

Türkçe İçin Genişletilebilir Sıklık Analiz Programı

An Extendible Frequency Analysis Tool for Turkish

Melek OKTAY¹, Atakan KURT¹, Mehmet KARA²

¹Bilgisayar Mühendisliği Bölümü
Fatih Üniversitesi
{moktay, akurt}@fatih.edu.tr

²Çağdaş Türk Lehçeleri ve Edebiyatları Bölümü
Fatih Üniversitesi
mkara@fatih.edu.tr

Özet

Türkçe metinlerin analiz edilmesi, Türk dilinde, edebiyatında ve çok geniş bir spectrumda önem taşımaktadır. Metinlerdeki dil yapılarını elle saymak çok karmaşık bir iştir. Bu nedenle, Türkçe dökümanları ve sözlükleri işleyen ve analiz eden bir bilgisayar uygulaması gereklidir. Bu çalışmada, Türkçe için bir sıklık analiz programı geliştirilmiştir. Genişletilebilen metin analiz programı, Türkçe metinleri ses, hece, ek, kök, kelime gibi çeşitli dil elemanlarının sıklık dağılımlarını kullanarak analiz eder. Araç Java programlama dili ile geliştirilmiştir ve PCMEF yapısına göre tasarlanmıştır. Geliştirilen program yeni dillerin eklenmesi için kolaylıklar sağlar ve çeşitli Türkçe lehçelerini de aynı şekilde eklemek zor değildir.

Abstract

The analysis of Turkish texts is significant in Turkish language, literature and a wide spectrum of areas. It is a complicated task to count language structures manually. A computer application that processes and analyzes Turkish text documents or document sets (corpus) is beneficial. In this paper, the text processing and analyzing tool is developed to analyze the texts and computes various phonetic, syllable, affix, stem, word, sentence frequencies. The text processing and analyzing tool developed can analyze Turkish texts using the frequency distributions of various language elements such as phonemes, syllables, affixes, words etc. The tool is developed with Java programming language and it is implemented according to PCMEF architecture. The framework provides facilities for adding new languages and it is not difficult to extend to do the same for some Turkish dialects.

Anahtar Kelimeler: PCMEF, Ergin, Java ve Türkçe Metin Analizi.

1. Giriş

Yazılım mimarisi geliştirmek ve geliştirilen yazılım mimarisinin genişletilebilir olması oldukça zahmetli ve zor bir iştir. Bunun içindir ki son yıllarda yazılım firmaları spesifik

bir yazılım geliştirmek yerine kendileri için yazılım mimarisi iyi tasarlanmış frameworkler geliştirmektedirler.

Biz bu çalışmamızda Türkçe dili için genişletilebilir sıklık analiz programı geliştirdik. Geliştirilen yazılım şu anda Türkçe dili için sorunsuz çalışmaktadır ve yazılımın genişletilebilir olması ileride program içerisinde küçük değişiklikler yapılarak diğer Türk Lehçelerine uygulanabilir olması manasına gelmektedir.

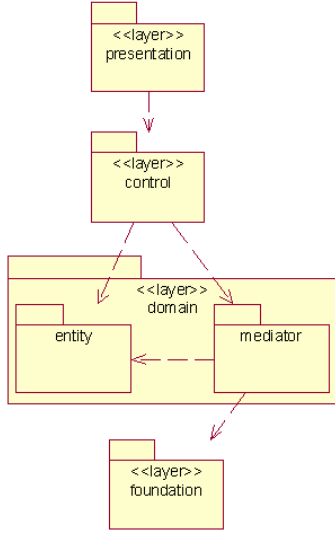
Program basit istemci-sunucu mimarisine göre dizayn edilmiştir. Bunun için kullanıcılar geliştirilen yazılımı kendi bilgisayarlarına kurduktan sonra çeşitli modüller için sunucuya bağlanarak çalışabilirler. Fakat bu yöntemin bazı dezavantajları olabilir. Örneğin kullanıcının bilgisayar özellikleri, geliştirilen yazılımın hız ve hafıza özellikleri gibi performansı etkileyecek özellikleri karşılayamayabilir; ya da kullanıcı kullanmış olduğu her bilgisayara bu yazılımı kurmak istemeyebilir. Bu gibi önemli nedenlerden ötürü yazılımın web üzerinden çalışabilir olması, yazılımın kullanılabilirliğini arttıracak ve kullanıcı dünyanın neresinde olursa olsun basit bir internet tarayıcısı sayesinde yazılımı kullanabilecektir. İşte tam bu noktada yazılımın genişletilebilir bir mimariye sahip olması, geliştirilen programın üzerinde çok fazla değişiklik yapmadan web üzerinde çalışabilir hale gelmesine yardımcı olacaktır. PCMEF [1,2,3] meta-mimarisi de bu bağlamda amacımıza uygun bir mimari olarak seçilmiştir.

Geliştirilen yazılım sayesinde Türkçe metinleri bilgisayarla işleyerek bu metinlerin; ses, hece, ek, kök, kelime sıklık ve oranları yardımıyla analizlerinin yapılmasına imkan sağlayacak bir uygulama geliştirilmiştir. Bu yolla Türkçe metinlerin analiz edilmesi, sadece Türkçe ve edebiyat alanında değil; eğitim-öğretim, psikoloji, sosyoloji, siyaset bilimi, işletme, suç bilimi, tıp gibi alanlarda da çok işe yarayacaktır. Dolayısıyla Türkçe metinleri hiç olmazsa ses, hece, ek, kelime, kelime grubu gibi açılardan inceleyen bir uygulama birçok alanda kullanılabilir [4].

Makalenin geriye kalan kısmı şu şekilde devam etmektedir: Bölüm 2’de PCMEF meta-mimarisi detaylı bir şekilde anlatılmıştır. Mimariye etki eden faktörler 3. Bölüm’de açıklanmıştır. Bölüm 4’te geliştirilen metin analiz programının yapısı, mahiyeti ve içeriği anlatılmıştır. Bölüm 5’te Sonuç ile bitirilmiştir.

2. PCMEF

PCMEF meta-mimarisi beş katmandan oluşmaktadır. Bunlar sırası ile Presentation (prezentasyon), Control (kontrol), Mediator (arabulucu), Entity (varlık) ve Foundation (altyapı) katmanlarıdır. PCMEF meta-mimarisi bu katman isimlerin başharflerinin birleşmesi ile adlandırılır. Bu katmanlar ve aralarındaki bağımlılık Şekil-1’de detaylı olarak gösterilmiştir.



Şekil 1: PCMEF meta-mimarisi.

Presentation (prezentasyon) katmanı geliştirilen programın kullanıcı arayüzleri ile alakalı kodları içermektedir. Mesela Java programlama dilinde arayüz geliştirme kütüphanesi olan Swing kodları burada bulunmaktadır; ya da geliştirilen programın web üzerinde çalışmasını sağlayan Servlet kütüphanesi kullanılarak geliştirilen uygulamaların sınıfları bu katman içerisinde bulunur. Bir önemli husus da bu katmandaki sınıfların hiçbirinde mantık kodu, uygulama için gerekli algoritmaların bulunmamasıdır. Bu sebepten ötürü, belirli bir alan için geliştirilen uygulama çok fazla bir değişiklik gerektirmeden başka bir alana rahatlıkla uyarlanabilir.

Control (kontrol) katmanı prezentasyon katmanı vasıtası ile kullanıcıdan alınan isteklerin yerine getirilmesini sağlayan katmandır. Mantık kodları, algoritmik çözümler gibi uygulamalar için hayati öneme sahip fonksiyonları gerçekleştirmek bu katmanın görevidir.

Entity (varlık) ve Mediator (arabulucu) katmanının bir diğer domain katmanı içerisinde bulunmaktadır. Varlık katmanı bünyesinde içerisinde verileri tutan, nesnelerin en basit hali olan iş-nesneleri bulunmaktadır. Bu iş-nesneleri çalışan programın ihtiyaç duyduğu verileri içermektedir. Arabulucu katmanı da isminden de anlaşılacağı gibi Kontrol, Varlık ve Altyapı katmanları arasında arabulucu özelliğini içerir. Bununla beraber kendi ismini taşıyan tasarım kalıbıyla aynı işi görmektedir [5].

Foundation (altyapı) katmanı veritabanı, dosyalar, web servisleri vb. sistemlerle bağlantıyı sağlar. Bu katmanın haricindeki katmanlar veritabanı bağlantısı ya da dosya okuma gibi işlemleri kendileri yapamazlar. Veritabanı sorgulamaları,

dosya okuma ve yazma gibi işlemlerin sınıfları bu katman içerisinde yer alır.

PCMEF’in barındırdığı katmanlara ek olarak Acquaintance (tanıdık) katmanı bulunmaktadır. Bu katman her katmanda bulunabilecek olan, geliştirilen programa ait arabirimleri (interface) içerir.

PCMEF’in katmanlarının yanında çok önemli olan tasarım prensipleri açıklanmıştır [1] ve bu meta-mimariyi kullanan herkesin bu prensiplere uyması gerekmektedir. Bu prensiplerden en önemlileri Aşağı-Doğru Bağımlılık Prenbi, Yukarı-Doğru Bilgilendirme Prensi, Komşular Arası Haberleşme Prensi, Döngü Eleme Prensi gibi prensiplerdir[1,2,3].

PCMEF meta-mimarisine uyulması sonucu genişletilebilir ve desteklenebilir bir program mimarisine sahip olunur. Bu da günümüzde geliştirilen programlar için çok önemli bir özelliktir. Bu mimari kullanılarak geliştirilen masaüstü yazılımları çok rahatlıkla farklı platformlarda çalışabilir hale getirilebilir. Örneğin sadece prezentasyon katmanında bulunan sınıflar üzerinde yapılacak değişikliklerle uygulama web üzerinde çalışabilir hale getirilebilir; ya da konsolda çalışan bir uygulama rahatlıkla kullanıcı arabirimi (Java Swing, Java Applet) hale getirilebilir; çünkü prezentasyon katmanındaki sınıflar diğer katmandaki sınıflara kullanıcı arabirimi olarak bağlı değildir. Mimariye Etki Eden Faktörler

Türkçe için geliştirilen yazılımın mimarisini etkileyen çeşitli faktörler olmuştur. Bunlardan başlıcaları Türkçe dilinin özellikleri [6,7,8] XML ve XML teknolojileri, yazılımın geliştirilmesinde kullanılacak programlama dili olan Java’nın özellikleri, unicode, internalizasyon faktörleridir.

2.1. XML ve XML Teknolojileri

XML kolay okunabilen ve kullanıcının kendi etiket kütüphanesini oluşturabileceği bir standarttır [9]. Geliştirilen yazılımın ihtiyacı olan bütün bilgiler XML dosyasında tutulmuştur ve program çalıştırıldığı zaman bir kereye mahsus bu bilgiler XML teknolojilerinden XPath kullanılarak dosyalardan okundu. Okunan bu bilgiler Türkçe dilinin özellikleri ve program için gerekli olan konfigürasyon bilgileridir; mesela Türkçe dilindeki harfler ve bu harflerin özellikleri (sesli, sessiz, yuvarlak, kapalı vb.). Bunun sonucu olarak geliştirilen yazılım hiçbir şekilde statik bir bilgiye dayanmamaktadır. Bu özellik sayesinde Türkçe için geliştirilen bu yazılım diğer Türk Lehçeleri (Azerice, Türkmençe vb.) için de genişletilebilir.

2.2. Java ve Unicode

Yazılım Java programlama dili kullanılarak geliştirilmiştir. Bu programlama dili, üzerinde çalıştığı işletim sisteminden bağımsız olması, nesnel tabanlı programlama yapılabilmesi, ve unicode’u desteklediği için seçilmiştir.

Unicode [10] 16-bit’lik bir karakter setidir ve dünyadaki hemen hemen bütün karakterleri kapsar. Temel olarak bilgisayarlar içerik olarak sayılar üzerinde işlem yaparlar ve bilgisayardaki karakterler sayılar şeklinde kodlanırlar. Unicode herbir karaktere benzersiz 16 bitlik bir numara verir. Böylelikle dünyadaki bütün karakter kümelerini hemen hemen kapsayabilir. Fakat karakterlerin düzgün bir

şekilde gösterimi için gerekli olan uygun fontun bilgisayar üzerinde kurulu olması gerekmektedir (Fontun kendisi de bir yazılımdır, sayıları karakterlere çevirip uygun bir şekilde görünmesini sağlarlar).

Java programlama dili kendi içerisinde JDK1.1'den itibaren unicode desteklediği için uygulama çalışırken metinler üzerinde herhangi bir bozulmaya sebebiyet vermez.

2.3. İnternalizasyon

Geliştirilen yazılımın farklı dil bilen kullanıcılar tarafından kullanılabilmesi için program geliştirilirken bazı hususlar göz önünde bulundurulmalıdır. Bu husulardan bir tanesi kullanıcıya kendi dilinde hizmet edilmesidir. Bunun için özkaynak demetleri (resource bundles) kullanılmıştır. Şu anda yazılım Türkçe ve İngilizce arayüzleri üzerinden hizmet verebilmektedir, istenildiği takdirde diğer dillerdeki arayüzler rahatlıkla eklenebilir.

Ayrıca internalizasyonda bir önemli husus yerel (Locale) bilgisidir. Yani kullanıcının çalıştığı bilgisayar üzerinde hangi dili kullandığını ve dünyanın hangi bölümünde bulunduğunu belirten bilgidir. Kullanıcının kullandığı dile göre zaman, saat, sayı formatları ve karakter dizilerinin sıralamaları değişir. Örneğin, America, Zulu, Ant, Zebra, Ångström kelimelerini bir İngiliz America, Ångström, Ant, Zebra, Zulu diye sıralarken, bir İsveçli America, Ant, Zebra, Zulu, Ångström diye sıralar. Geliştirdiğimiz yazılımda kullanıcının sistemindeki yerel bilgi alınarak buna göre bir sıralama ve zaman, saat ve sayı formatlaması yapılır. İnternalizasyon ile alakalı diğer önemli bilgilerin detayı için bakınız [11].

3. Metin Analiz Programı

Türkçe metinleri bilgisayarla işleyerek bu metinlerin ses, hece, ek, kök, kelime, cümle sıklık ve oranları yardımıyla analizlerinin yapılmasına imkan sağlayacak bir uygulama PCMEF meta-mimarisine göre geliştirilmiştir. Bu yolla Türkçe metinlerin analiz edilmesi, sadece Türkçe ve edebiyat alanında değil; eğitim-öğretim, psikoloji, sosyoloji, siyaset bilimi, işletme, suç bilimi, tıp gibi alanlarda da çok işe yarayacaktır. Dolayısıyla Türkçe metinleri hiç olmazsa ses, hece, ek, kelime, kelime grubu, cümle, paragraf (cümle grupları) gibi açılardan inceleyen bu uygulama birçok alanda kullanılabilir.

Örneğin zaman içerisinde değişik sebeplerden ses, yapı, anlam değişikliklerine uğrayan Türkçede meydana gelen değişiklikleri, kalabalık metin kümelerini (corpora) inceleyerek analiz edebiliriz. Bu analizin en önemli dayanak noktası; belirlenen metin kümelerindeki ses, hece, kelime, cümle vs. sıklıklarını ortaya koymak olacaktır. Bu sıklıkları belirlemek, bir araştırmacının kısa zamanda tek başına altından kalkabileceği bir iş değildir. Günümüzde Türkçenin temel metinlerinin birçoğu bilgisayar ortamına geçmiş ve yeni üretilen metinler ise ya doğrudan bilgisayar ortamında veya internette oluşmakta ya da kısa zamanda sayısal ortama aktarılmaktadır. Dolayısıyla bilgisayarlı bir Türkçe sıklık çalışması, metin analizlerinde hem süreyi çok azaltacak, hem de hataları en aza indirecektir. Ayrıca sonuçlar sayısal ortamda oluşturulacağı için elde edilen veriler başka bilgisayar uygulamaları ve kişiler tarafından daha ileri söz dizimi ve anlam analizleri için doğrudan kullanıma hazır olacaktır.

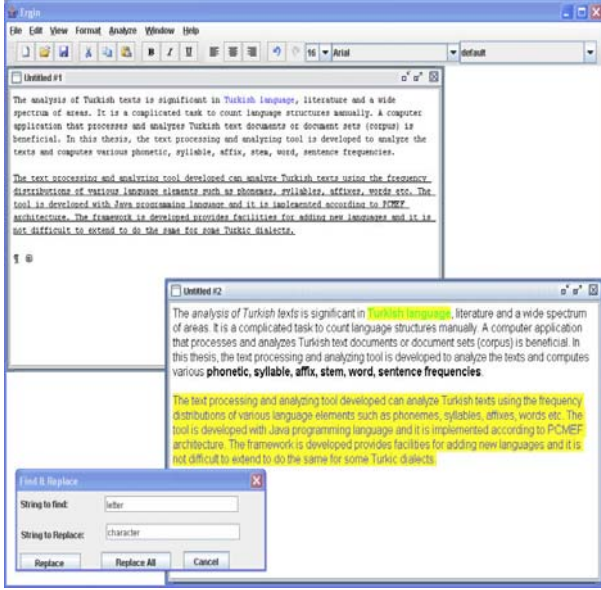
Benzer programlar İngilizce ve diğer diller için geliştirilmiş olsa da, bu programların Türkçe için kullanılmasında bazı önemli engeller bulunmaktadır. Türkçenin alfabesi ve sesleri, heceleme kuralları, kelime yapısı (kökler ve ekler), cümle yapısı ve anlam-kavram örgüsü İngilizce ve diğer dillerden farklıdır. Bu sebeplerden dolayı yabancı diller için geliştirilmiş uygulamalar Türkçe metinler için kullanılamamakta, kullanılsa da tam ve güvenilir sonuç almak mümkün olamamaktadır.

Türkçe metinleri analiz eden bu uygulamanın Türkçenin değişik lehçelerinde yazılmış metinleri -uygulama bu işleme uygun hale getirildiğinde- işleyebilmesi mümkün olabilir. Çünkü Azerice, Türkmençe gibi bazı lehçeler Türkiye Türkçesi ile önemli oranlarda benzeşmektedirler. Programın gerçekleştirilmesinin temel amaçlarından biri de Türkçe metin işleme uygulamasına hiç olmazsa Türkiye Türkçesi dışında bir başka lehçede daha metin işleme özelliğini kazandırmaktır. Bu iş için Türkmençe düşünülmüştür; çünkü Türkmençe, hem Türkiye Türkçesine yakın bir lehçedir hem de bu lehçe üzerine yapılmış tezler ve bilimsel çalışmalarla bu iş için gerekli altyapı bir dereceye kadar hazırlanmıştır. Pek tabii lehçeler üzerine yapılmış çalışmaların sınırlı olmasından dolayı, bunlarla ilgili sıklık analizlerinin Türkiye Türkçesi için yapılan sıklık analizlerine göre bazı yönlerden eksiklikleri olabilecektir. Fakat uygulama genişletilebilir (extensible) olarak geliştirildiği için diğer lehçelerin zaman içerisinde programa eklenmesi kısa zaman içerisinde yapılabilecektir.

Geliştirilen yazılımın masaüstü kısmında çalışan modülü Şekil-2'de gösterilmiştir.

Uygulama var olan metin editörlerinin (Microsoft Word, Open Office vb.) döküman yönetimi ve metin işleme gibi temel işlemlerin hepsini yapabilmektedir. Örneğin kullanıcının bilgisayarında var olan bir metnin açılması, üzerinde yapılan değişikliklerin kaydedilmesi, metinlerin fontlarının değiştirilmesi gibi işlemler rahatlıkla yapılabilmektedir. Bu program şimdilik "txt" ve "rtf" uzantılı metin dosya tipleri desteklemektedir. Bunun yanında herhangi bir kaynaktan kopyalanan metinlerin editörün açılan penceresine yapıştırılması ile de analiz yaptırılabilir.

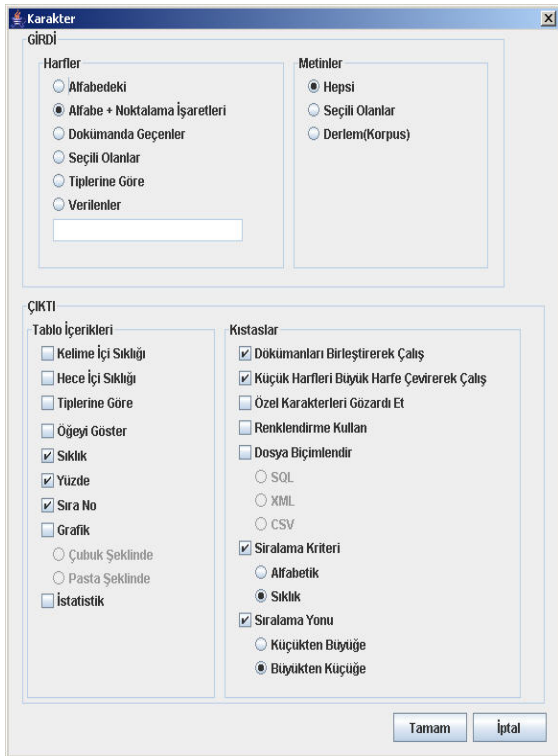
Söz konusu program, aynı anda birden fazla metin üzerinde çalışılabilir; istenilirse metinler birleştirilerek de tek bir metin halinde analiz edilip sonuçlar bir pencerede görülebilir. Bunun yanında, metinler birleştirilmeden her biri farklı bir metin şeklinde analiz edilip, sonuçlar farklı pencerelerde kullanıcıya gösterilebilir. Bu gibi seçenekler tamamen kullanıcının isteği doğrultusunda belirlenir ve buna göre işleme tabi tutulur.



Şekil 2: Genel Kullanıcı arabirimi..

Uygulamanın döküman yönetimi haricinde bir de metin analizi modülü vardır. Yapılacak metin analizi 5 ana başlık altında incelenebilir. Bu ana başlıklar sırası ile harf, hece, kelime, cümle ve paragraftır.

Örnek olarak, harf modülü seçildiği zaman kullanıcıdan girdi bölümünde işleme tabi tutulacak karakter ve metin tiplerinin seçimi istenilmektedir. Kullanıcı burada analizini yapmak istediği seçenekleri seçerek yeni açılan çıktı penceresinden istediği sonuçları görebilir. Karakter modülündeki örnek girdi penceresi ve fonksyonları Şekil-3'te gösterilmiştir.



Şekil-3: Karakter Penceresi Girdi Modülü

Karakter modülünde kullanıcı metin içerisindeki karakterlerin sıklıklarını, yüzdeleri ve sıra numaralarını hesaplatılabilir. Bunun haricinde çeşitli fonksiyonlar mevcuttur. Bunlar:

- **Kelime İçi Sıklığı:** Bir karakterin kelime içinde kaçınıcı sırada/sıralarda yer aldığı tespit etmeye yarar. Örnek çıktı için Şekil-4'te gösterilmiştir.

KELİME İÇİ SIKLIK																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Toplam	Yüde
E	9	29	5	14	14	4	19	2	8	9	5	3	1	0	1	0	0	0	0	0	117	411,40
İ	8	9	8	16	17	5	9	6	9	2	3	5	2	3	1	0	0	0	0	0	103	410,11
A	4	30	7	9	5	9	9	2	6	2	2	0	1	0	0	1	0	0	0	0	87	18,54
L	1	2	11	10	9	9	3	8	4	5	0	1	0	0	0	0	0	0	0	0	63	16,18
T	9	0	11	9	5	11	1	6	4	1	1	2	0	0	0	0	0	0	0	0	60	15,89
R	0	2	15	4	5	6	3	7	0	6	3	0	1	2	1	1	0	0	0	0	58	15,09
K	11	3	13	1	3	10	0	8	1	3	3	0	0	0	0	0	0	0	0	0	36	15,50
S	14	2	10	5	5	5	3	7	0	0	2	0	0	0	0	0	0	0	0	0	53	15,20
B	0	8	8	0	9	2	9	5	3	2	0	2	2	0	1	0	0	0	0	0	51	15,00
M	7	0	7	0	9	9	2	4	3	0	1	0	0	0	2	0	0	0	0	0	44	14,22
Ü	4	15	0	9	5	0	6	0	1	0	0	0	0	0	0	0	0	0	0	0	40	13,93
D	3	4	3	0	3	2	4	1	3	1	1	1	1	1	1	0	0	0	0	0	31	13,04
Y	10	1	4	1	4	6	0	2	1	0	0	0	0	0	0	0	0	0	0	0	29	12,85
U	0	7	0	5	1	2	1	1	3	0	4	0	0	0	0	0	0	0	0	0	26	12,55
I	0	4	1	4	6	0	1	1	0	1	1	1	0	1	0	0	0	0	0	0	25	12,45
B	13	0	1	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	19	11,86
F	13	0	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	11,86

Şekil-4: Kelime İçi Sıklığı

- **Hece İçi Sıklığı:** Bir karakterin kelime içerisinde kaçınıcı hece/hecelerde yer aldığı tespit etmeye yarar. Şekil-5'te hece içi sıklığı analizi gösterilmiştir.

HECE İÇİ SIKLIĞI							
	1	2	3	4	5	6	Toplam
E	9	108	0	0	0	0	117
İ	8	95	0	0	0	0	103
A	7	80	0	0	0	0	87
L	48	0	15	0	0	0	63
T	54	0	5	1	0	0	60
R	29	1	28	0	0	0	58
K	30	2	24	0	0	0	56
S	39	2	12	0	0	0	53
N	16	5	29	1	0	0	51
M	27	0	17	0	0	0	44
Ü	4	36	0	0	0	0	40
D	31	0	0	0	0	0	31
Y	27	0	2	0	0	0	29
U	0	26	0	0	0	0	26
I	0	25	0	0	0	0	25
B	19	0	0	0	0	0	19
F	19	0	0	0	0	0	19

Şekil 5: Hece İçi Sıklığı

- **Tiplerine Göre:** Ünlü ve ünsüzlerin türlerine göre sıklığını tespit etmeye yarar. Örnek çıktı için Şekil-6'ya bakınız.

TIPLERİNE GÖRE HARF SIKLIĞI

Sesli Karakter Sayısı	:	419
Sessiz Karakter Sayısı	:	548
Sesli --> Kalın	:	152
Sesli --> İnce	:	267
Sesli --> Düz	:	332
Sesli --> Yuvarlak	:	87
Sesli --> Geniş	:	225
Sesli --> Dar	:	194
Sessiz --> Sedalı	:	333
Sessiz --> Sedasız	:	215
Sessiz --> Şürekli	:	367
Sessiz --> Şüreksiz	:	181
Sessiz --> Akıcı	:	245
Sessiz --> Sızıcı	:	122
Sessiz --> Patlayıcı	:	181
Sessiz --> Dudak	:	69
Sessiz --> Diş Dudak	:	35
Sessiz --> Diş	:	199
Sessiz --> Diş Dudak	:	15
Sessiz --> Ön Damak	:	206
Sessiz --> Art Damak	:	13
Sessiz --> Gırtlak	:	11

Şekil-6: Tiplerin Göre Ünlü-Ünsüz Sıklığı

Programın üretmiş olduğu çıktılar kullanıcının isteğine göre şekillendirilebilmektedir. Örnek olarak kullanıcı isterse karakterlerin aşağıdan yukarıya doğru sıklık derecesine göre sıralayabilir; ya da bunun yerine alfabetik bir sıralama yapabilir. Bunun yanında kullanıcının isteğine göre metin içerisindeki bütün harfler büyük harfe çevrilebilir.

Aynı şekilde hece, kelime, cümle ve paragraf için gerekli modüller geliştirilmiştir. Bu modüllerin fonksyonları ve detayları için bakınız [4,11].

Bu bölümde anlatılan modüllerin çıktıları masaüstü kullanıcı arabirimi olan Java Swing kütüphanesi kullanılarak geliştirilmiştir. Aynı uygulamanın Web üzerinden sorunsuz bir şekilde çalışması için kullandığımız genişletilebilir PCMEF meta-mimarisi sayesinde Java'daki Servlet kütüphanesi kullanılarak, uygulamanın Web üzerinde sorunsuz bir şekilde çalışması sağlanmıştır.

4. Sonuçlar

Türkiye Türkçesine ait metinlerdeki ses, hece, ek, kelime, cümle ve paragraf sıklıklarını analiz edecek bir uygulamanın geliştirilmesini amaç edinen bu çalışma, birçok dil örgüsünün incelenbilmesine imkan sağlayacak biçimde tasarlanmıştır.

Bilgisayar dünyasında önemli bir yer tutan yazılım mühendisliği prensiplerine uyularak bir uygulama gerçekleştirilmiştir. Uygulama geliştirilirken ileride çıkabilecek gereksinimlere cevap verebilmek için genişletilebilir ve desteklenebilir bir yazılım mimarisi geliştirilmiştir. Bu mimari oluşturulurken PCMEF meta-mimarısından ilham alınmıştır. Bu meta-mimarinin prensiplerine uyularak geliştirilen uygulama aynı zamanda masaüstü uygulamalarda ve Web üzerinde sorunsuz bir şekilde çalışmaktadır.

İleride, geliştirilen bu uygulama Türkçenin değişik lehçelerinde yazılmış metinleri işleyebilir hale getirilecektir. Çünkü Azerice, Türkmence gibi bazı lehçeler Türkiye Türkçesi ile önemli oranlarda benzeşmektedirler.

5. Kaynaklar

- [1] L. A. Maciaszek, B. L. Liong, Practical Software Engineering, Addison Wesley, 2004.
- [2] L. A. Maciaszek, Roundtrip Architectural Modeling, Conf. in Research and Practice in Information Technology Series, Vol. 107, 2005.
- [3] L. A. Maciaszek, Developing Supportable Enterprise Information Systems – Architectural, Managerial and Engineering Imperatives, Int. Conf. on Software Maintenance, pp. 721-722, 2005
- [4] Oktay M., Kurt A., Kara M., "Türkçe İçin Bir Sıklık Analizi Programı", 38th International Congress of Asian and North African Studies (ICANAS 38), ANKARA/TÜRKİYE, Sep. 2007.
- [5] E. Gamma, R. Helm., R. Johnson, J. Vlissides, Design Patterns, Elements of Reusable Software, Addison Wesley, 1995.
- [6] Adalı, O., Türkiye Türkçesinde Biçimbirimler, Papatya Yayıncılık, 2004
- [7] Aksan, D., Anlambilimi ve Türk Anlambilimi, Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Yayınları, Ankara, 1978.
- [8] Toklu, M. O., Dilbilime Giriş, Akçağ Yayınları, 2005
- [9] McLaughlin, B., Edelson. J., Java and XML, O'Reilly 2006.
- [10] Jukka, K. K., Unicode Explained, O'Reilly, New York, 2006
- [11] Oktay, M. "A Text Processing and Analysis Tool for Turkish", Fatih Üniversitesi Master Tezi, 2007