



F KLAVYE İÇİN METİN ANALİZİ TABANLI KELİME TAMAMLAMA SİSTEMİ

Bergen KARABULUT*, Halil Murat ÜNVER

Kırıkkale Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Kırıkkale, Türkiye

Anahtar Kelimeler

*Alternatif Teknoloji,
Kelime Tamamlama,
Kelime Tahmin,
Tuşlama Tasarrufu,
F Klavye.*

Öz

Kelime tamamlama, kullanıcının basması gereken tuş sayısını azaltmayı hedefleyen bir yardımcı teknoloji aracıdır. İlk çalışmalarda, genellikle bir alternatif ve destekleyici iletişim aracı olarak ele alınmıştır. Fakat son dönemlerde kelime tamamlama sistemleri; mobil cihazlar, makine çevirisi ve arama motorları gibi farklı alanlarda da önem kazanmıştır. Bu çalışmada, Standart Türkçe Klavye olan F klavye için metin analizi tabanlı bir kelime tamamlama sistemi geliştirilmiştir. Geliştirilen sistemi test etmek için zabıt kâtiplerinin yoğun klavye kullanımından dolayı Adalet Bakanlığı tarafından yapılan zabıt kâtipliği sınavlarına ait 160 metin kullanılmıştır. Bu metinlerdeki kelimelerin *unigram* ve *bigram* frekansları çıkarılmış ve bir veritabanında tutulmuştur. Sistem, bu veritabanını kullanarak yazma işlemi esnasında kullanıcıya 8 adet alternatif kelimedenden oluşan bir tahmin listesi sunmaktadır. Tanımlanan tuş kombinasyonları yardımıyla kullanıcı tahmin listesinden bir kelime seçebilmekte ve seçilen kelime sistem tarafından otomatik olarak tamamlanmaktadır. Geliştirilen sistemin performansı tuşlama tasarrufu açısından değerlendirilmiştir. Tüm metinler arasından rastgele seçilen 15 metin ile yapılan test işleminde ortalama %50,98 tuşlama tasarrufu sağlanmıştır.

TEXT ANALYSIS BASED WORD COMPLETION SYSTEM FOR F KEYBOARD

Keywords

*Alternative Technology,
Word Completion,
Word Prediction,
Keystroke Saving,
F Keyboard.*

Abstract

Word completion is an assistive technology tool that aims to reduce the number of keys that the user must press. In early studies, it generally handled as an augmentative and alternative communication tool. However, recently word completion systems have also gained importance in different fields such as mobile devices, machine translation and search engines. In this study, a word completion system based on text analysis was developed for F keyboard, Standard Turkish Keyboard. 160 texts of clerk recruitment exams performed by the Ministry of Justice were used to test the developed system due to the intensive keyboard use of the court-house employees. Words' *unigram* and *bigram* frequencies of these texts were extracted and stored in a database. By using this database, the system presents a prediction list of 8 alternative words to the user during the writing process. With the help of defined key combinations, the user can select a word from the prediction list and the selected word is completed automatically by the system. The performance of the system was evaluated in terms of keystroke saving. 50.98% of average keystroke saving rate was provided at the test process done with 15 randomly selected texts among the all texts.

Alıntı / Cite

Karabulut, B., Ünver, H. M., (2020). F Klavye için Metin Analizi Tabanlı Kelime Tamamlama Sistemi, Mühendislik Bilimleri ve Tasarım Dergisi, 8(1), 262-272.

Yazar Kimliği / Author ID (ORCID Number)

B. Karabulut, 0000-0003-0755-1289
H. M. Ünver, 0000-0001-9959-8425

Makale Süreci / Article Process

Başvuru Tarihi / Submission Date	09.03.2018
Revizyon Tarihi / Revision Date	18.06.2018
Kabul Tarihi / Accepted Date	27.11.2019
Yayın Tarihi / Published Date	20.03.2020

* İlgili yazar / Corresponding author: brgnkarabulut@gmail.com

1. Giriş (Introduction)

Kelime tamamlama, kullanıcının tuşlamaya başladığı kelimeyi olabildiğince doğru şekilde tahmin ederek kelimenin doğrudan yazılmasını sağlama işidir. Kullanıcı bir kelimeyi tuşlamaya başladığında veya hiç tuşlamadan, ekranda kelime tahminleri listelenir. Eğer doğru kelime tahmin edilirse, kullanıcının seçimi ile bu kelime doğrudan yazılır. Bu şekilde kelimenin daha az tuşlama ile yazılması sağlanır. Ancak, doğru kelime tahmin edilemezse kullanıcı tüm kelimeyi kendi yazarak tamamlar (Fazly ve Graeme, 2003; Li ve Graeme, 2005).

Kelime tamamlama ve kelime tahmin sistemleri temel olarak fiziksel engelli kişilerin yazma hızını artırmak için geliştirilmiştir (Tam ve Wells, 2009). Bu nedenle, 1980-2000 yılları arasında yapılan çalışmalarda kelime tamamlama, genellikle bir alternatif ve destekleyici iletişim aracı olarak ele alınmıştır. Ancak, 2000'li yıllardan sonra mobil cihazlarda kullanımıyla yeni bir alanda yer almaya başlamıştır. Bu günkü konumuyla kelime tamamlama birçok farklı alanda yer almaktadır. Alternatif ve destekleyici iletişim sistemleri, cep telefonları, arama motorları ve kısa mesaj servisleri için önem arz etmektedir (HaCohen-Kerner ve Greenfield, 2012). Literatürdeki çalışmalar dikkate alınarak, kelime tamamlama ve tahmin sistemlerinin kullanım alanları, aşağıda belirtildiği gibi sıralanabilir.

- Alternatif ve Destekleyici İletişim (Li ve Graeme, 2005)
- Cep telefonu/Mobil Cihazlar (Van den Bosch ve Bogers, 2008)
- Kısa Mesaj Servisi(SMS) (Karunathne vd., 2013)
- Makine Çevirisi (Foster vd., 1996)
- Kişisel cep bilgisayarı(PDA) (Tam ve Wells, 2009)
- Ekran Klavyesi (Anson vd., 2006)
- Arama Motorları (HaCohen-Kerner ve Greenfield, 2012)

Geliştirilen kelime tamamlama veya kelime tahmin sistemleri değerlendirilirken farklı ölçütler kullanılmıştır. Bu ölçütler; isabet oranı (HR, Hit Rate), tuşlama tasarrufu (KS, Keystroke Saving), doğruluk (Acc, Accuracy) ve tamamlamaya kadar tuşlama (KuC, Keystroke Until Completion) olarak belirtilmiştir (Fazly ve Graeme, 2003). Ancak, kelime tahmin ve kelime tamamlama sistemlerinin performansının değerlendirilmesinde temel değerlendirme ölçütü olarak tuşlama tasarrufu kullanılmaktadır (Trnka, 2010). Kısaca KS olarak ifade edilen tuşlama tasarrufu, kullanıcının kelime tamamlama kullanarak yazarken tuş basımından elde ettiği tasarruf oranını vermektedir.

Tuşlama tasarruf oranı, Eşitlik(1) yardımıyla hesaplanmaktadır (Wandmacher ve Antoine, 2008).

$$ksr_n = \left(1 - \frac{k_p}{k_a}\right) \cdot 100 \quad (1)$$

Formülde kullanılan terimler: ksr_n : Tuşlama tasarruf oranı, n : Tahmin listesinin uzunluğu, k_p : Kelime tahmin kullanarak yazarken basılan toplam tuş sayısı, k_a : Kelime tahmin olmadan yazarken basılan toplam tuş sayısı.

KS değeri arttıkça sistemin verimliliği de artmaktadır. Bu sebeple, yapılan çalışmalar genellikle bu değeri arttırmaya yönelik olmaktadır. Bunun dışında kelime tamamlama ve tahmin sistemlerinin gerçek kullanıcılar ile farklı aygıtlarda kullanımının katkılarını değerlendiren çalışmalar da bulunmaktadır (Tam ve Wells, 2009; Anson vd., 2006; Handley-More vd., 2003). Anson vd. (2006), yaptıkları çalışmada ekran klavyelerinde kelime tamamlama ve kelime tahmininin yazma hızına etkisini araştırmışlardır. Çalışmada, 20-38 yaş aralığında, 5 kadın ve 5 erkek katılımcı ile bir test yapılmıştır. 7 katılımcı kelime tahmini ile en hızlı yazma hızına ulaşırken, 2 katılımcı kelime tamamlama ile en hızlı yazma hızına ulaşmıştır. Sadece bir katılımcı her iki metotla da bir hız artışı elde edememiştir. Sonuç olarak bu çalışma ile kelime tamamlama ve tahmin sistemlerinin ekran klavyeleri ile kullanımlarında hız artışı sağladığı tespit edilmiştir. Tam ve Wells (2009), cep bilgisayarlarında (PDA) kelime tahmin listesi kullanımının faydalarını değerlendirmiştir. Cep bilgisayarlarında kelime tahmin aracı olarak WordQ yazılımı kullanılmıştır. Çalışmada, 11-14 yaş aralığında 21 katılımcı yer almıştır. Sonuçta, cep bilgisayarında kelime tahmin listesi kullanımı ile olumlu sonuçlar elde edileceği belirtilmiştir. Belirtilen bu çalışmalar, kelime tamamlama ve tahmin kullanımının KS veya diğer formüsel değerlendirme ölçütlerinin dışında kullanıcıya sağlayacağı faydaları göstermektedir.

Bu çalışmanın amacı, Türkçeye özgü bir klavye olan F klavye ile birlikte kullanılmak üzere metin bilgisinden yararlanan bir kelime tamamlama sistemi geliştirilmesidir. Öncelikle KS açısından faydalı bir sistem geliştirilmesi hedeflenmiştir. Bununla birlikte dile özgü klavyenin ergonomik katkılarından ve metinlerin anlamlı bilgilerinden

yararlanarak kelime tamamlamanın faydası artırılmaya çalışılmıştır. Çalışma ile ilgili daha detaylı bilgi Karabulut (2016) tarafından yapılan yüksek lisans tez çalışmasında sunulmuştur.

2. Kaynak Araştırması (Literature Survey)

Kelime tamamlama üzerine yapılan basit yöntemlere dayalı ilk çalışmalar 1980'li yıllarda başlamıştır (Fazly ve Graeme, 2003). İlk zamanlarda yapılan çalışmalar, kelime tamamlamayı genellikle bir alternatif ve destekleyici iletişim aracı olarak ele alırken son zamanlarda farklı alanlarda kullanımına yönelik çalışmalar da yapılmıştır. Foster vd. (1996), makine çevirisi için yeni bir yaklaşım sunmuşlar ve bu yaklaşımın uygunluğunu belirlemek için otomatik kelime tamamlama sistemi kullanmışlardır. Van Den Bosh ve Bogers (2008), kelime tamamlama sisteminin mobil cihazlarda kullanımı için çalışma yapmışlardır. Karunaratne vd. (2013), Sinala dili için mobil kısa mesajlarda kullanılmak üzere bir cümle tahmin sistemi geliştirmişlerdir. Geliştirdikleri yaklaşım sayesinde tahmin sistemi kullanılmadan mesaj yazımına kıyasla mesaj yazma süresi kısaltılmıştır. Darragh vd. (1990) kelime tahmini ile kullanıcının yazma hızını artırmak ve iletişimi kolaylaştırmak için etkileşimli klavye çalışması yapmışlardır. Bununla birlikte kelime tamamlama veya kelime tahmin sisteminin farklı aygıtlarda kullanımı üzerine çok sayıda patent çalışması bulunmaktadır (Miller, 1998; Chronqvist, 2010; Unruh, 2010; Thorn, 2010; Bower vd., 2011; Tocci, 2012).

Kelime tamamlama sistemlerinde, kullanıcıya sağlanan faydayı artırmak için farklı metotlar kullanılmıştır. Özellikle kullanılan corpus üzerinde farklı yöntemler denenerek daha doğru tahminler listelenmeye ve daha fazla tuş tasarrufu elde edilmeye çalışılmıştır.

Garay-Vitoria ve Gonzalez-Abascal (1997), cümlenin sözdizim analizine dayalı kelime tahmini üzerine çalışma yapmışlardır. Çalışmada geliştirilen metot, kullanıcının sözlüğüne göre kendini adapte edebilme özelliğine sahiptir. Kelime tahmini için önerilen bu yöntem ile özellikle kullanılan sözdizim analizinden dolayı sadece istatistiksel olan yöntemlerden daha iyi sonuç elde edildiği belirtilmiştir. Diğer bir çalışmada Palazuelos vd. (1998), İspanyolca için hem istatistiksel kurallar hem de gramer kurallarını uygulayan kelime tahmin sistemi üzerine çalışma yapmışlardır. Geliştirilen tahminci, sadece bir anahtar (switch) kullanılarak yazabilecek durumda olan, motor becerileri yetersiz kişilerin yazma oranını artırmak için birkaç yazılım uygulaması içermektedir. Sistemde kelime tahminin bulunması ile bu kişilerin yazma oranının 2-3 wpm (words per minute- dakika başına kelime) değerinden 8-10 wpm değerine kadar arttığı saptanmıştır. Hunnicutt ve Carlberger (2001), İsveç dili için gramer olarak daha uygun kelimeler önererek kullanıcıya daha az zihinsel yük sağlayan ve KS açısından önceki tahmincilerden iyi olan bir kelime tahmin sistemi geliştirmeyi hedeflemişlerdir. Kelime tahmini için Markov modelleri ve sezgisel yaklaşımlar kullanılmıştır. Geliştirilen program testlerinde %46 KS değeri elde edilmiştir.

Fazly (2002) ve Fazly ve Graeme (2003), kelime tamamlama sisteminde sözdizim kurallarını dikkate alarak bir çalışma yapmışlardır. Kelime tamamlamada önerilen kelimelerin sözdizimsel olarak daha uygun olmasını sağlamak için sözcük türü (part-of-speech) ile n-gram bilgisini kombine eden iki algoritma sunulmuş ve test edilmiştir. Geliştirilen algoritmalarından biri ile temel algoritmalara kıyasla KS değerinde artış elde etmişlerdir. Al-Mubaid (2003), çalışmasında kelime tahmin problemi için yeni bir yaklaşım sunmuştur. Çalışmada, verilen cümle bağlamına uygun kelimeyi tahmin etmek için bir veri madenciliği aracı kullanılmıştır. Yaklaşımın, benzer metotlara kıyasla daha az eğitim metni gerektirdiği ve tahminleri %93 doğruluk oranıyla ürettiği saptanmıştır.

Li ve Graeme (2005), kelime tamamlama üzerine dikkat çeken çalışmalardan birini yapmışlardır. Çalışmada, fiziksel engelli kişiler için kullanılan kelime tamamlamaya semantik bilginin katkısı araştırılmıştır. Sadece n-gram kullanımı yerine anlamsal olarak daha uygun kelimeler tahmin etmek n-gram istatistikleri ile semantik bilgiyi kombine eden bir yaklaşım sunulmuştur. Sadece n-gram veya sadece sözcük türü bilgisini kullanan kelime tamamlama sistemlerine kıyasla önemli performans elde edilmiştir. Sistem tuş tasarrufunda %14,63 'e kadar artış sağlanmıştır. Trnka vd. (2006), kelime tahminlerini iyileştirmek için bir çalışma yapmışlardır. Çalışmada bir dil modeli ile konu modelleme (topic modelling) adapte eden iki metot geliştirmişlerdir. Geliştirilen metotlar kelime tahminine uygulanmış ve trigram üzerinde belirgin bir gelişme sağlanmıştır. Geliştirilen kelime tahmininde konu modelleme ile her bir kelime için basılması gereken tuş sayısı ortalama 2,5 'e düşürülmüştür. Miller ve Wolf (2006), kelime tamamlama sisteminde saklı anlam analizi (LSA, latent semantic analysis) yönteminden faydalanmışlardır. Çalışmada tuşlanan kelimenin anlamsal içeriğini modellemek için LSA kullanılarak dil bağımsız bir kelime tamamlama algoritması geliştirilmiştir. Geliştirilen algoritma ile %56 KS ve %42 isabet oranına ulaşılmıştır.

Al-Mubaid (2007), kelime tahmini için bağlam öznitelikleri (context features) ve makine öğrenmesi tabanlı yeni bir yaklaşım sunmuştur. Çeşitli özellik-çıkarma metotları ile yüksek ayırt edici özellikler çıkarılmış ve bunlarla kelime tahminci eğitilerek problem bir öğrenme-sınıflandırma işine dönüştürülmüştür. Test edilen sistem ile benzer çalışmalara kıyasla önemli sonuçlar elde edilmiştir. Bazı testlerde doğru tahmin oranı %91 değerine

ulaşmıştır. MacKenzie ve Zhang (2008), göz ile yazmada kelime tahmininin kullanıldığı bir çalışma yapmışlardır. Çalışmada, kelime tahmin ve harf tahmini olmak üzere iki yazma tekniği ve bir sabitleme algoritması sunulmuştur. Harf tahmini işleminde, kelime tahminine benzer şekilde muhtemel 3 harf belirlenip ekran klavyesinde görüntülenmiştir. Sonuçta, harf tahminin kelime tahmini kadar iyi hatta bazı durumlarda daha iyi sonuç verdiği belirtilmiştir.

Ha-CohenKerner ve Greenfield (2012), İbrani dilinde temel kelime tamamlama ve tahmin işlemleri için tuş tasarrufunu geliştirecek bir çalışma yapmışlardır. Önerilen modelde 177M kelime, morfolojik analiz, çeşitli n-gram modelleri ve diğer araçlar kullanılmıştır. Çalışmada, İbranice ve diğer diller için yapılan kelime tahmin sistemlerine göre daha yüksek tuş tasarruf oranı sağlanmıştır. Ayrıca çalışmada iki önemli sonuç elde edilmiştir. Birincisi, çalışılan dilin corpusu ne kadar büyük olursa tahminler o kadar iyi olmaktadır. İkincisi, dil modeli sadece bir corpusa dayanırsa morfolojik analiz faydalı olmaktadır. Stoop ve Van Den Bosch (2014), kelime tahmin için farklı bir bilgi kaynağı kullanan bir çalışma yapmışlardır. Sistemde, bilgi kaynağı olarak kişinin kendine özgü ve ayırt edici dil kullanımının (idiolect) yanı sıra sosyal gruplar içi dil kullanımından (sociolect) faydalanılmıştır. Çalışmada veri toplama ve test işlemleri için Twitter kullanılmıştır. Çalışma sonucunda, bir dizi Soothsayer kullanıcılarında %50'den daha fazla tuş tasarrufu sağlandığı belirtilmiştir.

Belirtilen bu çalışmaların yanı sıra geliştirilmiş ve kullanımda olan paket programlar bulunmaktadır. Bu programlardan bazıları ilkokullar ve sağlık merkezlerinde yaygın olarak kullanılmaktadır. Öne çıkanlardan bazıları WordQ (<http://www.goqsoftware.com/wordQ.php>) ve Co:Writer (<http://donjohnston.com/cowriter/>) adı verilen paket programlardır.

3. Materyal ve Yöntem (Material and Method)

3.1. F Klavye (F Keyboard)

F klavye, Türkçeye özgü geliştirilmiş olan Standart Türkçe Klavyedir. 1940'lı yıllarda Dr. İhsan Sıtkı Yener, dile özgü harf düzeninin gerekliliğini fark etmiş ve Türkçeye özgü harf düzeni oluşturma çalışmalarına başlamıştır (MEB, 2012). Bu çalışmalar ancak 1955 yılında resmi olarak desteklenmiştir. Oluşturulan bir komisyon, Türk Dil Kurumu tarafından sağlanan dil verileriyle bilimsel bir çalışma yürütmüş ve F Klavye oluşturulmuştur. Geliştirilen bu klavye ile dünya çapında yapılan yarışmalarda önemli başarılar elde edilerek dile özgü klavyenin başarımı gösterilmiştir.

Bu çalışmada geliştirilen sistem bir klavye ile adapte çalışmak üzere tasarlanmıştır. Türkçe metinlerin kullanıldığı çalışmada Türkçe yazı yazmaya uygun bir klavye olan F klavyeye yer verilmiştir. Bu şekilde dile özgü klavye kullanımı ile kelime tamamlamanın sağlayacağı katkının artırılması hedeflenmiştir.

3.2. Dil Verisi (Corpus)

Kelime tamamlama sistemleri, tahmin listesi oluşturabilmek için dil verisine ihtiyaç duymaktadır. Bu çalışmada gerekli kelime verilerinin, kullanıcının sisteme yükleyeceği metinlerden elde edilmesi esas alınmıştır. Bu şekilde metinlerin sağlayacağı anlamlı bilgilerin kelime tamamlama sistemine katkısını artıracığı ve gereksiz kelime verisi tutulmasını engelleyeceği düşünülmüştür.

Geliştirilen kelime tamamlama sisteminde kelime verisi oluşturma işlemi kullanıcıya bırakılmıştır. Ancak çalışmada, sistemin testi için bir çalışma alanı belirlenmiş ve bu alana ait metinlerden yararlanılmıştır. Adliyelerdeki yoğun klavye kullanımı ve sürekli benzer kelimelerle metinlerin yazılması dikkate alınmış ve adli metinler kullanılmıştır. Çalışmada, Türkiye Cumhuriyeti Adalet Bakanlığının zabıt kâtabi alım sınavlarında kullandığı 2014 ve 2015 yıllarına ait 160 metin sistemin eğitimi için kullanılmıştır (AB, 2014; AB, 2015). Yine bu metinler arasından rastgele seçilen 15 metin ile sistemin testi gerçekleştirilmiştir.

3.3. Tahmin Metotları (Prediction Methods)

Kelime tamamlama sistemlerinde tahmin listesinin oluşturulması için farklı yöntemler kullanılmıştır. Ayrıca tuş tasarrufunu artırmak için farklı yaklaşımlar geliştirilmiştir. Bunlardan bazıları istatistiksel n-gram dil modelleri, sözdizimsel cümle analizi, semantik analiz, konu modelleme ve saklı anlam analizi yöntemleridir. Bu çalışmada, metin analizi işlemiyle metinlerden kelime frekans bilgileri çıkarılmış ve bu frekans bilgileri tahmin listesinin oluşturulmasında kullanılmıştır. Burada kelimelerin unigram ve bigram frekans bilgilerinden yararlanılmıştır.

3.3.1. Unigram Frekansı (Unigram Frequency)

Kullanılan tüm metinler içerisinde, kelimelerin kullanım sayısıdır. Unigram frekans bilgileri, kullanıcının tuşlamaya başladığı kelimeyi tahmin etme işleminde kullanılmaktadır. Kullanıcının tuşladığı harflerle devam eden aday kelimelerden en yüksek frekansa sahip belirli sayıda kelime, tahmin listesinde görüntülenmektedir. Unigram frekanslarının elde edilmesinde uygulanan adımlar Şekil 1' de gösterilen algoritma da verilmiştir.

```

1: Girdi:  $M$ , kullanılan metin.
2: Semboller:  $kelimeleriCikar()$ , parametre olarak verilen bir metnin kelimelerinin listesini döndüren fonksiyondur,  $length()$ , verilen listenin eleman sayısını döndüren fonksiyondur,  $unigramKelimeSay()$ , parametre olarak verilen listede istenilen kelimenin frekansını bulan fonksiyondur.
3:  $T \leftarrow \emptyset$ 
4:  $unigramFrekans \leftarrow \emptyset$ 
5:  $T \leftarrow kelimeleriCikar(M)$ 
6: for  $i \leftarrow 1 \rightarrow length(T)$ 
7:   if  $T_i \notin unigramFrekans$  then
8:      $frekans \leftarrow unigramKelimeSay(T, T_i)$ 
9:      $unigramFrekans \leftarrow unigramFrekans \cup \{(T_i, frekans)\}$ 
10:   end if
11: end for
12: Çıktı:  $unigramfrekans$ 

```

Şekil 1. Metindeki kelimelerin unigram frekanslarının elde edilme algoritması (The algorithm of obtaining the unigram frequencies of the words in the text)

3.3.2. Bigram Frekansı (Bigram Frequency)

Kullanılan tüm metinler içerisinde, tüm olası kelime çiftlerin kullanım sayısıdır. Bigram frekans bilgileri, kullanıcının yazdığı bir kelimedenden sonra gelecek olan kelimenin tahmininde kullanılmaktadır. Yazılmış olan kelimedenden sonra gelebilecek olan aday kelimelerden en yüksek frekansa sahip belirli sayıda kelime, tahmin listesinde görüntülenmektedir. Bigram frekanslarının elde edilmesinde uygulanan adımlar Şekil 2' de gösterilen algoritma da verilmiştir.

```

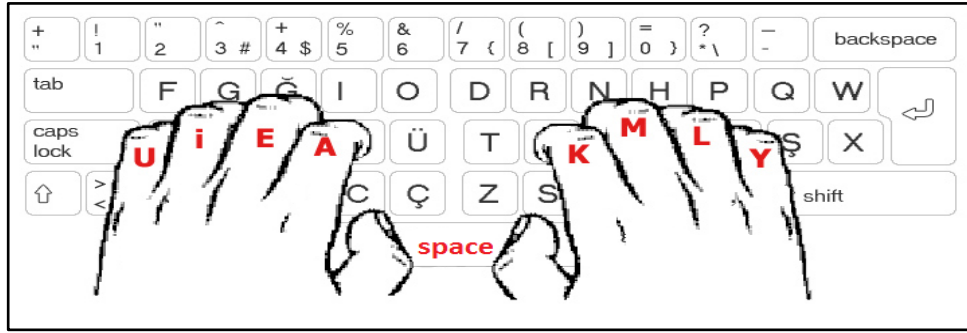
1: Girdi:  $M$ , kullanılan metin.
2: Semboller:  $kelimeleriCikar()$ , parametre olarak verilen bir metni kelimelerine ayıran ve kelime listesini döndüren fonksiyondur;  $length()$ , verilen listenin eleman sayısını döndüren fonksiyondur;  $bigramKelimeSay()$ , parametre olarak verilen listede, verilen iki kelimenin ardışık olarak yer alma frekansını bulan fonksiyondur.
3:  $T \leftarrow \emptyset$ 
4:  $bigramFrekans \leftarrow \emptyset$ 
5:  $T \leftarrow kelimeleriCikar(M)$ 
6: for  $i \leftarrow 1 \rightarrow (length(T) - 1)$ 
7:   if  $\{(T_i, T_{i+1})\} \notin bigramFrekans$  then
8:      $frekans \leftarrow bigramKelimeSay(T, T_i, T_{i+1})$ 
9:      $bigramFrekans \leftarrow bigramFrekans \cup \{(T_i, T_{i+1}, frekans)\}$ 
10:   end if
11: end for
12: Çıktı:  $bigramFrekans$ 

```

Şekil 2. Metindeki kelimelerin bigram frekanslarının elde edilme algoritması (The algorithm of obtaining the bigram frequencies of the words in the text)

3.4. Tahmin Listesi (Prediction List)

Tahmin listesi, kelime tamamlama sisteminde kullanıcının tuşlama bilgisine bağlı olarak sistemin kullanıcıya sunduğu aday kelimeleri barındırmaktadır. Çalışmada, kullanıcıya 8 elemanlı bir aday kelime listesi sunulmaktadır. Tahmin listesinin 8 elemanlı olması sistemin F klavye ile bütünleşik kullanımından kaynaklanmaktadır. Şekil 3 'te verildiği üzere F klavye kullanıcıları parmaklarını, klavyede orta sırasında bulunan 8 harf tuşuna konumlandırmaktadır. Bunlar sol elde U, İ, E, A harf tuşları ve sağ elde K, M, L, Y harf tuşlarıdır. Her iki elin başparmakları ise boşluk (space) tuşuna konumlanmaktadır. Bu bilgi dikkate alınarak kullanıcıya belirtilen 8 harf tuşuna karşılık gelecek olan 8 aday kelime içeren bir tahmin listesi sunulmaktadır. Tahmin listesinde bulunan kelimeler, kullanıcının tuşlamasına bağlı olarak belirlenen en yüksek frekansa sahip 8 aday kelimedir.



Şekil 3. F Klavye Parmak Konumları (Finger Positions for F Keyboard)

$i=1, 2, 3, 4, 5, 6, 7, 8$ ve $F_i = i$. tahminin frekans değeri olmak üzere tahmin listesi;
 $F_1 \geq F_2 \geq F_3 \geq F_4 \geq F_5 \geq F_6 \geq F_7 \geq F_8$ olacak şekilde en yüksek frekansa sahip 8 kelimeyi içermektedir.

Tahmin listesi kelimeleri ekranda, 8 harf tuşu ile paralel olacak şekilde yatay olarak sıralanmaktadır. Belirlenen 8 aday kelimenin ekranda hangi sırada görüntüleneceği parmakların aktif kullanılabilme durumuna göre belirlenmiştir. Burada sağ elini yoğun kullanan klavye kullanıcıları düşünülerek buna uygun şekilde sıralamayı sağlayacak öncelik değerleri belirlenmiştir. Örneğin, tahmin listesi için ele alınacak 8 kelimedeki frekans en yüksek olan kelime, klavyede en rahat kullanıldığı düşünülen ve en yüksek öncelik atanan sağ el işaret parmağına gelecek şekilde konumlandırılmıştır. Tablo 1’de belirtilen bu öncelik değerleri ve tahminleri temsil eden harf tuşları verilmiştir. Tahmin listesi için ele alınan 8 kelime, ekranda bu öncelik değerlerine göre sıralanacaktır.

Tablo 1. Öncelik Sıralaması (Priority Order)

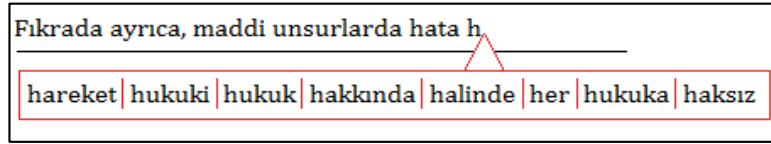
Tuş	Tahmin Frekansı (F_i)	Öncelik
K	1.tahmin (F_1)	1
M	2.tahmin (F_2)	2
A	3.tahmin (F_3)	3
E	4.tahmin (F_4)	4
L	5.tahmin (F_5)	5
Y	6.tahmin (F_6)	6
İ	7.tahmin (F_7)	7
U	8.tahmin (F_8)	8

Geliştirilen sistemin kullanımı esnasında tahminler belirtilen şekilde frekans ve öncelik değerleri temel alınarak ekranda listelenecektir. Ancak kullanıcı doğru kelimeyi tahmin listesinde gördüğünde bu kelimeyi seçebilmelidir. Bunun için bir seçim metodu belirlenmelidir. Seçim metodu için de F klavye kullanımı dikkate alınmıştır. Şekil 3 incelendiğinde 8 parmak harf tuşlarına denk gelirken başparmakların space tuşuna basmakla görevli olduğu görülmektedir. Parmakların bu konum bilgileri kullanılarak seçim işlemi için ikili tuş kombinasyonları belirlenmiştir. Belirtilen 8 harf tuşunun temsil ettiği tahminlerden seçim yapabilmek için space tuşu ile ilgili harf tuşuna birlikte basılmalıdır. Tablo 2’de belirtilen tuş kombinasyonları verilmiştir.

Tablo 2. Tuş Kombinasyonları (Key Combinations)

Tahmin (Frekans- F_i)	Tuş Kombinasyonu
1.tahmin(F_1)	Space + K
2.tahmin (F_2)	Space + M
3.tahmin (F_3)	Space + A
4.tahmin (F_4)	Space + E
5.tahmin (F_5)	Space + L
6.tahmin (F_6)	Space + Y
7.tahmin (F_7)	Space + İ
8.tahmin (F_8)	Space + U

Örnek: Şekil 4’te kelime tahmin listesi oluşturulması ve bu listeden kelime seçimine örnek verilmiştir.



Şekil 4. Kelime Tahmin Listesi Örneği (An Example Prediction List)

Yukarıda yazılmakta olan kelime için 8 kelime tahmini yatay olarak verilmiştir. Bu kelimelerin frekans sıralaması şu şekildedir;

$$F_{halinde} \geq F_{her} \geq F_{hakkında} \geq F_{hukuk} \geq F_{hukuka} \geq F_{haksız} \geq F_{hukuki} \geq F_{hareket}$$

Kullanıcı, burada görülen kelimelerden birini seçebilmek için aşağıda belirtilen tuş kombinasyonlarından ilgili olanını kullanmalıdır.

"halinde"	→ Space + K
"her"	→ Space + M
"hakkında"	→ Space + A
"hukuk"	→ Space + E
"hukuka"	→ Space + L
"haksız"	→ Space + Y
"hukuki"	→ Space + İ
"hareket"	→ Space + U

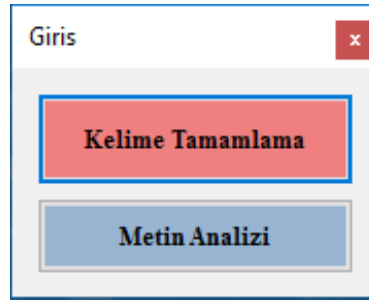
3.5. Programlama Dili ve Veri Tabanı (Programming Language and Database)

Çalışmada öngörülen kelime tamamlama sisteminin geliştirilmesi için bir uygulama geliştirme platformu, programlama dili ve veritabanı belirlenmiştir. Bu aşamada görsel programlamaya uygun bir platform ve hızlı sorgu yapılabilen bir veritabanı seçilmeye çalışılmıştır. Sistem, Visual Studio 2013 uygulama geliştirme platformunda, C# programlama dili kullanılarak geliştirilmiştir. Kelime verilerinin tutulması ve gerektiğinde erişilmesi için MySql veritabanı kullanılmıştır.

4. Deneysel Sonuçlar (Experimental Results)

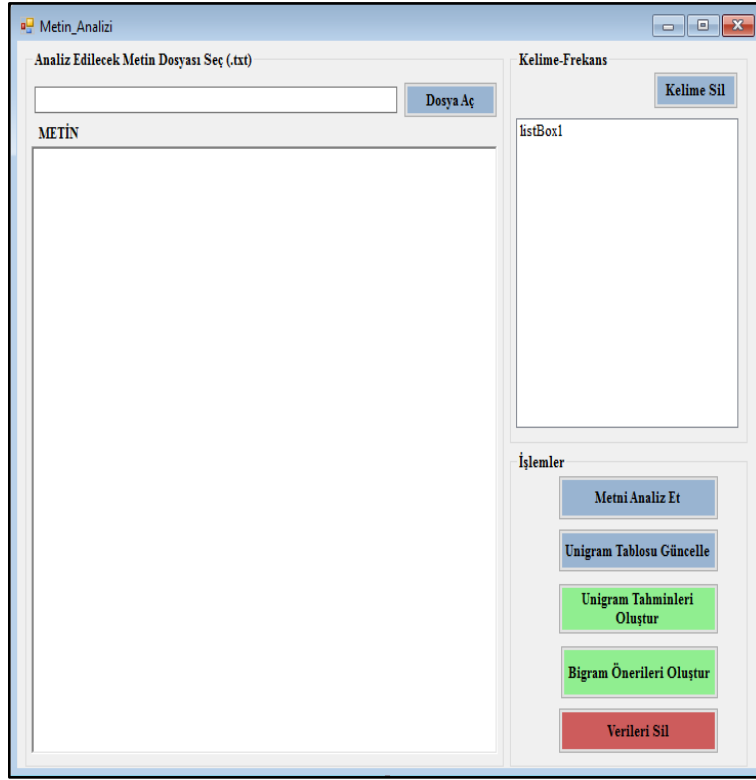
Materyal ve Yöntem bölümünde detaylı olarak belirtilen araçlar kullanılarak öngörülen kelime tamamlama sistemi geliştirilmiştir. Sistemde Giriş, Metin Analizi ve Kelime Tamamlama olmak üzere 3 arayüz bulunmaktadır.

Giriş Arayüzü: Bu arayüzde bulunan iki buton ile kullanıcının seçim yapmasına olanak sağlanmıştır. Kullanıcı kendi metinlerini kullanarak kelime veritabanını oluşturmak için metin analizi butonunu ve kelime tamamlama işlemini başlatmak için ise kelime tamamlama butonunu kullanır. (Şekil 5)



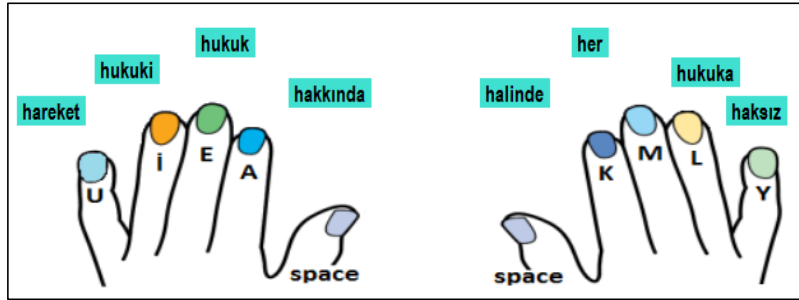
Şekil 5. Giriş arayüzü (Main Interface)

Metin Analizi Arayüzü: Bu arayüz kullanıcının istediği metinleri analiz ederek veritabanını oluşturmasına olanak sağlamaktadır. (Şekil 6) Sistemde kullanılmak üzere txt uzantılı dosyada bulunan metinler seçilebilmektedir. Arayüz ile seçilen metinlerdeki kelimelerin unigram ve bigram frekans bilgileri çıkarılır ve bu bilgiler veritabanında ilgili tablolara aktarılır. Ayrıca kullanıcı istemediği kelimeleri veritabanına aktarılmadan silbilmektedir.



Şekil 6. Metin Analizi Arayüzü (Text Analysis Interface)

Kelime Tamamlama Arayüzü: Bu arayüzde kelime tahminleri listelenmektedir. Arayüz imleç takibi ile konumunu kullanıcının odaklandığı noktaya göre ayarlamaktadır. Bununla birlikte klavye dinlemesi ile kullanıcının tuşlamaları takip edilmekte ve uygun tahminler yatay olarak listelenmektedir (Şekil 7).

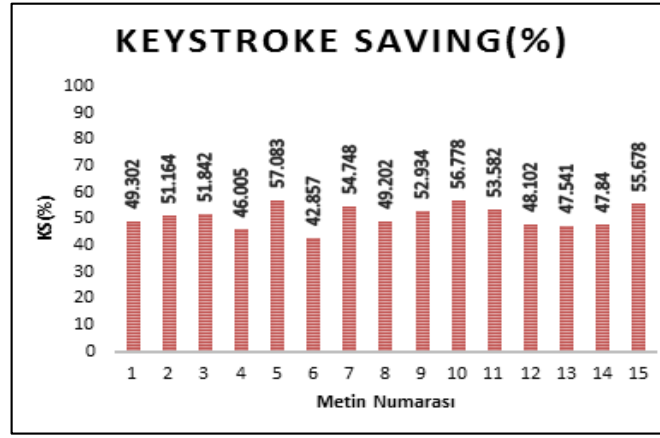


Şekil 7. Kelime Tamamlama Arayüzü (Word Completion Interface)

Kelime tamamlama arayüzüne kelime seçimi için gerekli tuş kombinasyonları eklenmiştir. Materyal ve Yöntem kısmında detaylı olarak verilen tuş kombinasyonlarından ilgi olanı kullanarak kelime seçimi yapılabilmektedir. Seçilen kelime doğrudan yazılarak tamamlanmaktadır.

5. Sonuç ve Tartışma (Result and Discussion)

Tuşlama Tasarrufu- KS: Geliştirilen sistemin testi için 160 metin üzerinde metin analizi işlemi yapılmış ve bu şekilde sistemin kullanımı için gerekli kelime verileri sağlanmıştır. Sistem, benzer çalışmaların değerlendirilmesinde yaygın olarak kullanılan ve sistemin kullanıcıya sağladığı tuş tasarrufunu ölçen KS açısından değerlendirilmiştir. Bu işlem için kullanılan 160 metin arasından rastgele 15 metin seçilmiştir. Bu metinlerin her biri için sistemin sağladığı tuş tasarruf oranı hesaplanmıştır. Hesaplanan bu oranlar Şekil 8' de grafiksel olarak verilmiştir. Elde edilen bu sonuçlar değerlendirildiğinde 57% değerine ulaşabilen tuş tasarrufu sağlanabildiği görülmüştür. Genel olarak değerlendirildiğinde ise 15 metin için ortalama 50,97% oranında tuş tasarrufu sağlanmıştır. Bu da tuş basımının neredeyse yarı yarıya azaltılabileceğini göstermektedir.



Şekil 8. Tuşlama Tasarrufu Oranları (Keystroke Saving Rates)

Tahmin Listesi Eleman Sayısı: Mevcut çalışmalar incelendiğinde, tahmin listesinin genellikle 5 elemanlı olduğu görülmüştür. Bazı çalışmalarda, tahmin listesi eleman sayısının fazla olmasının, kullanıcının daha geniş alana bakmasını gerektirerek harcanacak olan zihinsel çabayı artırabileceği belirtilmiştir. Ancak bu çalışmada, kullanıcının tuşlamasına bağlı olarak frekansı yüksek kelimeler işaret parmaklarına doğru, yani tahmin listesinin ortasına doğru kaymaktadır. Bu özellik, kelimenin kullanıcının görüş alanına girmesini sağlamaktadır. Kullanıcı, 8 elemanlı tahmin listesinin tamamına bakmadığında bile tuşlanmakta olan kelime, tuşlamaya bağlı olarak orta noktaya doğru kaymaktadır. Ayrıca sisteme eklenen imleç takibi özelliği ile tahmin listesi imlecin konumuna göre ilerlemekte ve her durumda kullanıcının görüş alanında bulunmaktadır. Bu şekilde tahmin listesi eleman sayısının fazla olmasından kaynaklanabilecek problemler azaltılmaya çalışılmıştır.

Tahmin Listesinden Kelime Seçimi: Geliştirilen sistemde, kullanıcının tahmin listesinden kelime seçme işlemini daha kolay yapmasını sağlamak için harf tuşları ve boşluk tuşunun kombinasyonlarının kullanıldığı yeni bir seçim yöntemi oluşturulmuştur. Mevcut çalışmalardan farklı olarak bu kombinasyonlar ile kullanıcı, klavyede yazarken parmaklarını harflerden uzaklaştırmadan ve dikkatini tuş arama ile dağıtmadan seçim işlemini yapabilecektir. Ayrıca belirlenen bu tuş kombinasyonları özellikle normal kullanıcılar için zamanla öğrenilebilir niteliktedir. Kullanıcı bir süre sonra klavyeye bakmadan kolaylıkla seçim işlemini yapabilecektir. Bu şekilde kelime tamamlama sisteminin kullanımında önemli bir kolaylık sağlanacağı öngörülmektedir.

Geliştirilen sistemde tahmin listesi elemanları, yatay olarak sıralanmaktadır. Bu yatay sıralama oluşturulurken parmakların aktif kullanım durumları dikkate alınmıştır. Frekansı yüksek kelime daha rahat kullanılan parmak ile eşleştirilmiştir. Bu da tahmin listesinden kelime seçim işlemini kolaylaştırması beklenen bir diğer faktördür.

Metin Bilgisi Kullanımı: Yapılan çalışmalarda genellikle doğrudan dile ait sözlükler kullanılmıştır. Bunun dışında metin bilgilerinden veya Twitter gibi sosyal medya verilerinden yararlanan çalışmalar da bulunmaktadır. Bu çalışmada metin bilgisinden yararlanılması esas alınmıştır. Kişilerin normal hayatta ve çalışma hayatında kullandıkları kelimeler, sözlük kelimeleriyle kıyaslandığında daha dar bir çerçevede ve genellikle benzer kelimelerdir. Ayrıca sıklıkla benzer metinler yazılmaktadır. Bu durumda, kullanılacak olan kelime tamamlama sisteminin genel sözlükler yerine yazılan metinlerden elde edilecek olan anlamlı kelime bilgilerini kullanmasının daha uygun bir yaklaşım olacağı düşünülmüş ve çalışmada metinlerden faydalanılmıştır. Bu şekilde daha az ve daha anlamlı kelime verisi ile verimli bir kelime tamamlama sistemi geliştirilebileceği öngörülmüştür.

Türkçe sondan eklemeli bir dildir ve metinlerde kelimeler genellikle aldıkları eklerle birlikte bulunmaktadır. Kelime tamamlama sisteminde temel amaç olabildiğince tuş tasarrufu sağlamaktır. Bu sebeple metinlerden çıkarılan kelimeler ekleriyle birlikte tutulmuştur. Bunun sistemin sağlayacağı tuş tasarrufunu artırması beklenmektedir.

Kelime Tahmin Yöntemi: Yapılan çalışmalarda kelime tahminlerini oluşturmak için kullanılan yöntemlerden bazıları kelime frekans bilgisi, semantik analiz, istatistiksel n-gram dil modelleri ve sözdizimsel analiz yöntemidir. Bu çalışmada, tahminlerin belirlenmesi için basit bir yöntem olan kelimelerin unigram ve bigram frekansları kullanılmıştır. Çalışmada alana veya kişiye özgü metinlerin gerekli anlamsal bilgiyi taşıdığı düşünülmüş ve bu sebeple karmaşık yöntemlerden kaçınılmıştır. Sonuçta, geliştirilen sistem ile kayda değer bir tuş tasarruf oranı elde edilmiştir. Bu oran belirtilen yaklaşımın uygunluğunu doğrular niteliktedir.

Sonuç olarak, hem alternatif ve destekleyici iletişim aracı olarak fiziksel, zihinsel veya öğrenme engelli kullanıcılar hem de normal klavye kullanıcıları için Türkçe yazı yazmaya uygun, metin bilgisinden faydalanan ve dile özgü

klavye olan F klavye ile birlikte çalışan bir kelime tamamlama sistemi geliştirilmiştir. Sistem yaklaşık **50%** tuşlama tasarrufu oranı ile başarılı bir performans göstermiştir. Bunun dışında mevcut çalışmalardan farklı olarak dile özgü klavye ile adapte edilmesi yeni bir yaklaşım sunmaktadır. Metin bilgisi kullanımı ile kişiye veya çalışma alanına göre özelleşebilen bir tasarım sunulmuştur. Gelecek çalışmalarda, sistemin aktif kullanımı esnasında da kelimelerle ilgili bilgi toplaması sağlanarak tasarım kullanıcı etkileşimli hale getirilebilir. Bunun yanı sıra kullanıcı görüşleri dikkate alınarak sistem yapılandırılabilir.

Çıkar Çatışması (Conflict of Interest)

Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir. No conflict of interest was declared by the authors.

Kaynaklar (References)

- AB, 2014. Adalet Bakanlığı, Zabıt Kâtipliği Uygulama Sınavı Metinlerine İlişkin Duyuru, <http://www.pgm.adalet.gov.tr/duyuru/2014/eylul/zbt.htm> (Erişim Tarihi: 20.07.2015)
- AB, 2015. Adalet Bakanlığı, Zabıt Kâtipliği Uygulama Sınavı Metinlerine İlişkin Duyuru <http://www.pgm.adalet.gov.tr/duyuru/2015/subat/metin.htm> (Erişim Tarihi: 20.07.2015)
- Al-Mubaid, H., 2003. Context-based word prediction and classification. In: Proceedings of the 18th International Conference on Computers and their Applications CATA'2003, Hawaii-USA, 384-388.
- Al-Mubaid, H., 2007. A Learning-Classification Based Approach for Word Prediction. *Int. Arab J. Inf. Technol*, 4, 264-271.
- Anson, D., Moist, P., Przywara, M., Wells, H., Saylor, H. and Maxime, H., 2006. The effects of word completion and word prediction on typing rates using on-screen keyboards. *Assistive technology*, 18, 146-154.
- Bower, J., Furuuchi, K., Liu, S., Morimoto, K., Robbins, D., Laughlin, C. and Davis, P., 2011. Context based word prediction. U.S. Patent No. 7,912,700.
- Chronqvist, F. A., 2010. Electronic device with text prediction function and method. U.S. Patent No. 12/363,799.
- Darragh, J. J., Witten, I. H. and James, M. L., 1990. The reactive keyboard: A predictive typing aid. *Computer*, 23, 41-49.
- Fazly, A., 2002. The use of syntax in word completion utilities. MSc Thesis, University of Toronto, Canada.
- Fazly, A. and Graeme, H., 2003. Testing the efficacy of part-of-speech information in word completion. In: Proceedings of the 2003 EACL Workshop on Language Modelling for Text Entry Methods. Association for Computational Linguistics, Budapest-Hungary, 9-16.
- Foster, G., Isabelle, P. and Plamondon, P., 1996. Word completion: A first step toward target-text mediated IMT. In: Proceedings of the 16th conference on Computational Linguistics- Association for Computational Linguistics., 1, 394-399.
- Garay-Vitoria, N. and Gonzalez-Abascal, J., 1997. Intelligent word-prediction to enhance text input rate (a syntactic analysis-based word-prediction aid for people with severe motor and speech disability). In: Proceedings of the 2nd international conference on intelligent user interfaces, New York-USA, 241-244.
- Handley-More, D., Deitz, J., Billingsley, F. F., & Coggins, T. E., 2003. Facilitating written work using computer word processing and word prediction. *American Journal of Occupational Therapy*. 57 (2): 139-151.
- HaCohen-Kerner, Y. and Greenfield, I., 2012. Basic word completion and prediction for Hebrew. In *String Processing and Information Retrieval*, Springer Berlin Heidelberg, 237-244.
- Hunnicut, S., and Carlberger, J., 2001. Improving word prediction using Markov models and heuristic methods. *Augmentative and Alternative Communication*, 17, 255-264.
- Karabulut, B., 2016. F Klavye için Metin Analizi Tabanlı Kelime Tamamlama Sistemi Tasarımı. MSc Thesis, University of Kırıkkale, Turkey.
- Karunarathne, M. S., Nanayakkara, L. D. J. F. and Ponnampereuma, K., 2013. Sentence prediction on SMS in Sinhala language. *International Journal of Scientific and Research Publications*, 3, 392.
- Li, J. and Graeme, H., 2005. Semantic knowledge in word completion. In: Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility, ACM, Baltimore-USA, 121-128.
- Miller, J. W., 1998. Word prediction system. U.S. Patent No. 5,805,911.
- MacKenzie, I. S. and Zhang, X., 2008. Eye typing using word and letter prediction and a fixation algorithm. In: Proceedings of the 2008 symposium on Eye tracking research & applications, Savannah-Georgia, 55-58.
- MEB, 2012. Millî Eğitim Bakanlığı, Hızlı Yazma, http://kozanhem.meb.k12.tr/meb_iys_dosyalar/01/11/113116/dosyalar/2012_12/05104825_hzlyazadaletmegep.pdf (Erişim Tarihi: 06.09.2015)
- Miller, T. and Wolf, E., 2006. Word completion with latent semantic analysis. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on Pattern Recognition, IEEE, Hong Kong-China*, 1252-1255.
- Palazuelos, S. E., Aguilera, S., Rodrigo, J. and Godino-Llorente, J. I., 1998. Grammatical and statistical word prediction system for Spanish integrated in an aid for people with disabilities. In *ICSLP, Sydney-Australia*.
- Stoop, W., van den Bosch, A. P. J., 2014. Using idiolects and sociolects to improve word prediction. Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg-Sweden, 318-327.
- Tam, C., and Wells, D., 2009. Evaluating the benefits of displaying word prediction lists on a personal digital assistant at the keyboard level. *Assistive technology*, 21, 105-114.
- Tocci, M., 2012. Predictive word completion, U.S. Patent No. 13/162,319.
- Thorn, O. K., 2010. Word prediction. U.S. Patent No. 7,698,326.
- Trnka, K., Yarrington, D., McCoy, K. and Pennington, C., 2006. Topic modelling in fringe word prediction for AAC. In: Proceedings of the 11th international conference on intelligent user interfaces, Sydney-Australia, 276-278.

- Trnka, K., 2010. Word prediction techniques for user adaptation and sparse data mitigation. PhD Thesis, University of Delaware, USA.
- Unruh, E., 2010. Text entry with word prediction, completion, or correction supplemented by search of shared corpus. U.S. Patent No. 12/943,856.
- Van den Bosch, A. and Bogers, T., 2008. Efficient context-sensitive word completion for mobile devices. In: Proceedings of the 10th international conference on Human computer interaction with mobile devices and services ACM, 465-470.
- Wandmacher, T., & Antoine, J. Y., 2008. Methods to integrate a language model with semantic information for a word prediction component. arXiv preprint arXiv:0801.4716.