium on, 1--4 , 2013
- [13] Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, Isabell M. Welp (2010) Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment.
- [14] Finn Arup Nielsen DTU Informatics, Technical University of Denmark, Lyngby, Denmark. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs

- [15] Songbo Tan * , Jin Zhang (2007). An empirical study of sentiment analysis for chinese documents.
- [16] Ray Chen, Marius Lazer (2012). Sentiment Analysis of Twitter Feeds for the Prediction of Stock Market Movement.
- [17] Gantt Şeması ve CMP Teknikleri ile Proje yönetimi, Planlaması ve Uygulaması Doc. Dr. Birdoğan Karadeniz Teknik Üniversitesi İktisadi ve İdari Bilimler Fakültesi İşletme Bölümü
- [18] S. B. Imandoust and M. Bolandraftar, "Application of K-Nearest Neighbor (KNN)
- [19] ” Scare Quotes”, [internet]. Erişim: https://en.wikipedia.org/wiki/Scare_quotes. [Nisan 2017]
- [20] Taş Cem, Aziz Nesin “Yaşar ne yaşar ne yaşamaz” Adlı eserinde mizah hiciv ve ironi incelemesi, T.C. Uludağ Üniversitesi Sosyal Bilimler Enstitüsü Yabancı Diller Eğitim Anabilim dalı Alman Dili Eğitimi Bilim dalı, Bursa 2007