

PAMUKKALE ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR YARDIMI İLE KİTAP SONU İNDEKSİ HAZIRLANMASI

**YÜKSEK LİSANS TEZİ
Mehmet BOZUYLA**

Anabilim Dalı : Bilgisayar Mühendisliği

Programı : Bilgisayar Mühendisliği

Tez Danışmanı: Doç. Dr. Sezai TOKAT

HAZİRAN 2011

YÜKSEK LİSANS TEZ ONAY FORMU

Pamukkale Üniversitesi Fen Bilimleri Enstitüsü 091281035 nolu öğrencisi Mehmet Bozuyla tarafından hazırlanan “**BİLGİSAYAR YARDIMI İLE KİTAP SONU İNDEKSİ HAZIRLANMASI**” başlıklı tez tarafımızdan okunmuş, kapsamı ve niteliği açısından bir Yüksek Lisans tezi olarak kabul edilmiştir.

Tez Danışmanı : Doç. Dr. Sezai TOKAT (PAÜ)
(Jüri Başkanı)

Jüri Üyesi : Doç. Dr. M. Vefa NALBANT (PAÜ)

Jüri Üyesi : Yrd. Doç. Dr. Gürhan GÜNDÜZ (PAÜ)

Pamukkale Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu’nun
..... tarih ve sayılı kararıyla onaylanmıştır.

Fen Bilimleri Enstitüsü Müdürü

Prof. Dr. Nuri KOLSUZ

Bu tezin tasarımı, hazırlanması, yürütülmesi, araştırmalarının yapılması ve bulgularının analizlerinde bilimsel etiğe ve akademik kurallara özenle riayet edildiğini; bu çalışmanın doğrudan birincil ürünü olmayan bulguların, verilerin ve materyallerin bilimsel etiğe uygun olarak kaynak gösterildiğini ve alıntı yapılan çalışmalara atfedildiğine beyan ederim.

Mehmet Bozuyla

İmza

ÖNSÖZ

Bu çalışma ile özellikle Türk dili edebiyatı alanında hazırlanan kitap sonu indekslerini otomatik olarak bilgisayar ortamında hazırlayan bir programın gerçekleştirilmesi üzerinde durulmuştur. Bu amaçla, bağlı liste yapıları kullanılarak hazırlanan algoritmalar ve nesne yönelimli programlama teknikleri ile etkin ve zengin içerikli bir bilgisayar yazılımı hazırlanmıştır. Bu çalışmanın gerçekleşmesinde yardımcı olan tez danışmanım Doç. Dr. Sezai TOKAT’a, dil bilim ve dizinleme konularında yardım ve bilgilerini esirgemeyen Pamukkale Üniversitesi Türk Dili ve Edebiyatı Bölümü öğretim üyelerinden Doç. Dr. Vefa NALBANT’a, Uşak Üniversitesi Türk Dili ve Edebiyatı Bölümü öğretim üyelerinden Yrd. Doç. Dr. Talip YILDIRIM’a ,Pamukkale Üniversitesi Bilgisayar Mühendisliği Bölümü öğretim üyelerinden Yrd. Doç. Dr. Gürhan GÜNDÜZ’e ve Pamukkale Üniversitesi Türk Dili ve Edebiyatı Bölümü öğretim üyelerinden Yrd. Doç. Dr. Bilge Özkan NALBANT ‘a teşekkür ediyorum.

Ayrıca bu çalışmamı sonuçlandırmam sırasında verdikleri desteklerle bana yardımcı olan aileme ve arkadaşlarıma teşekkürlerimi bir borç bilirim.

Haziran 2011

Mehmet Bozuyla

İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	iii
ÖZET.....	ix
SUMMARY	xi
1. GİRİŞ	1
1.1 Tezin Amacı	1
1.2 Literatür Özeti	2
1.3 Tezin Akışı	3
2. DİZİNLEME İŞLEMLERİ.....	5
2.1 Ayırıştırma.....	5
2.1.1 Ayırıştırmada kullanılan yöntemler	5
2.1.1.1 Geleneksel (Elle yazılan fişleme yöntemi) ayırıştırma	5
2.1.1.2 Bilgisayar tabanlı ayırıştırma	5
2.1.2 Ayırıştırma notasyonu	7
2.1.2.1 Satır numarası ekleme	7
2.1.2.2 Çekim eklerinin ayrılması	9
2.1.2.3 Sesteş kelimelerin ayırıştırılması	9
2.1.2.4 Kelimenin dil özelliğinin eklenmesi	10
2.1.2.5 Ses olayları bulunan kelimenin ayırıştırılması	10
2.1.2.6 Birleşik fiillerin belirtilmesi	11
2.1.3 Ayırıştırma süreci	12
2.2 Dizinleme	13
2.2.1 Dizinleme nedir	13
2.2.2 Dizinleme yazılımları.....	13
2.2.2.1 Concordance.....	13
2.2.2.2 Cibakaya.....	14
2.2.2.3 Dizin hizmet programı	15
2.2.2.4 Türkçe metin sıklık çözümleyicisi	16
2.2.2.5 Diğer yazılımlar	17
2.2.3 Dizinleme notasyonu.....	17
2.2.3.1 Satırların hazırlanması	17
2.2.3.2 Madde başlarının hazırlanması	20
2.2.4 Dizinleme süreci	20
2.2.5 Dizinleme türleri	22
2.2.5.1 Bağlam (Concordance) sözlüğü	22
2.2.5.2 Gramatiksel dizin	24
2.2.5.3 Sıkıştırılmış gramatiksel dizin	26
2.2.5.4 Ek dizin	27
2.2.5.5 Sıklık sözlüğü.....	29
Harf sıklığı	30
Ünlü ünsüz harf sıklığı	30
Kalın ince ünlü harf sıklığı.....	31
Düz yuvarlak ünlü harf sıklığı.....	31

Geniş dar ünlü harf sıklığı.....	32
Sedalı sedasız ünsüz harf sıklığı	32
Sürekli süreksiz ünsüz harf sıklığı	32
Hece içi harf sıklığı	33
Hece sıklığı.....	34
Kelime sıklığı	34
Tersten sıralı kelime sıklığı.....	35
Hece tipi sıklığı	35
Kelime tipi sıklığı.....	36
Harf tipi sıklığı	36
3. HAZIRLANAN YAZILIM	37
3.1 Giriş	37
3.2 Mevcut Sistem.....	38
3.3 Problem Belirleme.....	39
3.3.1 Problemler	39
3.3.2 Amaçlar	39
3.3.3 Projenin kapsamı.....	40
3.4 Kullanıcı Durum Modeli	40
3.5 Veritabanı Modeli	41
3.6 Yazılım Aşamaları.....	42
3.7 Yazılım Araçları	43
3.8 Kullanıcı Arayüzleri ve Program Tanıtımı.....	44
3.8.1 Giriş.....	44
3.8.2 Görüntü işleme	45
3.8.3 Ayrıştırma	46
3.8.4 Dizinleme	47
3.8.5 Sıklık analizi	48
3.8.6 Ayarlar	49
3.8.6.1 Font ayarları	49
3.8.6.2 Harf ayarları	50
3.8.6.3 Sıralama ayarları	51
3.8.6.4 Renk ayarları	51
3.8.6.5 Veritabanı ayarları.....	52
3.8.7 Yardım	52
3.8.8 Güvenlik.....	54
3.8.9 Performans analizi.....	57
4. SONUÇ VE ÖNERİLER.....	61
KAYNAKLAR	63
EKLER.....	66

TABLO LİSTESİ

Tablolar

2.1 : Satır numarası ekleme	7
2.2 : İsim soylu kelimeler	9
2.3 : Fiil soylu kelimeler	9
2.4 : Kelimelere anlam özelliğinin eklenmesi	10
2.5 : Kelimelere dil özelliğinin eklenmesi	10
2.6 : Ses olayları bulunan kelimenin ayrıştırılması	10
2.7 : Birleşik fiillerin ayrıştırılması	11
2.8 : Satırların hazırlanması	18
2.9 : Madde başlarının hazırlanması	20
2.10 : Ek dizin hazırlama biçimi	28
2.11 : Harf sıklığı tablosu	30
2.12 : Ünlü ünsüz harf sıklığı tablosu	31
2.13 : Kalın ince ünlü harf sıklığı tablosu	31
2.14 : Düz yuvarlak ünlü harf sıklığı tablosu	31
2.15 : Geniş dar ünlü harf sıklığı tablosu	32
2.16 : Sedalı sedasız ünsüz harf sıklığı tablosu	32
2.17 : Sürekli süreksiz ünsüz harf sıklığı tablosu	33
2.18 : Hece içi harf sıklığı tablosu	33
2.19 : Hece sıklığı tablosu	34
2.20 : Kelime sıklığı tablosu	35
2.21 : Tersten sıralı kelime sıklığı tablosu	35
2.22 : Hece tip sıklığı tablosu	36
2.23 : Kelime tip sıklığı tablosu	36
2.24 : Harf tip sıklığı tablosu	36
3.1 : Yazılım aşamaları tablosu	43
3.2 : Bilgisayar özellikleri ve dizinleme türüne göre performans analizi	57
3.3 : Programlara göre dizinleme türleri performans analizi	58
3.4 : Algoritmalarla göre dizinleme türleri performans analizi	59
3.5 : Programlara göre dizinleme aşamalarında gereken toplam süreler	59
A.1 : Transkripsiyon harfleri	66

ŞEKİL LİSTESİ

Şekiller

2.1 : Ayırıştırma süreci.....	12
2.2 : Concordance yazılımının ekran görüntüsü	14
2.3 : Cibakaya yazılımının ekran görüntüsü	15
2.4 : Dizin hizmet programının ekran görüntüsü	16
2.5 : Metin sıklık çözümleyicisinin ekran görüntüsü	17
2.6 : Dizinleme süreci	21
2.7 : Kümeleme algoritması akış diyagramı	22
2.8 : Heceleme algoritması akış diyagramı	29
3.1 : Dizinleme yazılımı bölümleri	38
3.2 : Kullanıcı durumu modeli seviye - 0.....	40
3.3 : Kullanıcı durumu modeli seviye - 1.....	41
3.4 : Veritabanı modeli.....	42
3.5 : Program giriş form görünümü.....	44
3.6 : Görüntü işleme form görüntüsü	45
3.7 : Ayırıştırma form görünümü	46
3.8 : Dizinleme form görünümü.....	47
3.9 : Sıklık sözlüğü form görünümü	48
3.10 : Program ayarları form görünümü	49
3.11 : Hedef dosya font ayarları form görünümü.....	50
3.12 : Harf ayarları form görünümü.....	50
3.13 : Sıralama ayarları form görünümü	51
3.14 : Renk ayarları form görünümü.....	51
3.15 : Veritabanı düzenleme form görünümü	52
3.16 : Yardım dosyası	53
3.17 : İletişim formu.....	53
3.18 : Kayıt formu	55
3.19 : Reflector program arayüzü.....	56
3.20 : Reflector kod gösterimini engelleme	56
A.1 : Osmanlıca divan ve Transkrip edilmiş metin.....	67

ÖZET

BİLGİSAYAR YARDIMI İLE KİTAP SONU İNDEKSİ HAZIRLANMASI

Kitap sonu indeksleri, kitap içerisinde bulunan kelimeleri biçimsel olarak sınıflandırmak amacıyla hazırlanırlar. Özellikle Türkoloji alanında yaygın olarak kullanılan kitap sonu indeksleri, araştırmacıların kitap ile ilgili bilgilere ulaşmasını kolaylaştırmaktadır.

Yaptığımız çalışma ile kitap sonu indekslerinin bilgisayar ortamında otomatik olarak hazırlanması sağlanmıştır. Tasarlanan algoritmalar ve kullanılan yardımcı programlar yardımıyla kitap sonu indeklerini hızlı ve etkin bir şekilde hazırlayan bir yazılım gerçekleştirilmiştir.

Disiplinler arası bu çalışma, birbirinden bağımsız üç farklı sürece bölünerek hazırlanmıştır. Bu süreçler görüntü işleme, ayrıştırma, dizinleme olarak adlandırılmıştır. Görüntü işleme sürecinde resim dosyalarının metin dosyalarına dönüştürülmesi sağlanmaktadır. Ayrıştırma sürecinde ise mevcut ayrıştırma notasyonu iyileştirilmiş ve hazırlanan bu yeni notasyon ile ayrıştırma işleminin otomatik yapılması sağlanmıştır. Son olarak dizinleme sürecinde, ayrıştırma işlemi tamamlanan metinlerin dizinlenmesi yapılmaktadır. Bu süreç kapsamında mevcut dizinleme türlerine ek olarak yeni dizinleme türleri hazırlanmıştır. Ayrıca çalışmaya dâhil edilen sıklık sözlükleri yardımıyla kullanıcılara daha kapsamlı çalışmalar hazırlama olanağı sağlanmıştır. Böylelikle kapsam genişletilerek kullanıcının farklı çalışmalar yapması hedeflenmiştir.

Sonuç olarak dizin hazırlama konusunda kullanıcıların yükünü oldukça hafifleten bir yazılım ile kullanıcıların daha az sürede ve daha verimli çalışmalar hazırlaması sağlanmıştır.

Anahtar Kelimeler: Dizinleme, Türkoloji, İndeksleme, Gramatiksel Dizin, Sıklık Analizi

SUMMARY

COMPUTER-AIDED PREPARATION OF BACK-OF-THE-BOOK INDEX

The end of the book indexes are prepared to formally classify the words contained in the book. The end of the book indexes that are especially widely used in the field of Turkology, facilitate researchers to access information about the book.

In our study, the end of the book index has been prepared on computer automatically. A software that prepare end of the book indexes in a fast and effective way is implemented with the aid of the designed algorithms and the utility programs used.

This interdisciplinary study is prepared by dividing into three indepented processes. These processes are named as image processing, decomposition and indexing. In the image processing process, the conversion of image files into text files is provided. In the decomposition process, on the other hand, the existent decomposition notation is improved and the decomposition operation is done automatically with the newly prepared notation. Finally, the indexing of text for which the decomposition process has been completed, is obtained in the indexing process. In the context of this final process, new indexing types are prepared in addition to the existent indexing types. Also, with the help of the included “frequncy dictionaries” users are provided the opportunity to prepare more comprehensive studies. Thus, expanding the scope, users are aimed to make different studies.

As a result, with software that fairly alleviates the burden of the users in preparing indexes, users are provided to prepare studies in less time and more efficiently.

Key words: Indexing, Turkology, Concorder, Gramatical Index, Frequency Analysis

1. GİRİŞ

1.1 Tezin Amacı

Dizin bir çalışmadaki ses bilgisi, biçim bilgisi, sözcük bilgisi, cümle bilgisi, anlam bilgisi, sözlük bilgisi, söz varlığı, ağız atlası incelemeleri için gerekli örneklerle ulaşmanın en kolay yoludur. Bu yüzden sözlü metinlerden yazılı metin oluşturulduktan sonra yapılacak ilk iş metinden ayrıştırılmış dizin çalışması yapmaktır.(İleri 2010)

Bir kitabın değeri doğru hazırlanmış bir kitap sonu indeksi ile artmaktadır. Böyle bir indeks, okuyucuya araştırmalarında zaman kazandıracak ve eserin içerik olarak gerçek değerini ortaya koyacak için önemlidir.(Baydur 1988)

Dizinleme yazılımları, dizini yapılan kitapta aranan bilgiye hızlı ve doğru bir şekilde ulaşmayı amaçlar. Geçmişten günümüze dizinleme sürecini ele aldığımızda elle veya bilgisayar yardımıyla dizinleme yapıldığı görülmektedir. Elle yapılan dizinleme çalışmalarında kâğıt üzerinde işlemler gerçekleştirildiğinden dizinde yazım hataları oluşabilmektedir. Yine dizinleme işleminde kullanılan fişlerin kaybolmasına bağlı olarak bilgi eksiklikleri de ortaya çıkabilmektedir.

Bilgisayar ortamında kullanılan dizinleme süreci ile ilgili mevcut programlar incelendiğinde ayrıştırma, dizinleme ve sıklık analizi gibi kullanıcı isteklerinin hepsinin tek bir programda tam olarak karşılanamadığı görülmektedir. Font uyumsuzluğu, transkripsiyon harflerinde ortaya çıkan sorunlar, kullanıcının isteği dışında üretilen sonuçlar ve bunlara bağlı olarak ortaya çıkan zaman kaybı bu tür yazılımların güvenilirliğini ve etkinliğini olumsuz etkilemektedir. (Uçar 2009)

Transkripsiyon harfleri, Türk Latin alfabesinde karşılığı olmayan seslerin gösterilmesi amacıyla kullanılmaktadır.(EK 1.A) Eski Türk dili çalışmalarının çeşitli alfabelerle hazırlanması ve araştırmacıların transkripsiyon harflerini farklı simgelerle göstermesi hazırlanacak yazılımda font ve sıralama hatalarına sebep olabilmektedir. Amaç bu hataları ortadan kaldırarak kullanıcılara serbest çalışma yapabilecekleri bir

ortam sunmaktır. Ayrıca hazırlanacak yazılım sadece Türkiye’de değil tüm dünyada ki Türkologlar tarafından kullanılabilecektir.

Kitap sonu indeklerinin hazırlanmasında en önemli aşamalardan birisi ayrıştırma işlemidir. Bu süreçte metin içerisinde bulunan her bir kelimenin ek ve kök ayrımının belirtilmesi gerekmektedir. Bu ayrımlar belirtilirken çeşitli simgeler kullanılmaktadır. Dizinleme işlemi de bu simgelere göre gerçekleştirilir. Günümüzde ayrıştırma işlemi tamamen elle yapılmaktadır. Bu tez çalışması ile ayrıştırma sürecini otomatik yapan bir yazılım hedeflenmiştir.

Bu tez çalışmasında geliştireceğimiz yazılım ile transkrip edilmiş metinler üzerinde dizin işlemleri yapılırken görülen mevcut hataları ortadan kaldırmak ve kullanıcıya maksimum kullanım kolaylığı sağlayarak sonuca istenilen biçimde ve hızlı erişim sağlamak amaçlanmaktadır. Ayrıca ayrıştırma ve görüntü işleme ile metin tanıma gibi öncül süreçler ve sıklık analizi gibi artçıl süreçler ile eklenecek yeni özellikler sayesinde dizinleme sürecinde gerekli işlemler de ele alınacaktır.

Hazırlanan bu tez, Bilgisayar Bilimleri ve Türk Dili ve Edebiyatı gibi farklı bilim dallarını bir araya getiren disiplinlerarası bir çalışmadır. Disiplinler arası çalışmalar, farklı bilim dallarındaki araştırmacılar tarafından oluşturulur. Amaç bilgi alışverişi ile daha doğru ve kapsamlı sonuçlar elde etmektir. Hazırladığımız tez de bu amaç doğrultusunda oluşturulmuş disiplinlerarası bir çalışmadır.

1.2 Literatür Özeti

Bu tez çalışmasında mevcut yazılımlar ve yayınlanmış kitap sonu indeksleri incelenmiştir. Daha sonra bu çalışmalardan kaynak niteliği taşıyanlar seçilmiş ve bu tez çalışmasındaki yazılımın oluşturulmasında yararlanılmıştır.

Ceval Kaya tarafından hazırlanan Cibakaya dizin programı kitap sonu indeksleri ile ilgili en kapsamlı kaynaktır. Programın kullanımı ile ilgili hazırlanan yardım dosyası ayrıştırma ve dizinleme notasyonu ile ilgili bilgiler içermektedir. Hazırladığımız çalışmada da bu yardım dosyası esas alınarak eklenen yeni özelliklerle Cibakaya dizin programına göre daha kapsamlı bir yazılım oluşturulmuştur.(Kaya 2006)

İsa Sarı tarafından hazırlanan ve internet tabanlı olarak çalışan Dizin Hizmet Programı ayrıştırma işlemi elle yapılan metinlerin dizinlenmesi amacıyla kullanılmaktadır. Ayrıca Fatih Üniversitesinde gerçekleştirilen Metin Sıklık

Çözümleyicisi Web üzerinden çalışmakta olup girdi olarak aldığı metinlerin harf-hece-kelime bazında istatistiklerini hesaplamaktadır.(WEB_1)

Birebir edebi metinlerin dizinlenmesi amacıyla oluşturulmamış olup bu alandaki eksiklikten dolayı kullanılan çeşitli yazılımlar da vardır. Concordance, Text Master, Concorde ve TextStat bu yazılımlardan bazılarıdır. Bu yazılımlar metin içerisinde bulunan kelime ve harf sayıları ile ilgili istatistikler içermektedir.(WEB_2)

İsa Sarı tarafından yazılan ve “Türkoloji Çalışmalarında Teknolojinin Önemi” isimli makale kitap sonu indekslerinin hazırlanmasında bilgisayar yazılımlarının önemini ifade etmektedir. (Sarı 2008) Ayrıca 1988 yılında Doç. Dr. Gülbün Baydur tarafından yayınlanan “Kitap Sonu İndeksleri” adlı makale yazılım alanında olmayıp kitap sonu indekslerinin önemi, hazırlanışı ve Türkiye ile dünyada indeksleme ile ilgili çalışmaları içermektedir. (Baydur 1988)

Sonuç olarak literatüre ve bu konuda yapılmış ticari yazılımlara bakıldığında yapılacak çalışmanın kapsam açısından bir ilk olduğu düşünülmektedir. Cibakaya gibi yardımcı programlar ile indeks hazırlaması yapılırken gereksinimler tam olarak karşılanamamaktadır. Bahsi geçen bu program ayrıştırılmış metinlerin dizinlenmesinde kullanılmaktadır. Yani ayrıştırma işlemi kullanıcı tarafından yapılmaktadır. Ayrıca dizinleme işleminde üretilen sonuçlar kullanıcıların isteklerini karşılayamamaktadır. Bu ise düzeltme işlemlerinde kullanıcılara ek bir iş yükü getirmektedir. Tüm bu üzerinde durduğumuz eksiklikler hatalardan arındırılmış ve çözüme hızlı ulaşmayı hedefleyen bir yazılımı zorunlu hale getirmektedir. Yaptığımız literatür taramasında gördüğümüz önemli bir eksiklik de hem ülkemizde hem de yurtdışında bu alanda yapılan herhangi bir akademik yayın ve projenin bulunmamasıdır.

1.3 Tezin Akışı

Bu çalışma ile kitap sonu indekslerinin hazırlanması konusunda yeni bir yazılım gerçekleştirilmiştir. Geliştirilen algoritmalar ile kullanıcıların iş gücünü en aza indiren bir bilgisayar yazılımı hazırlanmıştır.

Birinci bölümde yapılan çalışma ile ilgili ön bilgiler ve literatür özeti verilmiştir.

İkinci bölümde geliştirilen sistemin özellikleri, mevcut sistemlerin avantaj ve dezavantajları anlatılmıştır. Çalışma kapsamında hazırlanan ayrıştırma, dizinleme ve

sıklık analizlerinin nasıl yapıldığı da bu bölümde kapsamlı olarak ifade edilmiştir. Ayrıca iyileştirilen yeni notasyonun özellikleri ve kullanım bilgileri de yine bu bölümde anlatılmıştır.

Üçüncü bölümde hazırlanan yazılımın tanıtımı yapılmıştır. Ayrıca yazılımın farklı platformlarda performans analizi ve güvenlik konusunda alınan önlemler de bu bölümde anlatılmıştır.

Son bölümde çalışma sonucunda elde edilen verilerin analizi yapılmış ve yapılacak farklı çalışmalar konusunda öneriler sunulmuştur.

2. DİZİNLEME İŞLEMLERİ

2.1 Ayırıştırma

Ayırıştırma, metinde bulunan kelimelerin yapım ve çekim eklerine göre ayrılması işlemidir. Amaç, biçimsel olarak aynı özellikte olan kelimeleri belirlemektir.

Bu tez çalışmasında kelimeler belirli standartlara göre ayrıştırılmaktadır. Bu standartlar sonraki bölümlerde ayrıntılı olarak anlatılacak olan Cibakaya dizin yazılımı esas alınarak belirlenmiştir (Kaya 2006). Yaptığımız çalışmada mevcut notasyon temel alınarak yeni bir notasyon elde edilmiştir.

2.1.1 Ayırıştırmada kullanılan yöntemler

2.1.1.1 Geleneksel (Elle yazılan fişleme yöntemi) ayırıştırma

Elle yapılan geleneksel ayırıştırma yöntemi bilinen en eski modeldir. Bu yöntemde belli standartların olduğu söylenemez. Bu alandaki yazılım eksikliğinden dolayı günümüzde bazı Türkologlar tarafından hala kullanılmaktadır. Model temel olarak kelimelerin bilgisayar ile değil el ile yapılmasını gerektirir.

Geleneksel ayırıştırma modelinde ayırıştırma fişler yardımıyla yapılır. Fiş, kelimelerin üzerine yazıldığı küçük kâğıt parçalarıdır. Ayırıştırma yapılırken herbir kelime ayrı ayrı fişlere yazılır. Kelimeler fişlere yazılırken ek-kök ayrımı yapılarak yazılırlar. Bu şekilde tüm fişlerin hazırlanmasıyla ayırıştırma tamamlanır. Bu aşamalarda herhangi bir bilgisayar yazılımı kullanılmaz.(WEB_3)

Geleneksel ayırıştırma işleminde karşılaşılan en büyük sorun fişlerin kaybolması ve hatalı fişlerin farkedilmesinin güç olmasıdır. Ayrıca geleneksel ayırıştırmanın süreç olarak uzun ve zor olması bu alanda çalışmaların az olmasına sebep olmaktadır.(Sarı 2008)

2.1.1.2 Bilgisayar tabanlı ayırıştırma

Bilgisayar yazılımları kullanılarak yapılan ayırıştırma modelidir. Bu ayırıştırma modelinin de geleneksel modelden pek fazla bir farkı yoktur. Temel fark ayırıştırma işleminde fiş yerine bilgisayarın kullanılmasıdır. Ayırıştırma işleminin yapıldığı

yardımcı yazılımlar ise genel olarak Office programlarıdır. Günümüzde ne yazık ki ayrıştırma yapmak amacıyla hazırlanan bilgisayar yazılımı yoktur. Hazırlanan yazılım ile bu alanda önemli bir eksikliğin giderilmesi amaçlanmaktadır.

Office yazılımları kullanılarak hazırlanan bilgisayar tabanlı ayrıştırma işleminde, bütün metin transkripsiyon harflerinde problem olmayacak bir fontta bilgisayar ortamına elle geçirilir. Yazım işlemi tamamlanan metin manuel olarak bilgisayarda ek ve köklerine ayrılır. Ek-kök ayrımı yapılırken dizinleme aşamasında kullanılacak yazılım referans alınır ve kullanılacak yazılıma göre ayrıştırma gerçekleştirilir.

Günümüzde transkrip edilmiş metinler üzerinde ayrıştırma işlemi için sadece Cibakaya dizin yazılımında özel bir notasyon geliştirmiştir. Diğer yazılımlar Türkoloji alanında kitap sonu indeklerinin oluşturulması amacıyla hazırlanmadığından programların ürettiği sonuç üzerinden ek işlemler yapmak gerekmektedir.

Cibakaya’da dizin hazırlayabilmek için isim ve fiil soylu kelimeler farklı karakterler kullanılarak ayrıştırılır. Ayrıca sesteş kelimeleri belirlemek ve kelimelerin dil özelliğinin eklenmesi ile ilgili de farklı karakterler kullanılmaktadır.(Kaya 2006) Cibakaya dizin yazılımı bu alanda en yaygın olarak kullanılan yazılım olmasından hazırladığımız yazılımda Cibakaya dizin programının kullandığı notasyonu referans aldık. Bu notasyon ayrıntılı olarak yeni eklediğimiz özelliklerle beraber sonraki bölümde anlatılacaktır.

Bilgisayar yazılımları kullanılarak yapılan ayrıştırmada her ne kadar bilgisayar kullanılsa da işlemler manuel yapıldığından kullanıcının hata yapması ve ayrıştırma işleminin uzun zaman alması bu modelin olumsuzlukları arasında sayılabilir. Kullanıcı taleplerini karşılayan bir yazılımın gerçekleştirilmesi bu tür olumsuzlukları ortadan kaldıracaktır.

Ayrıştırma işleminin tamamen bilgisayar tabanlı yapılması mümkün değildir. Çünkü aynı şekilde fakat farklı özellik ve anlamda olan yüzlerce kelime vardır. Ayrıca kelimenin zaman içerisinde şekilsel özelliğinin değişmesi bu alanda hazırlanacak bir yazılımın kesin sonuçlar elde etmesini engellemektedir. Hazırladığımız yazılım kapsamında geliştirilen bazı yöntemler ile kullanıcının iş yoğunluğu azaltılmaktadır. Örneğin aynı özellikte olan ve metin içerisinde tekrar eden kelimelerin otomatik

ayrıştırılması ve kelimelerin kitap içerisindeki konumlarını belirleyen satır numaralarının otomatik eklenmesi kullanıcı yükünü %70 oranında azaltmaktadır.

Ayrıca ayrıştırılan kelimelerin veritabanına kaydedilerek kullanıcıya özel bir kelime veritabanının oluşturulması sağlanmaktadır. Böylelikle kullanıcılara özel çözümler içeren bir ortam hazırlanmıştır.

2.1.2 Ayrıştırma notasyonu

2.1.2.1 Satır numarası ekleme

Satır numaraları kelimelerin metin içerisindeki konumlarını belirlemek amacıyla oluşturulurlar. Genel olarak her bir satır numarası, satırın bulunduğu yaprak, sayfa ve satır numarasını içerir. Satır numaraları belli standartlara göre verilir. Satır numarası hazırlanırken “/” karakterinin satır numarasının içerisinde kullanılması gerekmektedir. Bu karakter satır numarası için ayırt edici özelliktir. Tablo 2.1’de örnek satır numaraları ve bu satır numaralarına ait açıklamalar bulunmaktadır.

Tablo 2.1 : Satır numarası ekleme

001a/01	1. yaprak, yaprağın sol yüzü, 1. satır
045b/09	45. yaprak, yaprağın sağ yüzü, 9. satır
046/23	46. yaprak 23. satır (yüz(sayfa) yok)

Satır numaraları el ile teker teker verilebilirken otomatik olarak da eklenebilir. Mevcut yazılımlarda bulunmayan bu özellik ile kullanıcılar satır numarası eklemek gibi sıkıcı bir işten kurtarılmıştır. Ayrıca bu özellik kullanıcıya ayrıştırma sürecinde zaman da kazandırmaktadır.

Satır numaraları yayın haline getirilmiş kitaplara göre belirlenmektedir. Her bir kitabın sayfa ve satır sayılarının farklı olması ve bazı satırların tablo, şekil yâda boş olması otomatik satır ekleme işleminde karşılaşılan problemlerdir. Bu problemler geliştirilen notasyon ile ortadan kaldırılmıştır.

Satır numarasının otomatik eklenmesi basit birkaç adım ile gerçekleştirilebilir. Yazılım kullanılırken sayfa sonunun ve boş satırların belirtilmesi gerekmektedir. Sayfa sonunu belirtmek için “#” karakteri kullanılmaktadır. Ayrıca boş satırları ifade

etmek için ise satırın boş bırakılması yeterlidir. Böylelikle uzun ve sıkıcı bir süreç olan satır ekleme işlemi basit birkaç adım ile gerçekleştirilebilir.

Aşağıda örnek bir metin ve hazırlanan yazılım kullanılarak elde edilmiş sonuçlar görülmektedir. Metin içerisinde boş bırakılan satırlar kitapta metin bulunmayan satırları ifade etmektedir.

Metin

san.sız Һamd

bir Һūdā.yım-ğa

biz.ni yok.dın bar

ķıl-dı ‘aķl ve ‘imān

#

#

biz ķul.lar-ğa yār ķıl-dı

ve taķı nihāyet.siz

#

dürūd Һazret_i MuҺammed

Satır numarası eklenmiş metin

001a/01 san.sız Һamd

001a/03 bir Һūdā.yım-ğa

001a/06 biz.ni yok.dın bar

001a/07 ķıl-dı ‘aķl ve ‘imān

002a/01 biz ķul.lar-ğa yār ķıl-dı

002a/02 ve taķı nihāyet.siz

002b/01 dürūd Һazret_i MuҺammed

2.1.2.2 Çekim eklerinin ayrılması

Ayrıştırma işleminde herbir kelimenin ek ve kök ayrımı yapılırken, kelimenin isim veya fiil soylu kelime olduğu belirtilmelidir. Bu fark ayrıştırma esnasında kullanılan çeşitli karakterler yardımıyla belirlenmektedir.

Herhangi bir iş, oluş veya hareket bildirmeyen sözcüklere isim denir. Ayrıştırma yapılırken isim soylu sözcükle “.” simgesi kullanılarak belirtilir ve simgeden sonra sadece kelimenin çekim ekleri eklenir. Örneğin “gözlükçüde” kelimesi ayrıştırılırken “göz.lükçüde” değil, “gözlükçü.de” şeklinde olmalıdır. Tablo 2.2’de isim soylu kelimelerin ayrıştırılması ile ilgili örnekler verilmiştir.

Tablo 2.2 : İsim soylu kelimeler

Karanlıkta	Karanlık.ta
Gürültüde	Gürültü.de
Kitaplıkta	Kitaplık.ta

İş, oluş veya hareket bildiren sözcüklere fiil denir. Fiil soylu sözcükler ayrıştırılırken “-” simgesi kullanılır. Tablo 2.3’te fiil soylu kelimelerin ayrıştırılması ile ilgili örnekler verilmiştir.

Tablo 2.3 : Fiil soylu kelimeler

Gidiyordu	Gid-iyordu{git}
Bekleyecekmiş	Bekle-yecekmiş
Kırılmıştı	Kırıl-mıştı

2.1.2.3 Sesteş kelimelerin ayrıştırılması

Sesteş kelimeler yazılışları veya okunuşları aynı, anlamları farklı olan kelimelerdir. Geliştirilen notasyon sadece sesteş kelimelerin değil bütün kelimelerin anlamlarını belirtmek için kullanılabilir. Ayrıştırma işleminde kelimelerin anlamları “()” simgeleri kullanılarak belirtilebilir. Tablo 2.4’te ayrıştırma esnasında kelimelerin anlamlarının belirtilmesi örneklerle gösterilmiştir.

Tablo 2.4 : Kelimelere anlam özelliğinin eklenmesi

Yüzdü	Yüz-dü(Havuzda yüzmek)
Yüzdü	Yüz-dü(Deri yüzmek)
Yüzdü	Yüz.dü(Sayı)

2.1.2.4 Kelimenin dil özelliğinin eklenmesi

Diller birbirlerinden etkilenererek yapancı kelimeler içerebilmektedirler. Türkçe gerek bulundu coğrafya gerekse tarih boyunca farklı bölgelerde kullanılmasından diğer dillerden kelimeler barındırmaktadır. Ayırıştırma sürecinde metin içerisinde bulunan kelimelerin hangi dile ait olduğunu belirtmek mümkündür. Dil özelliğinin belirtilmesi için “[]” simgeleri kullanılmaktadır. Tablo 2.5’te kelimelerin dil özelliğinin nasıl eklendiği örneklerle açıklanmıştır. Örneklerde de görüldüğü gibi dil özelliği ile beraber anlam özelliği de eklenebilmektedir.

Tablo 2.5 : Kelimelere dil özelliğinin eklenmesi

Pardon	Pardon(Özür)[FR]
Muhabbetle	Muhabbet.le(Sevgi)[AR]

2.1.2.5 Ses olayları bulunan kelimenin ayrıştırılması

Türkçe de kökte ya da ekte değişmelere sebep olabilen ses olayları bulunmaktadır. Bu değişimler metin içerisinde belirtilmediği durumda hazırlanan dizinde hatalar oluşabilmektedir. Yaptığımız çalışmada ses olaylarıyla meydana gelen değişimleri belirtmek için “{ }” karakterleri kullanılmaktadır. Tablo 2.6’da ses olaylarıyla değişime uğrayan kelimelerin ayrıştırılması örneklerle açıklanmıştır.

Tablo 2.6 : Ses olayları bulunan kelimenin ayrıştırılması

Kitab.ı	Kitab.ı{Kitap}
Tarağ.a	Tarağ.a{Tarak}
Göğs.üne	Göğs.üne{Göğüs}
Akl.a	Akl.a(Düşünme)[AR]{Akıl}

Tablo 2.6’da da görüldüğü gibi birden fazla özellik bir kelime için belirtilebilir. Özellikler belirtilirken simgelerin sıralanışı önemli değildir. Yazılım tüm özellikleri hatasız değerlendirebilmektedir.

2.1.2.6 Birleşik fiillerin belirtilmesi

Birleşik fiil, bir eylemi karşılamak üzere isim veya fiile bir yardımcı fiilin getirilmesiyle yapılan kelime grubudur. Kıl-, ir-, turur-, er- ... birleşik fiillere örnek olarak sayılabilir.(WEB_4)

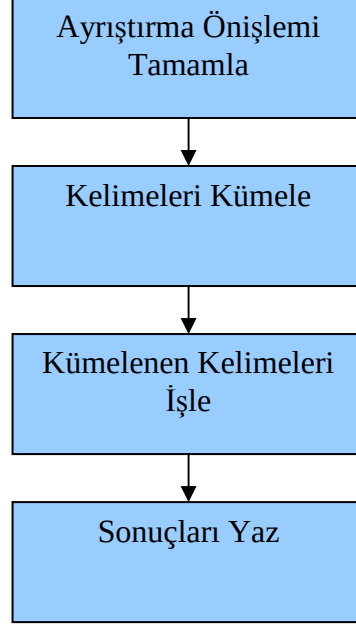
Birleşik fiillerin kullanıcı isteğine bağlı olarak dizin içerisinde kelime grubu olarak gösterilmesi mümkündür. Bu işlemin gerçekleştirilebilmesi için kullanıcı tarafından birleşik fiillerin ayrıştırma esnasında gösterilmesi gerekir. Hazırlanan yazılım içerisinde bazı birleşik fiiller üretici tarafından eklenmiştir. Kullanıcı kayıtlı olan birleşik fiillere ekleme, çıkarma veya düzenleme yapabilir. Ayrıca isteğe bağlı olarak ayrıştırma esnasında da birleşik fiilleri belirtilebilir. Ayrıştırma esnasında birleşik fiilleri belirtmek için “*” karakteri kullanılır. Tablo 2.7’de örneklerle birleşik fiillerin ayrıştırılması gösterilmiştir.

Tablo 2.7 : Birleşik fiillerin ayrıştırılması

Tururdu	*Turur-du
Kılacak	*Kıl-acak

2.1.3 Ayırıştırma süreci

Hazırlanan yazılımda ayırıştırma süreci Şekil 2.1 de görüldüğü gibi dört aşamada yapılmaktadır.



Şekil 2.1 : Ayırıştırma süreci

Ayırıştırma önışlem aşamasında metin kelimelere ayrılır. Ayrıca veritabanında kayıtlı kelimeler de ayırıştırma yapmak amacıyla kopyalanır.

Kelimelerin kümelenmesi aşamasında ayırıştırılmış kelimeler ayırıştırma notasyonuna göre şekil, özellik ve anlamlarına göre gruplandırılır. Diğer kelimeler ise ayırıştırılmak üzere farklı bir alana kopyalanır. Kelimelerin kümelenmesini sağlayan algoritmaya ait akış diyagramı Şekil 2.7’de gösterilmiştir.

Kümelenen kelimelerin işlenmesi aşamasında ayırıştırılmayı bekleyen kelimeler, veritabanından kopyalanan kelimelerle ayırıştırılmış kelimelerin kullanılmasıyla ayırıştırılır.

Sonuçların yazılması aşamasında kullanıcının belirlediği seçeneklere göre ayırıştırılmış kelimelerin ekrana yazdırılmasıdır.

2.2 Dizinleme

2.2.1 Dizinleme nedir

Dizinleme, ayrıştırma işlemi tamamlanan kelimelerin alfabetik olarak bulundukları sayfa ve satır numaralarıyla beraber sıralanmasıdır. Dizinleme işleminde amaç, madde başı olarak isimlendirilen herbir kelime kökünün alfabetik olarak sıralanması ve aynı kök ve özelliklere sahip kelimelerin kendi madde başlarının altında gösterilmesidir. İşlemlerin doğru gerçekleştirilmesi ayrıştırılan metnin hatasız olmasına bağlıdır.

Dizinleme işleminde çeşitli yöntemler kullanılmaktadır. Bu yöntemler kabaca geleneksel dizinleme ve bilgisayar tabanlı dizinleme olarak ikiye ayrılmaktadır.

Geleneksel dizinleme, geleneksel yöntemlerle ayrıştırılmış metinlerde kullanılır. Bu yöntemde literatür değeri taşıyan sabit kurallar bulunmamaktadır. Ayrıştırma esnasında hazırlanan fişlerin madde başlarına göre gruplandırılması ve gruplanan madde başlarının alfabetik olarak sıralanması ile dizinler hazırlanmaktadır. Geleneksel dizinleme yöntemi, uzun zaman almakla beraber fişlerin kaybolmasına bağlı olarak hatalı dizinler oluşmasına sebep olmaktadır.(Sarı 2008)

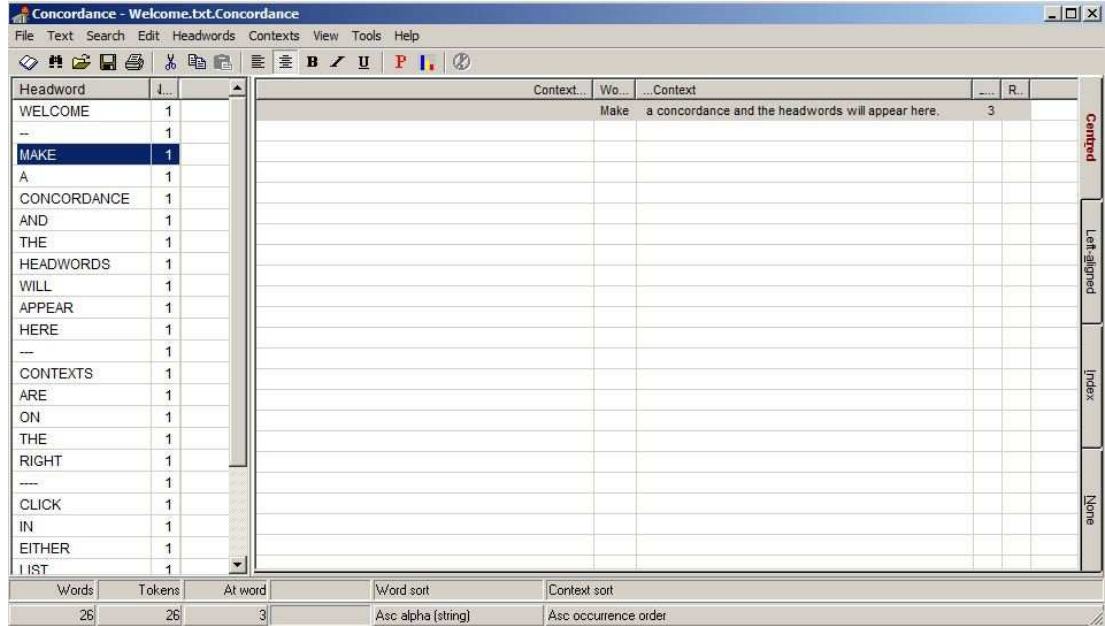
Bilgisayar tabanlı dizinleme yönteminde ise dizinleme işlemini gerçekleştirmek amacıyla bazı yazılımlar geliştirilmiştir. Bu yazılımlar kullanıcının isteklerini tam olarak karşılayamamasına rağmen bu alanda hazırlanan etkin bir yazılımın olmamasından hala yaygın olarak kullanılmaktadırlar.

2.2.2 Dizinleme yazılımları

2.2.2.1 Concordance

Concordance, kelimelerin ayrıştırılmasına bakmaksızın aynı kelimeleri satır numaralarına göre kendi içerisinde guruplayan ve yurt dışında üretilen ücretli bir yazılımdır. Bu yazılımda metinlerin ayrıştırılması sonuç üzerinde herhangi bir değişiklik oluşturmamaktadır. Herbir kelime ek veya köklerine bakılmaksızın grublandırılır. Program metin dosyalarını(*.txt) girdi olarak kabul etmektedir. Bu da transkripsiyon harflerinde font hatalarının oluşmasına sebep olmaktadır. Program madde başlarını alfabetik gruplarken kelimelerin geçtiği satırların tamamını madde

başlarının altında gruplamaktadır. Ayrıca kelimelere ait istatistiksel bilgileride hazırlamaktadır. Yazılım ekran görüntüsü Şekil 2.2’de görülmektedir.

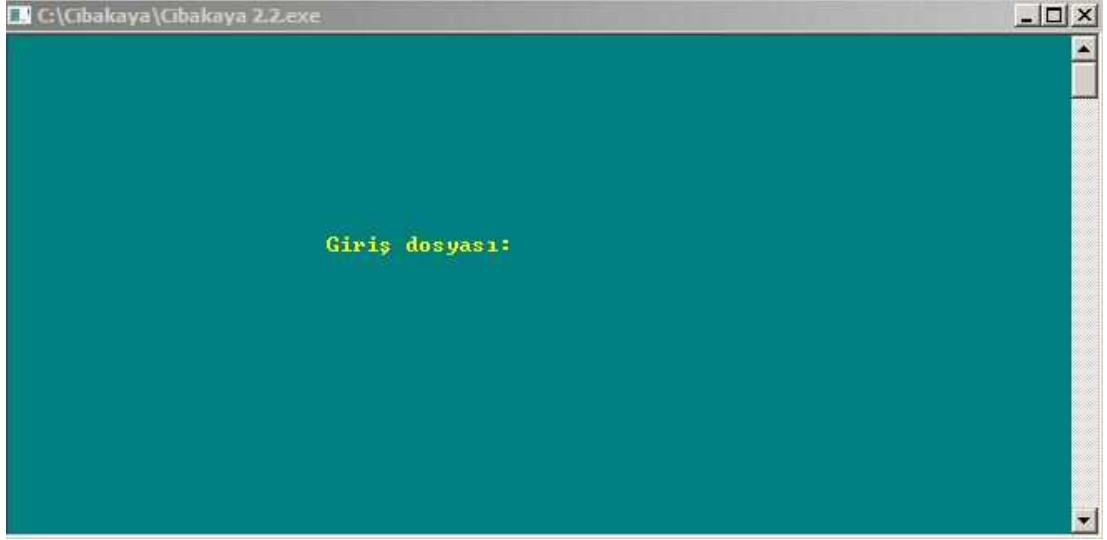


Şekil 2.2 : Concordance yazılımının ekran görüntüsü

Concordance yazılımı kelimelerin sıralanması ve satır içerisindeki konumunu güzel bir şekilde belirtmesine rağmen literatürde kabul edilen ek-kök ayrımını gerçekleştiremediğinden kullanıcı isteğini yansıtan sonuç üretememektedir. Kullanıcılar dizin işlemini tamamlayabilmek için Concordance yazılımının hazırladığı sonuç üzerinde herbir madde başının ayrıştırmasını tamamlamalıdır. Bu da kullanıcılara ek iş getirmekle beraber bu yazılıma olan güveni ve ilgiyi azaltmaktadır.(WEB_5)

2.2.2.2 Cibakaya

Cibakaya yazılımı Ceval Kaya tarafından geliştirilmiştir. Yazılım Türkoloji alanında dizin hazırlamak amacıyla oluşturulmuş olup ayrıştırma işlemi tamamlanmış metinler ile çalışır. Concordance gibi sadece metin dosyaları (*.txt) ile çalışmaya izin verir. Bu nedenle font hataları oluşmaktadır. Yazılım kişisel bilgisayarda çalışacak şekilde tasarlanmıştır. Kullanımı oldukça basit olan yazılım sadece giriş dosyasının alınmasını sağlayan basit bir kullanıcı arayüzünden oluşmaktadır.(WEB_6) Şekil 2.3’de örnek ekran görüntüsü görülmektedir.

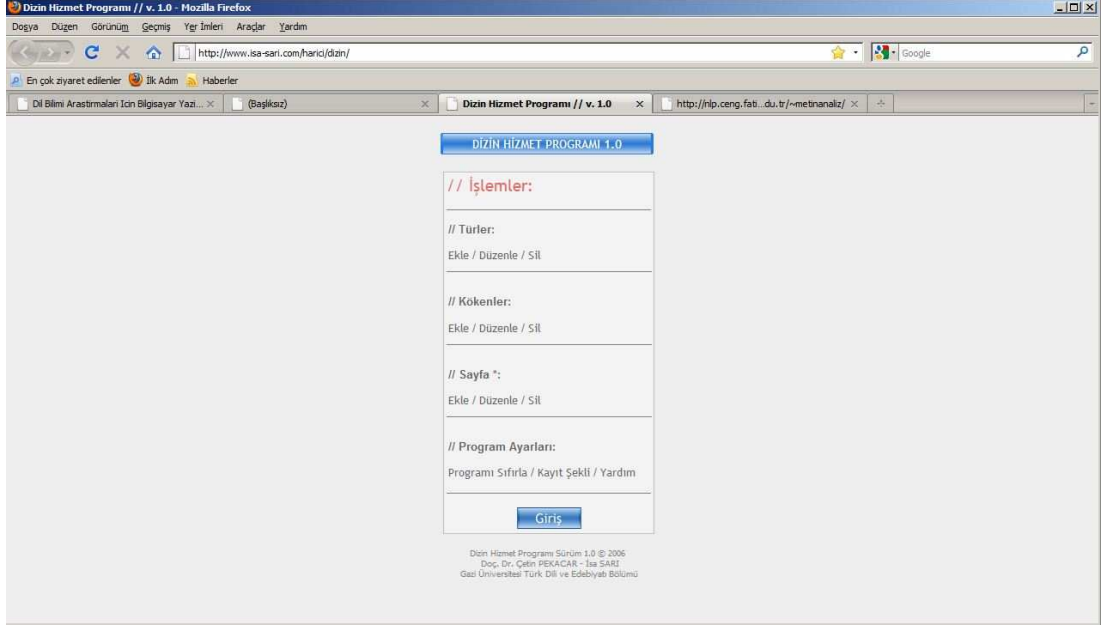


Şekil 2.3 : Cibakaya yazılımının ekran görüntüsü

Cibakaya yazılımında ayrıştırma özelliği bulunmamaktadır. Kullanıcılar dizin hazırlayacakları metinlerin ayrıştırma işlemini Office yazılımları ile manuel yapmak zorundadırlar. Ayrıca yalnız metin(*.txt) dosyaları ile çalışmaya izin verdiği için transkripsiyon harflerinde problemler oluşmaktadır. (Kaya 2006)

2.2.2.3 Dizin hizmet programı

Dizin hizmet programı internet tabanlı bir uygulamadır. Kullanılan veritabanı yardımıyla kelimelerin dizinlenmesi sağlanmaktadır. Veritabanı sürekli yeni kelimelerle zenginleşmektedir. Bu işlem bazı olumsuzlukları da beraberinde getirmektedir. Bu olumsuzluklara kelimelerin artmasına bağlı olarak performansın düşmesi ve yanlış kelimelerin kaydedilmesi sonucunda veritabanının bozulması sayılabilir. Şekil 2.4'te programın örnek ekran görünümü verilmiştir.(WEB_7)

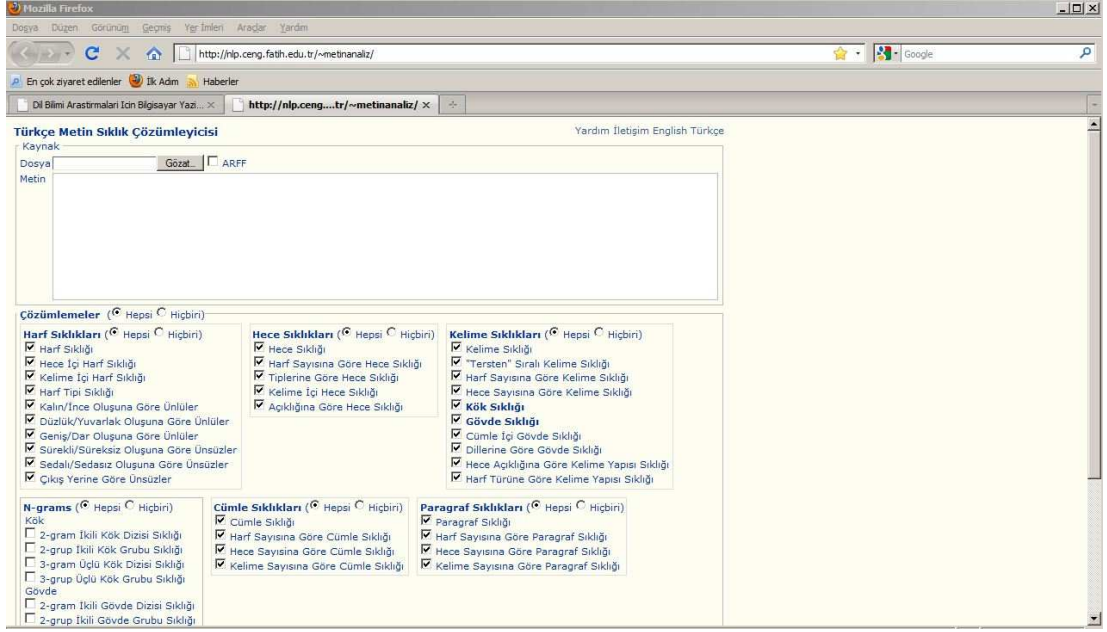


Şekil 2.4 : Dizin hizmet programının ekran görüntüsü

Geliştirdiğimiz yazılımda kullanıcı isteğine göre ayrıştırılan kelimelerin kaydedilmesi ve veritabanı üzerinde düzenlemelerin yapılması sağlanmıştır. Bu sayede daha sağlıklı sonuçlar elde edilmektedir. Ayrıca yazılım internet tabanlı olmadığı için performans internet tabanlı uygulamalara göre daha iyi olduğu görülmüştür.

2.2.2.4 Türkçe metin sıklık çözümleyicisi

İnternet tabanlı bir uygulamadır. İçerdiği sıklık sözlükleri yardımıyla metnin harf, hece, kelime ve satır bakımından detaylı istatistiğini hesaplamak amacıyla hazırlanmıştır. İnternet tabanlı bir uygulama olduğundan yüksek boyuttaki metin dosyalarında uzun zaman gerekmektedir. Ayrıca font uyuşmazlığına bağlı olarak transkripsiyon harflerinde hatalar oluşmaktadır. Şekil 2.5'te programın örnek ekran görünümü verilmiştir.(WEB_1)



Şekil 2.5 : Metin sıklık çözümleyicisinin ekran görüntüsü

Geliştirdiğimiz yazılım 14 ayrı sıklık sözlüğü yardımıyla kısa sürede hatasız sonuçlar vermektedir. Ayrıca metin sıklık çözümleyicisinde oluşan font hataları büyük oranda çözülmüştür.

2.2.2.5 Diğer yazılımlar

Birebir edebi metinlerin dizinlenmesi amacıyla oluşturulmamış olup bu alandaki eksiklikten dolayı kullanılan çeşitli yazılımlar da vardır. Text Master, Concorde ve TextStat bu yazılımlardan bazılarıdır. Concordance ve Cibakaya dizin yazılımları bu yazılımlara göre çok daha iyi sonuçlar vermektedir.

Bahsi geçen tüm yazılımlarda gerek ayrıştırma işleminde gerekse dizinleme sonucu oluşan metindeki düzeltmeler oldukça fazla zaman gerektirmektedir. Font hataları, sıralama hataları ve birleşik kelime gruplarında oluşan hatalar kullanıcılara zaman kaybı yaşatmaktadır.

2.2.3 Dizinleme notasyonu

2.2.3.1 Satırların hazırlanması

Dizin içerisinde bulunan ve madde başı olarak isimlendirilen her bir kelimenin metin içerisindeki konumunun beirlenmesi dizin hazırlama işleminin en önemli hedeflerinden birisidir. Madde başlarının metin içerisindeki konumları belirlenirken

madde başının altında göstermek üzere satırlar hazırlanır. Hazırlanan satırlar genel olarak ilgili madde başının sayfa ve satır numarasını ve aldığı ekleri içerir.

Satırlar hazırlanırken madde başının ilk harfi, madde başının özelliğine göre eklenecek karakter, madde başının metin içerisinde aldığı çekim ekleri ve son olarak madde başının metin içerisinde hangi sayfa ve satırda bulunduğunu belirten satır numarası eklenerek hazırlanır. Satır içerisinde madde başı olan kelimeler gösterilirken fiil soylu kelimeler “.-”, isim soylu kelimeler ise “.+” karakter gurupları kullanılarak hazırlanır.(Kaya 2006)

Tablo 2.8’da satırların hazırlanması ile ilgili örnekler verilmiştir.

Tablo 2.8 : Satırların hazırlanması

duvar.da	d.+da
gel-iyordu	g.-iyordu
Kitab_1	K.-1

Hazırladığımız yazılımda satırların hazırlanması örnek bir metin üzerinde aşağıda görülmektedir. Bu örnekte satırlar bütün olarak alınmıştır. İsteğe bağlı olarak farklı türlerde de dizin hazırlanabilir. İlerde dizinleme türleri ile ilgili detaylı bilgi verilecektir.

Örnek Dizin:

Metin

001a/01 Vay al duvağ.ımın{duvak} sahib.i{sahip}
001b/01 Vay aln.ımın{alın} baş.ımın umud.u{umut}
001a/02 Vay şah yiğid.im{yiğit} şahbaz yiğid.im{yiğit}
001b/02 Doy-uncaya dek yüz.üne(uzuv) bak-amadığım

Sonuç

al-1- [Tr]

Vay a. duvağımın sahibi 001a/01

aln-1- [Tr]

Vay a.+ımın başımın umudu 001b/01

bak - -1- [Tr]

Doyuncaya dek yüzüne b.-amadığım 001b/02

Baş -1- [Tr]

Vay alnımın b.+ımın umudu 001b/01

dek -1- [Tr]

Doyuncaya d. yüzüne bakamadığım 001b/02

Doy - -1- [Tr]

D.-uncaya dek yüzüne bakamadığım 001b/02

duvak -1- [Tr]

Vay al d.+ımın sahibi 001a/01

sahip -1- [Tr]

Vay al duvağımın s.+i 001a/01

şah -1- [Tr]

Vay ş. yiğidim şahbaz yiğidim 001a/02

Şahbaz -1- [Tr]

Vay şah yiğidim ş. yiğidim 001a/02

Umut -1- [Tr]

Vay alnımın başımın u.+u 001b/01

Vay -3- [Tr]

V. al duvağımın sahibi 001a/01 , 001a/02 , 001b/01

Yiğit -2- [Tr]

Vay şah y.+im şahbaz y.+im 001a/02

yüz : uzuv -1- [Tr]

Doyuncaya dek y.+üne bakamadığım 001b/02

2.2.3.2 Madde başlarının hazırlanması

Dizin içerisinde aynı kök ve özellikte olan kelimeleri sınıflandırmak amacıyla madde başı olarak isimlendirilen kelimelerin çekim eki hali almamış durumları kullanılır. Bir madde başı diğer madde başlarından şekil, özellik veya anlam bakımından farklı olmalıdır. Kelimeler ayrıştırma işleminde bu özelliklere göre farklı notasyonlarla ifade edilir. Bu notasyonlar önceki bölümlerde anlatılmıştır.

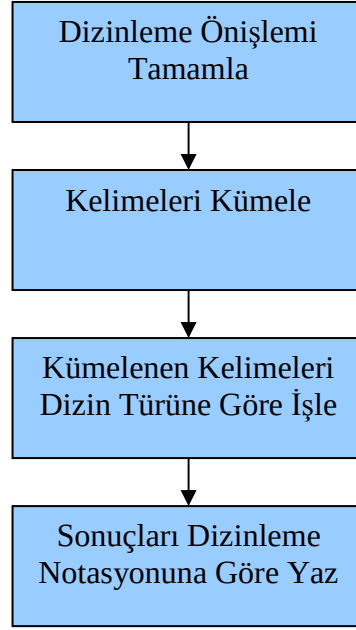
Madde başı olan kelimeler isim veya fiil soylu kelime olmalarına göre iki farklı şekilde belirtilirler. İsim soylu kelimeler madde başı olarak ifade edilirken herhangi bir karakter almazlar. Fakat fiil soylu kelimeler madde başı olarak ifade edilirken “-” karakteri kullanılır. Sesteş kelimeler madde başı ile beraber “:” karakteri kullanılarak ayrılır. Ve anlamları “:” karakterinden sonra madde başının yanına eklenir. Kullanıcılar isteğe bağlı olarak madde başı olan kelimelerin adetlerini ve hangi dile ait olduğunu madde başının yanına otomatik ekleyebilmektedirler. Tablo 2.9 madde başlarının hazırlanması ile ilgili örnek açıklamalar içermektedir.

Tablo 2.9 : Madde başlarının hazırlanması

Ben	<i>İsim soylu</i> anlamı belirtilmemiş madde başı
Ben : Leke	<i>İsim soylu</i> anlamı <i>leke</i> olan madde başı
Yaz : Mevsim	<i>İsim soylu</i> anlamı <i>mevsim</i> olan madde başı
Yaz - : Yazmak	<i>Fiil soylu</i> anlamı <i>yazmak</i> olan madde başı
İns:İnsan -3- [AR]	<i>İsim soylu</i> , anlamı <i>insan</i> olan, metin içerisinde 3 farklı yerde bulunan <i>Arapça</i> bir madde başı

2.2.4 Dizinleme süreci

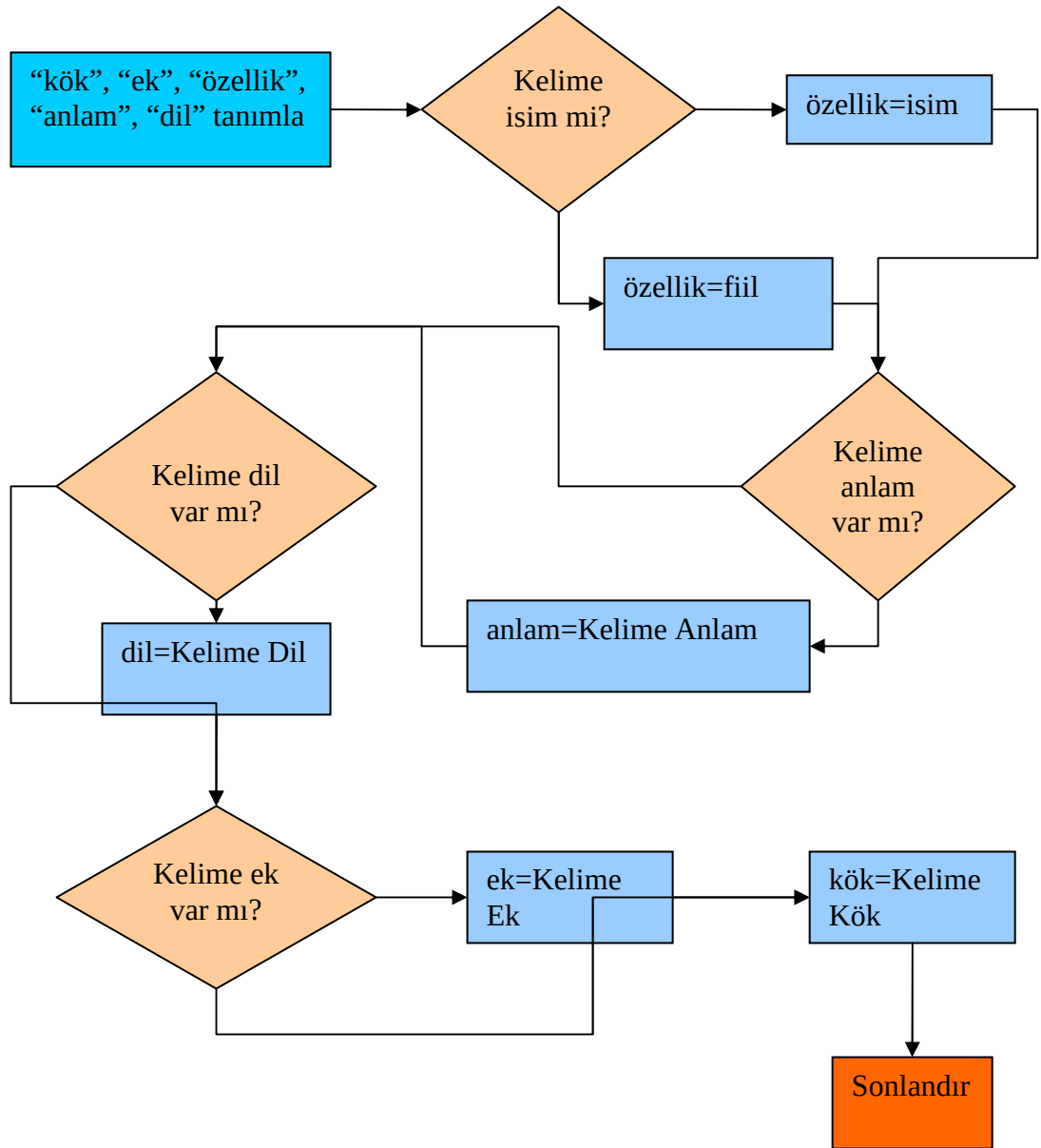
Hazırlanan yazılımda dizinleme süreci Şekil 2.6 de görüldüğü gibi dört aşamada yapılmaktadır.



Şekil 2.6 : Dizinleme süreci

Dizinleme önişlem aşamasında metin kelimelere ayrılır ve kelimelerin ayrıştırma standartlarına uygun olarak hazırlanıp hazırlanmadığı incelenir. Herhangi bir problem olması durumunda süreç durdurularak hatalı durum kullanıcıya düzeltmesi için sunulur.

Kelimelerin kümelenmesi aşamasında, tüm kelimeler ayrıştırma notasyonuna göre şekil, özellik ve anlamlarına göre gruplandırılır. Kelime özelliklerinin belirlenmesini sağlayan algoritmaya ait akış diyagramı Şekil 2.7’de gösterilmiştir.



Şekil 2.7 : Kümeleme algoritması akış diyagramı

Kümelenen kelimelerin işlenmesi aşamasında, seçilen dizin türüne göre kelimeler dizin oluşturacak şekilde hazırlanır.

Sonuçların dizinleme notasyonuna göre yazılması aşamasında ise kullanıcının belirlediği seçeneklere göre sonuç ekrana yazdırılır.

2.2.5 Dizinleme türleri

2.2.5.1 Bağlam (Concordance) sözlüğü

Dizin türleri içerisinde en kapsamlı dizinleme türüdür. Kelimeler madde başlarının altında ifade edilirken madde başının bulunduğu satırın tamamı dizine eklenir. Her

bir satır alt alta gelecek şekilde madde başının aldığı çekim ekinin alfabetik sıralamasına göre gruplandırılır. Bu dizin türünde aynı çekim eklerine sahip madde başları metin içerisinde farklı satırlarda bulunuyorsa dizin içerisinde farklı satırlarda ifade edilir. Böylelikle madde başları daha detaylı incelenebilir. Bağlam sözlüğü hazırlanırken diğer sözlük türlerinde olduğu gibi madde başları alfabetik olarak sıralanır. Metin içerisinde aynı satırda birden fazla sayıda madde başı bulunursa bu satır sadece bir sefer madde başının altında ifade edilir. Örnek metin ve program kullanılarak elde edilen sonuç aşağıda görülmektedir.

Metin:

1b/2 san.sız ẖamd bir H̱uḏā.yım̱ğa biz.ni yok.dın bar kıldı ‘aql ve ‘im̱ṉ

1b/3 biz ḵul.laṟğa yār kıldı ve taqı niẖayet.siz ḏüṟd H̱azret_i Muḥammed

Bağlam Sözlüğü:

‘aql

sansız ẖamd bir H̱uḏāyım̱ğa bizni yokdın bar kıldı ‘a. ve ‘im̱ṉ 1b/2

‘im̱ṉ

sansız ẖamd bir H̱uḏāyım̱ğa bizni yokdın bar kıldı ‘aql ve ‘i. 1b/2

bar

sansız ẖamd bir H̱uḏāyım̱ğa bizni yokdın b. kıldı ‘aql ve ‘im̱ṉ 1b/2

bir

sansız ẖamd b. H̱uḏāyım̱ğa bizni yokdın bar kıldı ‘aql ve ‘im̱ṉ 1b/2

biz

b. ḵullaṟğa yār kıldı ve taqı niẖayetsiz ḏüṟd H̱azret_i Muḥammed 1b/3

sansız ẖamd bir H̱uḏāyım̱ğa b.+ni yokdın bar kıldı ‘aql ve ‘im̱ṉ 1b/2

ḏüṟd

biz ḵullaṟğa yār kıldı ve taqı niẖayetsiz d. H̱azret_i Muḥammed 1b/3

ẖamd

sansız ẖ. bir H̱uḏāyım̱ğa bizni yokdın bar kıldı ‘aql ve ‘im̱ṉ 1b/2

H̱azret

biz ḵullaṟğa yār kıldı ve taqı niẖayetsiz ḏüṟd H̱.+_i Muḥammed 1b/3

H̱uḏā

sansız ẖamd bir H̱.+yım̱ğa bizni yokdın bar kıldı ‘aql ve ‘im̱ṉ 1b/2

ḵıl -

sansız Һamd bir Һүдәймға бизни жоқдн бар қ.-дн ‘ақл ve ‘имән 1b/2
biz қullarға yār қ.-дн ve тақн nihāyetsiz дүрүд Һазрет_и Muһammed 1b/3

қул

biz қ.+larға yār қылды ve тақн nihāyetsiz дүрүд Һазрет_и Muһammed 1b/3

Muһammed

biz қullarға yār қылды ve тақн nihāyetsiz дүрүд Һазрет_и M. 1b/3

nihāyet

biz қullarға yār қылды ve тақн n.+siz дүрүд Һазрет_и Muһammed 1b/3

san

s.+siz Һamd bir Һүдәймға бизни жоқдн бар қылды ‘ақл ve ‘имән 1b/2

тақн

biz қullarға yār қылды ve t. nihāyetsiz дүрүд Һазрет_и Muһammed 1b/3

ve

sansız Һamd bir Һүдәймға бизни жоқдн бар қылды ‘ақл v. ‘имән 1b/2
biz қullarға yār қылды v. тақн nihāyetsiz дүрүд Һазрет_и Muһammed 1b/3

yār

biz қullarға y. қылды ve тақн nihāyetsiz дүрүд Һазрет_и Muһammed 1b/3

yoқ

sansız Һamd bir Һүдәймға бизни y.+дн бар қылды ‘ақл ve ‘имән 1b/2

2.2.5.2 Gramatiksel dizin

Шәкил ve өзәллик оларак айнн олан кәлимәләрин мәтин ичәрисиндә булуңдуқларн сәтур илә бәрәбәр мәддә бәшннн әлтүндә гөстәрилмәсинә сағлаған дзин түрүдүр. Мәддә бәшларн алфәбәтик сараланнр. Бәғлам сөзлүғүндән фаркн мәддә бәшларннн әлтүндә сәтурлар һазрланнркен мәтин ичәрисиндә мәддә бәшннн гәчтиғи түм сәтурлар гөстәрилмәз. Айнн чәким әкләринә әлмиш олан мәддә бәшларн шәкил itibariyle айндыр. Бу кәлимәләрин мәтин ичәрисиндә булуңдуғу түм сәтурлар yеринә әрнәк оларак ilk гелән сәтур мәддә бәшннн әлтүндә гөстәрилр ve диғер сәтурларда сәтур numarasн илә ilgili сәтурнн yаннда гөстәрилр. Әрнәғин;

Metin:

1b/2 san.sız Һamd bir Һүдә.yıмға биз.ни yoқ.дн бар қыл-дн ‘ақл ve ‘имән

1b/3 biz қул.larға yār қыл-дн ve тақн nihāyet.siz дүрүд Һазрет_и Muһammed

Gramatiksel dizin:

‘aql

sansız Һamd bir Һүдәймға бизни жоқдın бар қıldı ‘a. ve ‘imān 1b/2

‘imān

sansız Һamd bir Һүдәймға бизни жоқдın бар қıldı ‘aql ve ‘i. 1b/2

bar

sansız Һamd bir Һүдәймға бизни жоқдın b. қıldı ‘aql ve ‘imān 1b/2

bir

sansız Һamd b. Һүдәймға бизни жоқдın бар қıldı ‘aql ve ‘imān 1b/2

biz

b. қullarға yār қıldı ve тақı nihāyetsiz дүрүд Һәзret_i Muḥammed 1b/3

sansız Һamd bir Һүдәймға b.+ni жоқдın бар қıldı ‘aql ve ‘imān 1b/2

dürüd

biz қullarға yār қıldı ve тақı nihāyetsiz d. Һәзret_i Muḥammed 1b/3

Һamd

sansız Һ. bir Һүдәймға бизни жоқдın бар қıldı ‘aql ve ‘imān 1b/2

Һәзret

biz қullarға yār қıldı ve тақı nihāyetsiz дүрүд Һ.+_i Muḥammed 1b/3

Һүдә

sansız Һamd bir Һ.+ymға бизни жоқдın бар қıldı ‘aql ve ‘imān 1b/2

қıl -

sansız Һamd bir Һүдәймға бизни жоқдın бар қ.-dı ‘aql ve ‘imān 1b/2,1b/3

құл

biz қ.+larға yār қıldı ve тақı nihāyetsiz дүрүд Һәзret_i Muḥammed 1b/3

Muḥammed

biz қullarға yār қıldı ve тақı nihāyetsiz дүрүд Һәзret_i M. 1b/3

nihāyet

biz қullarға yār қıldı ve тақı n.+siz дүрүд Һәзret_i Muḥammed 1b/3

san

s.+siz Һamd bir Һүдәймға бизни жоқдın бар қıldı ‘aql ve ‘imān 1b/2

тақı

biz қullarға yār қıldı ve t. nihāyetsiz дүрүд Һәзret_i Muḥammed 1b/3

ve

sansız Һamd bir Һūdāyımğa bizni yoқdın bar қıldı ‘aқl v. ‘imān 1b/2 , 1b/3

yār

biz қullarға y. қıldı ve тақı nihāyetsiz дүрүд Һазret_и Muһammed 1b/3

yоқ

sansız Һamd bir Һūdāyımğa bizni y.+dın bar қıldı ‘aқl ve ‘imān 1b/2

2.2.5.3 Sıkıştırılmış gramatiksel dizin

Metinde bulunan kelimeleri, diğer dizin türlerinden farklı olarak satırın tamamı yerine sadece kelime veya kelime grubu olarak madde başının altında gösteren dizin türüdür. Sıkıştırılmış gramatiksel dizin sadece madde başı olan kelimeyi ve geçtiği satır numarasını barındırır. Bu dizin türünde terkipler ve birleşik fiillerden oluşan kelime grupları ayrılmaksızın madde başı altında beraber gösterilir. Madde başlarının sıralanışı diğer dizin türlerinde olduğu gibi alfabetik sıraya göredir. Satırların sıralanışı ise madde başının aldığı çekim ekinin alfabetik sıralamasına göre yapılır.

Örneğin;

Metin:

1b/2 san.sız Һamd bir Һūdā.yımğa biz.nи yoқ.dın bar қıl-dı ‘aқl ve ‘imān

1b/3 biz қul.larға yār қıl-dı ve тақı nihāyet.siz дүрүд Һазret_и Muһammed

Sıkıştırılmış gramatiksel dizin:

‘aқl

‘a. 1b/2

‘imān

‘i. 1b/2

bar

b. қıldı 1b/2

bir

b. 1b/2

biz

b. 1b/3

b.+ni 1b/2

дүрүд

d. 1b/3

ḥamd

ḥ. 1b/2

Ḥazret

Ḥ.-i Muḥammed 1b/3

Ḥudā

Ḥ.+yımğa 1b/2

ḳıl -

ḳ.-dı 1b/2, 1b/3

ḳul

ḳ.+largā 1b/3

Muḥammed

Ḥazret-i M. 1b/3

nihāyet

n.+sız 1b/3

san

s.+sız 1b/2

taḳı

t. 1b/3

ve

v. 1b/2, 1b/3

yār

y. ḳıldı 1b/3

yok

y.+dın 1b/2

2.2.5.4 Ek dizin

Metinde bulunan kelime eklerinin dizinlemesi amacıyla hazırlanmıştır. Diğer dizin türlerinden farklı olarak kelimelerde bulunan bütün ekler madde başı olarak alınır ve ekin geçtiği kelime, madde başı altında önceden ifade edilen satır hazırlama formatına göre hazırlanır.

Dizin hazırlanırken ekleri ayırmak için “- +” karakterleri kullanılır.”-” karakteri önüne geldiği ekin fiil soylu sözcüklere gelen ek olduğunu gösterirken “+” karakteri isim soylu kelimelere gelen ek olduğunu ifade eder. Tablo 2.10’de ek dizin

hazırlamak için ayırıştırma esnasında kullanılan notasyon ile ilgili örnek ve açıklamalar bulunmaktadır.

Tablo 2.10 : Ek dizin hazırlama biçimi

Gel-iyor-du	Kelimenin kökü gel-, ekler –iyor , -du
Araba.sı+n+da	Kelimenin kökü araba, ekler –sı, -n, -da

Ayırıştırma işlemi tamamlanan ekler madde başı olmak üzere alfabetik olarak sıralanır. Satırlar ise sadece ekin bulunduğu kelime ve kelimenin geçtiği satır numarasından oluşur. Örneğin;

Metin

1b/2 san.sız hamd bir Hūdā.yımğa biz.ni yok.dın bar kııl-dı ‘aql ve ‘imān

1b/3 biz kul.larğa yār kııl-dı ve taķı nihāyet.siz dūrūd Hāzret_i Muḥammed

Ek dizin

-dı

ķıl-dı 1b/2

+dın

yok+dın 1b/2

+larğa

ķul+larğa 1b/3

+ni

biz+ni 1b/2

+sız

san+sız 1b/2

+siz

nihāyet+siz 1b/3

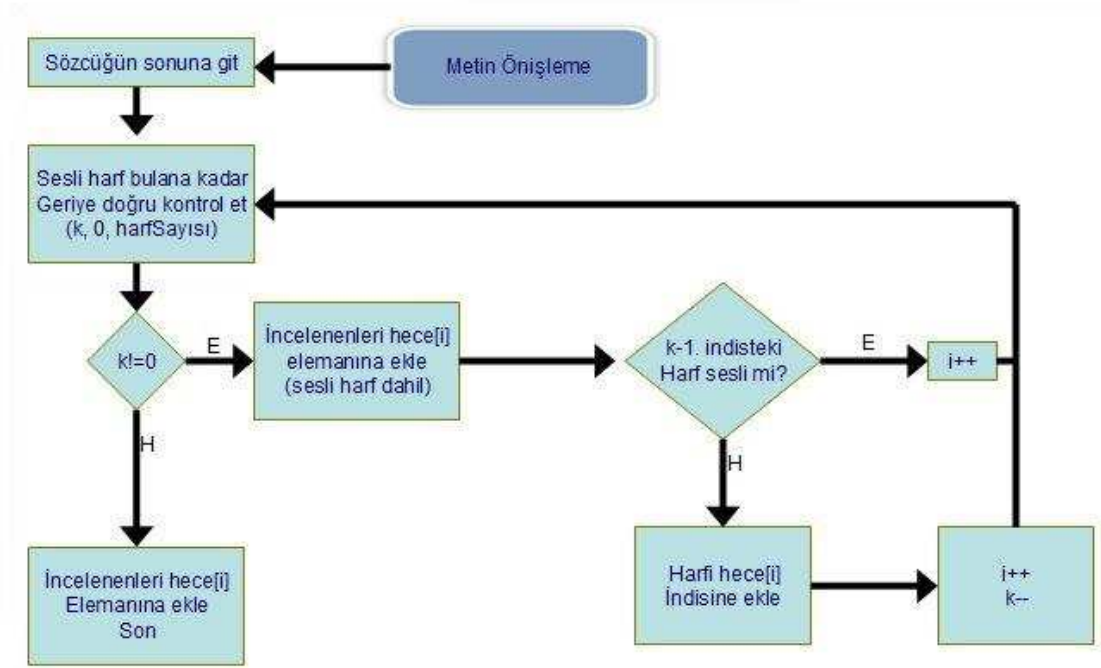
+yımğa

hūdā+yımğa 1b/2

2.2.5.5 Sıklık sözlüğü

Sıklık sözlüğü, metin içerisindeki kelime, hece ve harflere ait istatistikleri hesaplamak amacıyla hazırlanmıştır. Yazılım da bulunan sözlükler metin sıklık çözümleyicisi örnek alınarak hazırlanmıştır. Hazırladığımız yazılımda 14 farklı sıklık sözlüğü bulunmaktadır.

Sıklık sözlükleri oluşturulurken harflerin konumlarına ve özelliklerine göre çeşitli algoritmalar oluşturulmuştur. Örneğin kelimeleri hecelere ayırmak için Türkçe hece yapısı incelenmiş ve bu hece yapısına göre bir algoritma üretilmiştir. Türkçe’de heceler kelimenin sonundan başlayarak iki farklı biçimde konumlanırlar. Birincisi ünsüz + ünlü, ikincisi ünsüz + ünlü + ünsüz biçimleridir. Bu bilgi yardımıyla tüm kelimeler hecelere ayrılabilir. Şekil 2.8’de algoritmanın akış diyagramı gösterilmiştir. Diğer algoritmalar da bu tür yapıların kullanılmasıyla elde edilmiştir.(Hengirmen 1999)



Şekil 2.8 : Heceme algoritması akış diyagramı

Sıklık sözlüklerini daha iyi belirtmek amacı ile bir örnek metin seçilmiş ve bunun üzerinde örnek sıklık sözlükleri elde edilmiştir. Örnek metin, “Doyuncaya dek yüzüne bakamadığım” olarak alınmıştır.

Harf sıklığı

Metin içerisinde bulunan herbir harfin metin içerisinde kaç adet geçtiğini gösteren dizin türüdür. Harflerin sıralanışı metin içerisindeki yoğunluklarına göre çoktan aza doğru sıralanır. Ayrıca herbir harfin tüm kelimeler içerisinde yüzde kaçını teşkil ettiği de hesaplanır. Tablo 2.11’de örnek metin için yazılım tarafından üretilen örnek sonuç gösterilmiştir.

Tablo 2.11 : Harf sıklığı tablosu

HARF SIKLIĞI			
SIRA	KELİME	SIKLIK	%
1	a	5	%17,24
2	d	3	%10,34
3	y	3	%10,34
4	ı	2	%6,896
5	k	2	%6,896
6	m	2	%6,896
7	n	2	%6,896
8	ü	2	%6,896
9	e	2	%6,896
10	b	1	%3,448
11	o	1	%3,448
12	u	1	%3,448
13	c	1	%3,448
14	ğ	1	%3,448
15	z	1	%3,448
TOPLAM		29	%100

Ünlü ünsüz harf sıklığı

Metin içerisinde bulunan harflerin ünlü veya ünsüz olmalarına göre metin içerisinde kaç adet geçtiğini gösteren dizin türüdür. Ünlü ve ünsüz harfler, adet ve yüzdelerine göre hesaplanarak metin içerisindeki yoğunluklarına göre sıralanırlar. Ayrıca herbir harfin kaç adet olduğu da gösterilir. Tablo 2.12’de örnek metin için yazılım tarafından üretilen örnek sonuç gösterilmiştir.

Tablo 2.12 : Ünlü ünsüz harf sıklığı tablosu

ÜNLÜ-ÜNSÜZ HARF SIKLIĞI				
SIRA	TİPİ	SIKLIK	%	SIKLIK DAĞILIMI
1	Ünlüler(aıüeou)	13	%44	a(5), ı(2), ü(2), e(2), o(1), u(1),
2	Ünsüzler(dykmnbcğz)	16	%55	d(3), y(3), k(2), m(2), n(2), b(1),
TOPLAM		29		

Kalın ince ünlü harf sıklığı

Metin içerisinde bulunan ünlü harflerin kalın ya da ince olmalarına göre gruplandığı dizin türüdür. Bu dizin türünde de kalın veya ince olan ünlü harflerin adetleri ve yüzdeleri hesaplanır. Tablo 2.13’te örnek metin için yazılım tarafından üretilen örnek sonuç gösterilmiştir.

Tablo 2.13 : Kalın ince ünlü harf sıklığı tablosu

KALIN - İNCE ÜNLÜ HARF SIKLIĞI				
SIRA	TİPİ	SIKLIK	%	SIKLIK DAĞILIMI
1	Kalın(aıou)	9	%69	a(5), ı(2), o(1), u(1),
2	İnce(üe)	4	%30	ü(2), e(2),
TOPLAM		13	%100	

Düz yuvarlak ünlü harf sıklığı

Metin içerisinde bulunan ünlü harflerin düz ya da yuvarlak olmalarına göre gruplandığı dizin türüdür. Bu dizin türünde de düz veya yuvarlak olan ünlü harflerin adetleri ve yüzdeleri hesaplanır. Tablo 2.14’te örnek metin için yazılım tarafından üretilen örnek sonuç gösterilmiştir.

Tablo 2.14 : Düz yuvarlak ünlü harf sıklığı tablosu

DÜZ - YUVARLAK ÜNLÜ HARF SIKLIĞI				
SIRA	TİPİ	SIKLIK	%	SIKLIK DAĞILIMI
1	Düz(aıe)	9	%69	a(5), ı(2), e(2),
2	Yuvarlak(üiou)	4	%30	ü(2), o(1), u(1),
TOPLAM		13	%100	

Geniş dar ünlü harf sıklığı

Metin içerisinde bulunan ünlü harflerin geniş ya da dar olmalarına göre gruplandığı dizin türüdür. Bu dizin türünde de geniş veya dar olan ünlü harflerin adetleri ve yüzdeleri hesaplanır. Tablo 2.15’de örnek metin için yazılım tarafından üretilen örnek sonuç gösterilmiştir.

Tablo 2.15 : Geniş dar ünlü harf sıklığı tablosu

GENİŞ - DAR ÜNLÜ HARF SIKLIĞI				
SIRA	TİPİ	SIKLIK	%	SIKLIK DAĞILIMI
1	Geniş(üü)	5	%38	ı(2), ü(2), u(1),
2	Dar(aeo)	8	%61	a(5), e(2), o(1),
TOPLAM		13	%100	

Sedalı sedasız ünsüz harf sıklığı

Metin içerisinde bulunan ünsüz harflerin sedalı ya da sedasız olmalarına göre gruplandığı dizin türüdür. Bu dizin türünde de sedalı veya sedasız olan ünsüz harflerin adetleri ve yüzdeleri hesaplanır. Tablo 2.16’da örnek metin için yazılım tarafından üretilen örnek sonuç gösterilmiştir.

Tablo 2.16 : Sedalı sedasız ünsüz harf sıklığı tablosu

SEDALI - SEDASIZ ÜNSÜZ HARF SIKLIĞI				
SIRA	TİPİ	SIKLIK	%	SIKLIK DAĞILIMI
1	Sedalı(dymnbcğz)	14	%87	d(3), y(3), m(2), n(2), b(1), c(1),ğ(1),z(1)
2	Sedasız(k)	2	%12	k(2),
TOPLAM		16	%100	

Sürekli süreksiz ünsüz harf sıklığı

Metin içerisinde bulunan ünsüz harflerin sürekli ya da süreksiz olmalarına göre gruplandığı dizin türüdür. Bu dizin türünde de sürekli veya süreksiz olan ünsüz harflerin adetleri ve yüzdeleri hesaplanır. Tablo 2.17’de örnek metin için yazılım tarafından üretilen örnek sonuç gösterilmiştir.

Tablo 2.17 : Sürekli süreksiz ünsüz harf sıklığı tablosu

SÜREKLİ - SÜREKSİZ ÜNSÜZ HARF SIKLIĞI				
SIRA	TİPİ	SIKLIK	%	SIKLIK DAĞILIMI
1	Sürekli(ymnğz)	9	%56	y(3), m(2), n(2), ğ(1), z(1),
2	Süreksiz(dkbc)	7	%43	d(3), k(2), b(1), c(1),
TOPLAM		16	%100	

Hece içi harf sıklığı

Metin içerisinde bulunan harflerin hecelerde kaçınıcı harfi oluşturduklarını gösteren dizin türüdür. Harfler metin içerisindeki yoğunluklarına göre sıralanır. Tablo 2.18’de örnek metin için yazılım tarafından üretilen örnek sonuç gösterilmiştir.

Tablo 2.18 : Hece içi harf sıklığı tablosu

HECE İÇİ HARF SIKLIĞI							
SIRA	HARF	SIKLIK	%	1.HARF	2.HARF	3.HARF	4.HARF
1	a	5	%17,24	0(%0)	5(%100)	0(%0)	0(%0)
2	d	3	%10,34	3(%100)	0(%0)	0(%0)	0(%0)
3	y	3	%10,34	3(%100)	0(%0)	0(%0)	0(%0)
4	ı	2	%6,896	0(%0)	2(%100)	0(%0)	0(%0)
5	k	2	%6,896	1(%50)	0(%0)	1(%50)	0(%0)
6	m	2	%6,896	1(%50)	0(%0)	1(%50)	0(%0)
7	n	2	%6,896	1(%50)	0(%0)	1(%50)	0(%0)
8	ü	2	%6,896	0(%0)	2(%100)	0(%0)	0(%0)
9	e	2	%6,896	0(%0)	2(%100)	0(%0)	0(%0)
10	b	1	%3,448	1(%100)	0(%0)	0(%0)	0(%0)
11	o	1	%3,448	0(%0)	1(%100)	0(%0)	0(%0)
12	u	1	%3,448	0(%0)	1(%100)	0(%0)	0(%0)
13	c	1	%3,448	1(%100)	0(%0)	0(%0)	0(%0)
14	ğ	1	%3,448	1(%100)	0(%0)	0(%0)	0(%0)
15	z	1	%3,448	1(%100)	0(%0)	0(%0)	0(%0)
TOPLAM		29	%100	13	13	3	0

Hece sıklığı

Metin içerisinde bulunan hecelerin hesaplanarak yoğunluklarına göre gruplandırıldığı dizin türüdür. Tablo 2.19’de örnek metin için yazılım tarafından üretilen örnek sonuç gösterilmiştir.

Tablo 2.19 : Hece sıklığı tablosu

HECE SIKLIĞI			
SIRA	KELİME	SIKLIK	%
1	ba	1	%8,250
2	ca	1	%8,250
3	dek	1	%8,250
4	dı	1	%8,250
5	do	1	%8,250
6	ğım	1	%8,250
7	ka	1	%8,250
8	ma	1	%8,250
9	ne	1	%8,250
10	yun	1	%8,250
11	yü	1	%8,250
12	zü	1	%8,250
TOPLAM		12	%100

Kelime sıklığı

Metin içerisinde bulunan kelimelerin hesaplanarak geçiş sayılarına göre gruplandırıldığı dizin türüdür. Tablo 2.20’de örnek metin için yazılım tarafından üretilen örnek sonuç gösterilmiştir.

Tablo 2.20 : Kelime sıklığı tablosu

KELİME SIKLIĞI			
SIRA	KELİME	SIKLIK	%
1	bakamadığım	1	%25
2	dek	1	%25
3	doyuncaya	1	%25
4	yüzüne	1	%25
TOPLAM		4	%100

Tersten sıralı kelime sıklığı

Metin içerisinde bulunan kelimelerin hesaplanarak yoğunluklarına göre değil de son harfinden başlayarak alfabetik olarak gruplandırıldığı dizin türüdür. Tablo 2.21’de örnek metin için yazılım tarafından üretilen örnek sonuç gösterilmiştir.

Tablo 2.21 : Tersten sıralı kelime sıklığı tablosu

TERSTEN SIRALI KELİME SIKLIĞI			
SIRA	KELİME	SIKLIK	%
1	doyuncaya	1	%25
2	yüzüne	1	%25
3	dek	1	%25
4	bakamadığım	1	%25
TOPLAM		4	%100

Hece tipi sıklığı

Ünlü harflerin Vokal(V), ünsüz harflerin Consonant(C) olarak simgelenerek hece tipine göre yoğunlukların hesaplandığı dizin türüdür. Tablo 2.22’de örnek metin için yazılım tarafından üretilen örnek sonuç gösterilmiştir.

Tablo 2.22 : Hece tip sıklığı tablosu

HECE TİP SIKLIĞI			
SIRA	KELİME	SIKLIK	%
1	CV	10	%76,92
2	CVC	3	%23,07
TOPLAM		13	%100

Kelime tipi sıklığı

Ünlü harflerin Vokal(V), ünsüz harflerin Consonant(C) olarak simgelenerek kelime tipine göre yoğunlukların hesaplandığı dizin türüdür. Tablo 2.23’te örnek metin için yazılım tarafından üretilen örnek sonuç gösterilmiştir.

Tablo 2.23 : Kelime tip sıklığı tablosu

KELİME TİP SIKLIĞI			
SIRA	KELİME	SIKLIK	%
1	CVC	1	%25
2	CVCVCCVVCV	1	%25
3	CVCVCV	1	%25
4	CVCVCVCVCVC	1	%25
TOPLAM		4	%100

Harf tipi sıklığı

Ünlü harflerin Vokal(V) ünsüz harflerin Consonant(C) olarak simgelenerek harf tipine göre yoğunlukların hesaplandığı dizin türüdür. Tablo 2.24’te örnek metin için yazılım tarafından üretilen örnek sonuç gösterilmiştir.

Tablo 2.24 : Harf tip sıklığı tablosu

HARF TİP SIKLIĞI			
SIRA	KELİME	SIKLIK	%
1	C	16	%55,17
2	V	13	%44,82
TOPLAM		29	%100

3. HAZIRLANAN YAZILIM

3.1 Giriş

Bu tez çalışması ile kitap sonu indeksleri hazırlayan araştırmacılara çalışmalarında katkı sağlayacak yeni bir yazılım oluşturulması hedeflenmiştir. Bu hedef doğrultusunda literatürde mevcut dizinleme yöntemleri ve yazılımları incelenmiştir. Sonuç olarak dizinleme konusunda mevcut yöntem ve yazılımların hatalarının tespit edilmesi ve eksikliklerin yeni yöntemlerle giderilmesi sağlanmıştır. Programın en önemli özelliği görselliği ön plana çıkararak dizin hazırlama işlemini farklı araçlara ihtiyaç kalmadan kolay, hızlı ve yüksek doğrulukta yapma yeteneğine sahip olmasıdır. Programın kullanıma sunulması ile kullanıcılardan alınacak geridönüşlerle programın Türkoloji alanındaki önemli bir eksikliği kapatması amaçlanmaktadır.

Disiplinler arası çalışmalara en güzel örneklerden biri olan yazılım, Türk Dili ve Edebiyatı Bölümü ile ortaklaşa hazırlanmıştır. Yazılımın hazırlanmasında birçok zorluklarla karşılaşılmasına rağmen bu zorluklar aşılarak tez sonucunda Türkologların ihtiyaçlarına cevap verebilecek nitelikte bir program oluşturulmuştur.

Program beş ana bölümden oluşmaktadır. Hazırlanan bu bölümler birbirlerinden bağımsız çalışmaktadır. Herbir bölüm kendi işlevini gerçekleştirerek zor ve uzun zaman alan dizinleme işlemi arka plandaki yazılım kodları sayesinde kısa ve etkin bir şekilde gerçekleştirilebilmektedir. Yazılımın gerçekleştirilmesi esnasında yazılımın indirgenmesi ile oluşturulan bölümler Şekil 3.1’de gösterilmiştir.



Şekil 3.1 : Dizinleme yazılımı bölümleri

Tezde ayrıntılı olarak anlatılan Şekil 3.1’deki bölümler temel olarak dizin işlemini gerçekleştirmek için gereken tüm aşamaları kapsamaktadırlar.

Program 31 farklı sınıf ve form içermektedir. Bu form ve sınıflar görsel bir bütünlük içinde birleştirilerek paket program haline getirilmiştir. Ayrıca ayrıştırma işleminin daha kolay yapılması için basit bir veritabanı kullanılmıştır.

3.2 Mevcut Sistem

Kitap sonu indekslerinin hazırlanması için çeşitli yöntemler mevcuttur. Bu yöntemler ayrıntılı olarak önceki bölümlerde anlatılmıştır. Fakat yaygın olarak kullanılan yöntem;

- Metnin bilgisayar ortamına aktarılması
- Ofis programları gibi yardımcı programlarla bütün kelimelerin ayrıştırma işleminin yapılması
- Ayrıştırma işlemi tamamlanan kelimelerin Cibakaya veya Concordance gibi yardımcı yazılımlarla dizinlenmesi
- Dizin işlemi tamamlanan metinlerin hata kontrolünün yapılmasıdır.

Yukarda sıralanan işlemlerin yapılması normal uzunlukta bir metin için yaklaşık 6-8 ay zaman gerektirmektedir.

3.3 Problem Belirleme

Yazılım geliştirme aşamalarının en önemlilerinden birisi problem belirleme aşamasıdır. Bu aşamada, mevcut sistemlerin incelenmesi, problemlerin tespit edilmesi ve çalışmanın amaç ve kapsamının belirlenmesi gerekmektedir. Hazırladığımız çalışmada mevcut sistemler incelenerek problem ve eksikler belirlenmiş olup projenin kapsamı oluşturulmuştur. (Riordan 2002)

3.3.1 Problemler

Problemler belirlenirken mevcut sistemlerde karşılaşılan sorunlar incelenmektedir. Bu sayede hazırlanacak sistemin içeriği belirlenmiş olur. Hazırladığımız sistemde mevcut yazılımlarda ki problemler şunlardır;

- Ayırıştırma işleminin tamamen elle yapılması,
- Transkripsiyon harflerinin tanınmamasından kaynaklanan font hataları,
- Kitap sonu indekslerinin hazırlanmasını hedef edinen bilgisayar yazılımlarının eksikliği ve var olanların yetersizliği,
- Çalışma yaparken birden fazla yazılıma ihtiyaç duyulması,
- Dizin hazırlama sürecinin çok fazla zaman alması.

3.3.2 Amaçlar

Belirlenen problemler doğrultusunda geliştirilecek yazılımın içeriği bu aşamada belirtilmektedir. Hazırladığımız yazılımda da amaçlar şöyle özetlenebilir;

- Bu alanda çalışan Türkologların işlerini kolaylaştırmak,
- Görüntü işleme ile metinlerin bilgisayar ortamına aktarmak,
- Ayırıştırma işlemini otomatik yapılmasını sağlamak,
- Transkripsiyon harflerinden kaynaklanan hataları düzeltmek,
- Dizin hazırlama süresini minimuma indirmek,
- Farklı formatta dizin hazırlamak,
- Bütün işlemlerin farklı yazılımlara ihtiyaç duyulmadan gidermek.

3.3.3 Projenin kapsamı

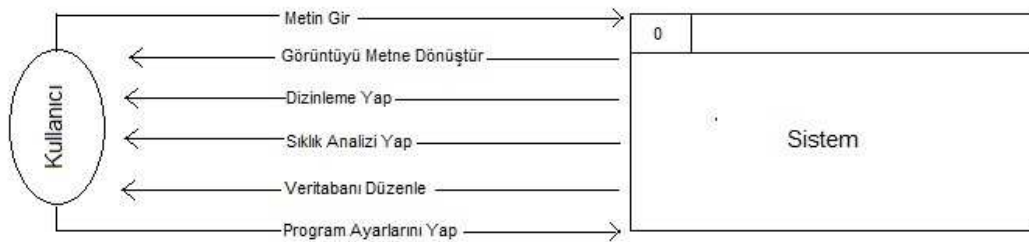
Geliştirilecek çalışmanın kapsamının belirlenmesi kullanıcı ve yazılımcı açısından önemlidir. Yazılımın neler yapıp neler yapamayacağı bu aşamada belirlenir. Hazırladığımız yazılımın kapsamı maddeler halinde şu şekildedir;

- Görüntü işleme ile metinlerin bilgisayar ortamına aktarılması,
- Ayırıştırma işleminin otomatik yapılması,
- Dizinleme işleminin farklı formatlarda hazırlanması,
- Metin sıklık analizinin yapılması,
- Veritabanı işlemleri ile kelimelerin ekleme silme ve düzenleme işleminin yapılması.

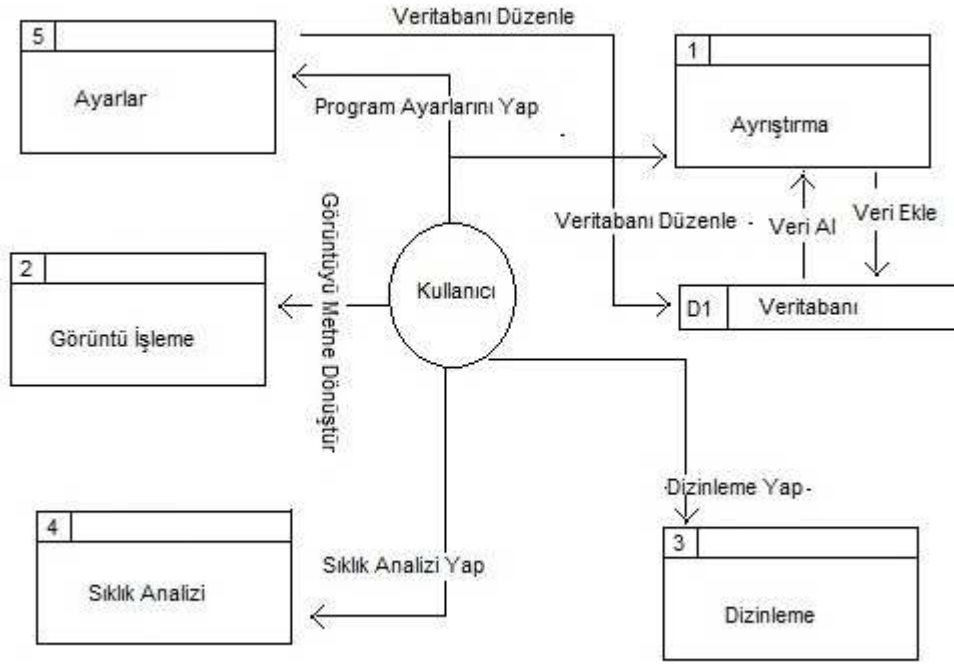
3.4 Kullanıcı Durum Modeli

Kullanıcı durum modeli, yazılımın hangi kullanıcı türleri tarafından kullanılacağını ve bu kullanıcı türlerinin görev ve sorumluluklarını belirlemek amacıyla oluşturulmaktadır.(WEB_8)

Hazırladığımız çalışma tek tür kullanıcı modeline göre tasarlanmıştır. Kullanıcının yazılımdaki rol ve sorumlulukları iki seviye halinde şekillendirilmiştir. Şekil 3.2 kullanıcının hazırlanan sistemde neler yapabileceğini göstermektedir. Şekil 3.3'te ise ayrıntılı olarak modüllerin birbirleri ve kullanıcı ile olan bağlantıları ifade edilmiştir.



Şekil 3.2 : Kullanıcı durumu modeli seviye - 0



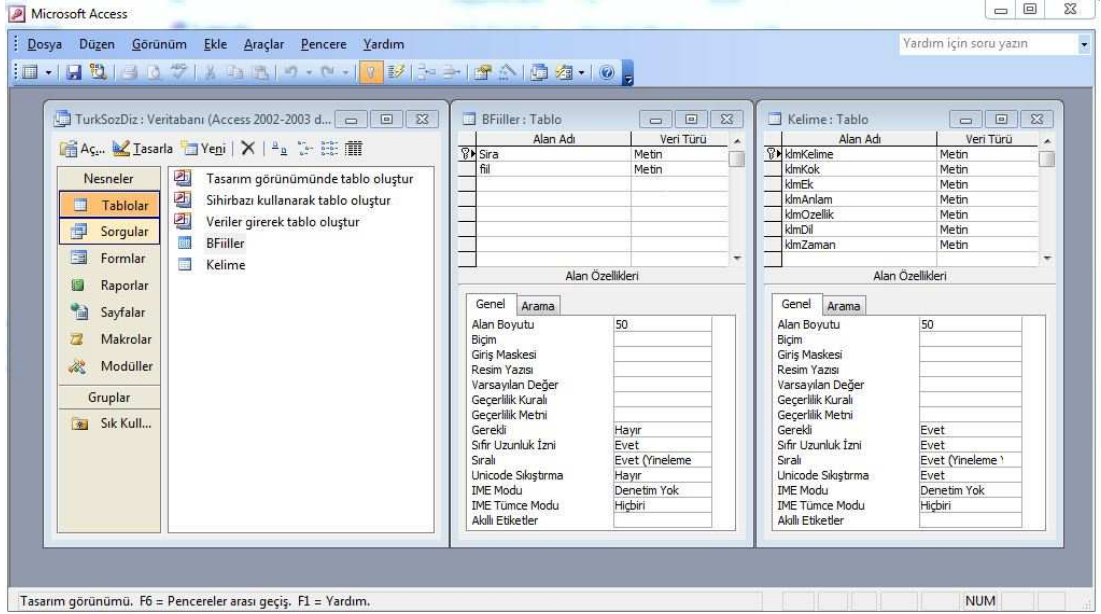
Şekil 3.3 : Kullanıcı durumu modeli seviye - 1

3.5 Veritabanı Modeli

Yazılımda hazırlanan veritabanı sadece ayrıştırma esnasında verileri kaydetmek ve birleşik fiilleri düzenlemek amacıyla kullanıldığından oldukça basit hazırlanmıştır. Hazırlanan veritabanı kelimelerin ayrıştırılması ve dizinlenmesi esnasında kullanılmaktadır. Kelimelerin tüm özellikleri kaydedilerek ayrıştırma esnasında kullanıcının yükünü hafifletmek amaçlanmıştır.

Hazırlanan veritabanı Access 2003 veritabanı aracı kullanılarak hazırlanmıştır. Bu aracı kullanmamızdaki sebepler; hazırlanan veritabanının Şekil 3.4'te görüldüğü gibi çok fazla ayrıntılı ve ilişkisel bir yapıda olmaması ve kullanıcıların Office araçlarını etkin olarak kullanmasıdır. (Gözüdeli 2007)

Veritabanı birbiri ile ilişkisi olmayan iki tablodan oluşmaktadır. Kelime tablosu ayrıştırma esnasında kaynak niteliği taşıyan kelimeleri ve bu kelimelerin özelliklerini içerirken birleşik fiil tablosu birleşik fiillerin kaydedilmesi amacıyla oluşturulmuştur. Şekil 3.4'te tablolar ve özellikleri gösterilmiştir.



Şekil 3.4 : Veritabanı modeli

3.6 Yazılım Aşamaları

Proje başlangıcında zamanımızı daha verimli kullanmak ve projenin daha sistemli oluşturulması için Tablo 3.1’de gösterilen yazılım aşamaları belirlenmiştir. Belirlenen bu aşamalar bir yandan projenin ne kadar süreceğini tahmin etmemize yardımcı olurken diğer yandan yapmamız gereken görevleri adım adım yaparak yazılım geliştirme sürecine uygun hareket etmemize yardımcı olmuştur.(WEB_9)

Tablo 3.1 : Yazılım aşamaları tablosu

Aşama adı	Ayrıntılı bilgi	Zamanla
	Proje ekibinin belirlenmesi	2 hafta
	Proje hakkında ön hazırlıkların yapılması	3 hafta
Gereksinim analizi	Kullanıcılarla anket yapılması	2 hafta
	Kullanıcılara yapılan anketin değerlendirilmesi	1 hafta
	Rol ve sorumlulukların belirlenmesi	2 hafta
Tasarım	İş modellemesi	1 hafta
	Diyafram tasarımı	1 hafta
	Sınıf tasarımı	1 hafta
	Veritabanı tasarımı	2 hafta
	Arayüz tasarımı	2 hafta
Geliştirme	Veritabanı kodlaması	2 hafta
	Yazılım kodlaması	12 hafta
	Dokümantasyon yapılması	2 hafta
Test ve eğitim	Veritabanı ve yazılım testlerinin yapılması	4 hafta
	Kullanıcıların eğitimi	1 hafta
	Genel testlerin tamamlanması	2 hafta
Sunum ve Bakım	Projenin sunumu ve bakım aşamasının başlaması	12 hafta
	Projenin sonlandırılması	
Toplam		52 hafta

Proje başlangıcında Tablo 3.1’de belirlediğimiz süreçler ve bu süreçler tamamlamak için gereken süreler, proje tamamlandığında büyük oranda doğru tahmin edildiği görülmüştür.

3.7 Yazılım Araçları

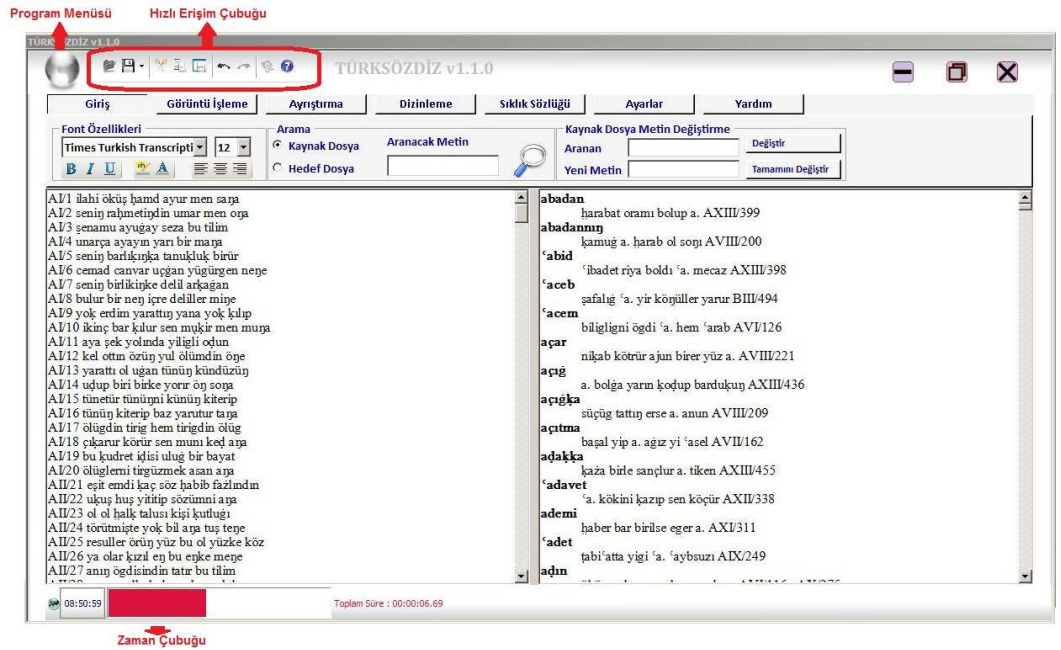
Proje Windows ortamında kullanılmak üzere kişisel bilgisayarlara uyumlu bir formatta hazırlanmıştır. Yapım aşamasında Microsoft Visual Studio 2010 yazılım

geliştirme aracı kullanılmıştır. Bu araç sunduğu basit çözümler, sahip olduğu kütüphanelerle ve yaygın kullanımıyla daha uygun bulunmuştur. Ayrıca veritabanı tasarımında Microsoft Office Access 2003 aracı kullanılmıştır.(Yanık 2002)

3.8 Kullanıcı Arayüzleri ve Program Tanıtımı

3.8.1 Giriş

Program ilk açıldığında Şekil 3.5'te gösterilen form ekrana gelir. Bu programın bütün temel işlevlerini kapsayan ana formdur. Form üzerinde de görüldüğü üzere çeşitli kontrol araçları ve çalışma bölümleri mevcuttur.



Şekil 3.5 : Program giriş form görünümü

Program menüsü olarak isimlendirilen alan dosya açma veya kaydetme gibi çeşitli seçenekler içerir.

Hızlı araç çubuğu, programın hızlı kullanılmasını ve estetiğini artırmak amacıyla hazırlanmıştır. Bu alanda var olan dosya aç, dosya kaydet, geri-ileri al gibi seçenekler kullanıcıların daha hızlı işlem yapmasına olanak sağlamaktadır.

Font özellikleri, kullanıcının çalışılacak metin ile ilgili her türlü font seçeneğini değiştirebileceği özellikler içermektedir. Bu özellik sayesinde diğer yazılımlarda karşılaşılan font hataları önlenmektedir. Ayrıca farklı font türleriyle de çalışma imkânı sağlanmaktadır.

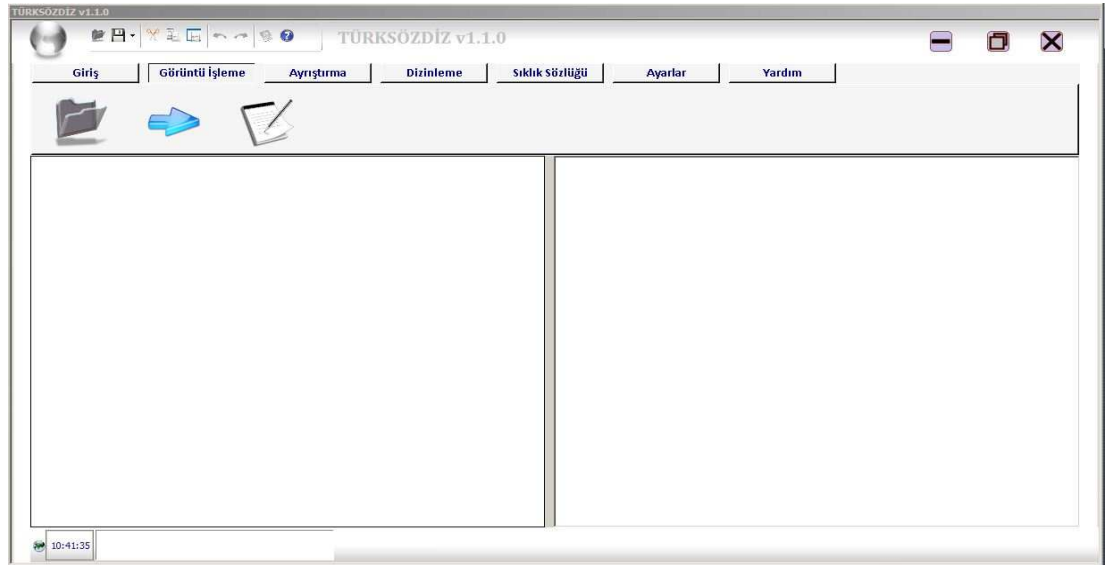
Arama, aranacak metinlerin yerlerini belirlemek amacıyla hazırlanmıştır. Kaynak dosya metin deęiřtirme ise metin üzerinde otomatik deęiřtirme yapma olanaęı saęlamaktadır.

Form üzerinde iki ayrı alan bulunmaktadır. Bu alanlardan sol tarafta bulunan ve kaynak dosya olarak isimlendirilen alan, dizinleme veya sıklık analizi yapmak için hazırlanan metinlerin gösterildięi alandır. Dizinleme ve sıklık analizi yapılmadan önce kaynak metin üzerindeki her türlü ön iřlem bu alanda gerekleřtirilmektedir. Saę tarafta gösterilen hedef dosya olarak isimlendirilen alan ise dizinleme ve sıklık analizi gibi sonuların gösterildięi alandır.

Görüntü iřleme, ayrıştırma, dizinleme, sıklık sözlüęü, ayarlar ve yardım seenekleri ise kullanıcının farklı iřlemleri yapabileceęi alıřma alanlarını gösterir. Bu seenekler ilerde ayrıntılı anlatılacaktır.

3.8.2 Görüntü iřleme

Giriř formu üzerinde bulunan görüntü iřleme sekmesi seildięinde řekil 3.6’da gösterilen form gösterilir. Sol tarafta resim dosyalarının görüntülendięi alan bulunurken sonular saę tarafta bulunan alanda gösterilir.



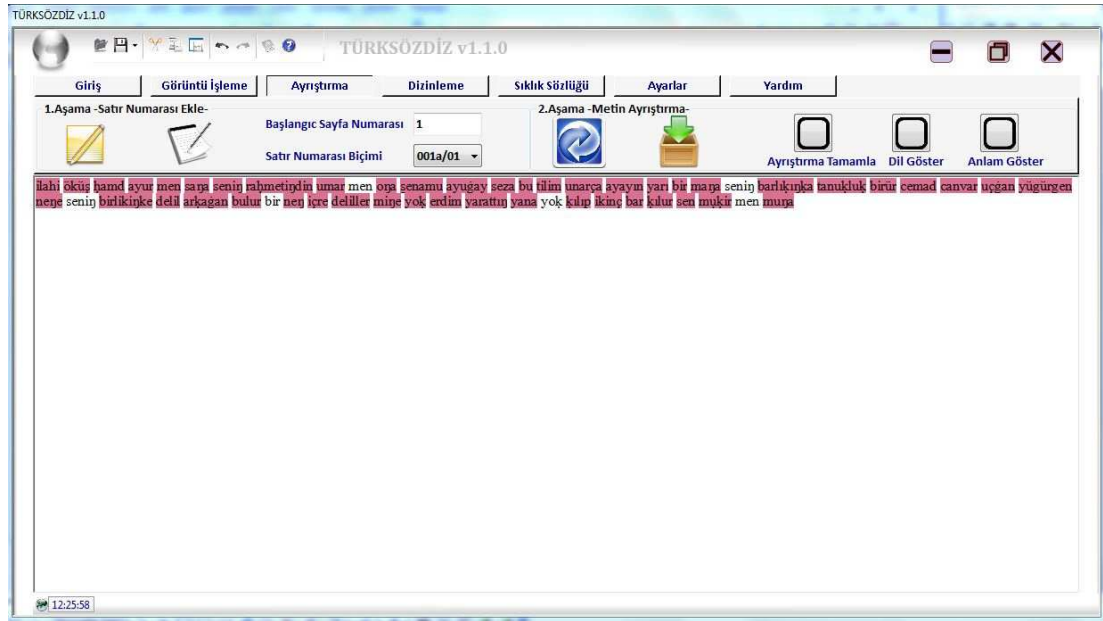
řekil 3.6 : Görüntü iřleme form görüntüsü

Resim dosyalarının metin dosyalarına dönüřtürölmesi için resim dosyalarının dosya aç ile ekrana alınması gerekmektedir. Ekrana alınan resim dosyaları tanıma düęmesine basılarak metne dönüřtürülebilir. Sonu formun saę tarafında görüntülenir.

Bu bölümde Office yazılımlar içerisinde bulunan MODI kütüphanesi referans alınmıştır. Karşılaşılan en büyük problem ise transkripsiyon harflerinin tanınmamasıdır. Farklı bir tez çalışması olarak transkripsiyon alfabesini tanıyan bir görüntü işleme uygulaması faydalı olacaktır.(WEB_10)

3.8.3 Ayırıştırma

Dizinleme işleminin öncesinde yapılan bütün ön işlemlerin gerçekleştirildiği bölümdür. Bu işlemlerin gerçekleştirilmesi için giriş formu üzerinde bulunan ayırıştırma sekmesi seçilmelidir. Ayırıştırma sekmesinin seçilmesi ile Şekil 3.7’de görüntülenen form gösterilir.



Şekil 3.7 : Ayırıştırma form görünümü

Ayırıştırma işlemi temel olarak kullanıcının önceki çalışmalarını kaynak olarak yeni hazırlanacak metinlerin otomatik ayrıştırılmasını sağlar. Kullanıcının ayrıştırması gereken kelimeler farklı bir renk ile işaretlenir. Bu sayede her seferinde bütün metni ayrıştırmak yerine sadece ilk defa karşılaştığı kelimeleri ayrıştırması yeterli olacaktır.

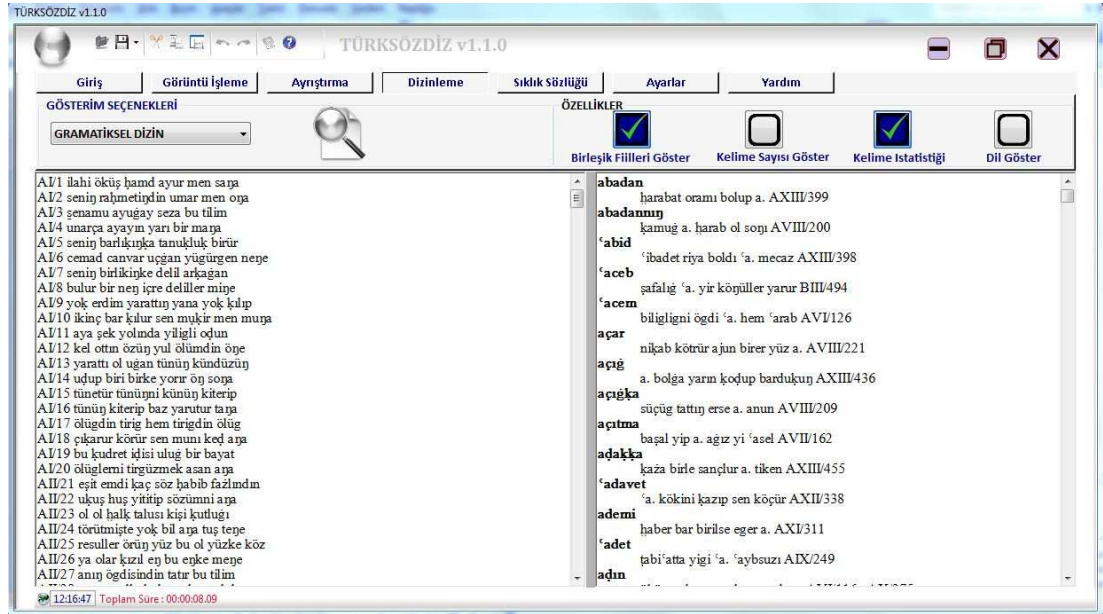
Ayırıştırma işlemi yapılacak metnin metin alanına elle ya da Word dosyasından alınması gerekmektedir.(WEB_11) Ekrana alınan metin üzerinde gerekli ön işlemler yapıldıktan sonra “Yenile” düğmesine basılarak otomatik ayrıştırma süreci başlatılabilir. Ayırıştırma için önerilen kelimelerin ayrıştırması tamamlandıktan sonra “Ayırıştırmayı Tamamladım” seçeneği aktif hale getirilerek “Yenile” düğmesine basılarak ayrıştırma işleminin ilk aşaması tamamlanır.

Ayrıştırma işleminin diğer bir aşaması satır numaralarının eklenmesidir. Satır numaralarını eklemek için eklenecek satır numarasının özellikleri gerekli seçeneklerle belirlendikten sonra “*Satır Numarası Ekle*” düğmesine basılarak otomatik satır numarası ekleme işlemi gerçekleştirilebilir. “*Satır Numarası Sil*” seçeneği ise herhangi bir hata olması durumunda eklenen satır numaralarının otomatik olarak ekrandan silinmesini sağlar. Ayrıştırma işlemi tamamlanmış ve satır numarası eklenmiş metin dizin hazırlamak için hazır hale gelir.

Ayrıştırma işlemi tamamlandığında istenilmesi halinde “*Veritabanına Yaz*” düğmesi yardımıyla ayrıştırılan kelimeler veritabanına kaydedilebilir ve ilerdeki çalışmalarda kaynak olarak kullanılabilir.

3.8.4 Dizinleme

Ayrıştırma işlemi tamamlanan metinde dizin hazırlamak için dizinleme sekmesi seçilmelidir. Dizinleme sekmesinin seçilmesi ile Şekil 3.8’de görüntülenen form görüntülenir. Bu form yardımıyla çeşitli formatlarda dizin hazırlanabilir.



Şekil 3.8 : Dizinleme form görünümü

Gösterim seçenekleri hazırlanacak dizin türünün belirlenmesini sağlar. Program dört farklı türde dizin hazırlanmasına olanak sağlar. Bu dizin türleri gramatiksel dizin, bağlam sözlüğü, sıkıştırılmış gramatiksel dizin ve ek dizindir. Önceki bölümde ayrıntılı olarak bu dizin türlerinin farkları ve özellikleri ifade edilmiştir. Burada sadece dizin hazırlama işleminin nasıl yapıldığı anlatılacaktır.

Özellikler bölümü, hazırlanacak dizinin hangi özellikleri içermesi gerektiğini belirtir. “Dil Gösterimi” seçeneği, madde başı olan kelimenin ait olduğu dili madde başının yanına eklenmesini sağlar. “Kelime Sayısı” ve “Toplam Kelime İstatistiği” seçenekleri metin içerisindeki kelimelerin istatistiğini hesaplamak için kullanılır. “Birleşik Fiilleri Göster” seçeneği sıkıştırılmış gramatiksel dizin türünde hazırlanan dizinlerde “bol-, ir-, kıl-,” gibi birleşik fiillerin kendinden önceki kelime ile beraber satır içerisinde gösterilmesini sağlar. Ayırıştırma veya dizinleme yapılırken programın donmasını engellemek için Thread yapısı kullanılmıştır. Böylelikle dizinleme yapılırken farklı işlemlerin yapılmasına da olanak sağlanmıştır.(WEB_12)

3.8.5 Sıklık analizi

Hazırlanan dizini zenginleştirmek için sıklık analizi yardımıyla farklı türlerde dizinlerde hazırlanabilmektedir. Form üzerinde sıklık sözlüğü sekmesi işaretlendiğinde Şekil 3.9’da gösterilen form oluşur.

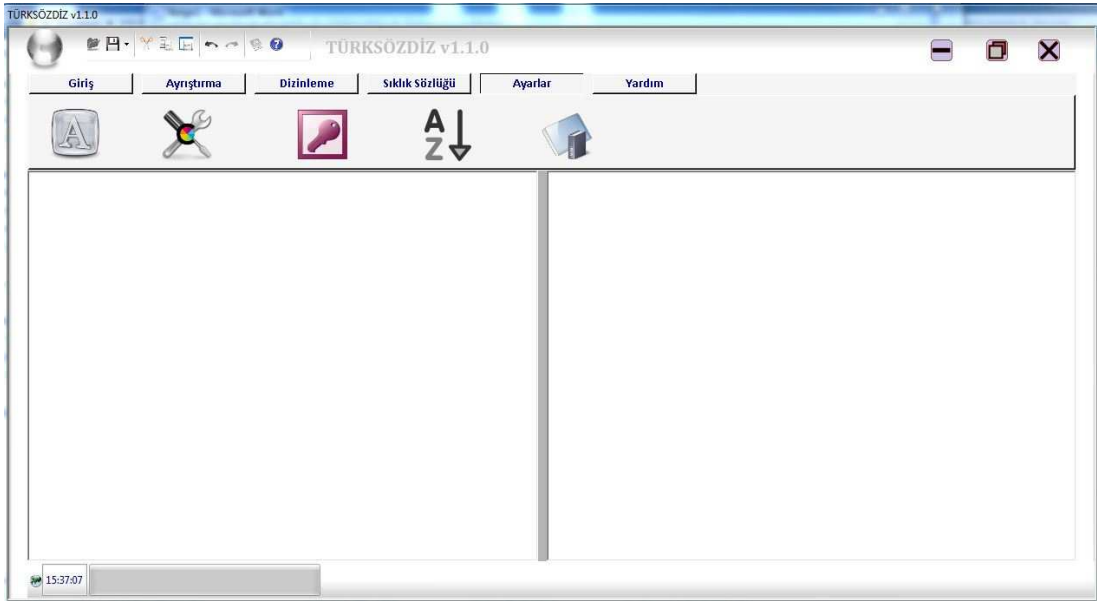
HARF SIKLIĞI			
SIRA	KELİME	SIKLIK	%
1	a	1777	%11,63
2	i	1524	%9,981
3	ı	1242	%8,134
4	e	1065	%6,974
5	l	861	%5,638
6	r	830	%5,435
7	n	728	%4,767
8	u	570	%3,733
9	t	551	%3,608
10	b	513	%3,359
11	k	469	%3,071
12	s	442	%2,894
13	ğ	422	%2,763
14	ü	412	%2,698
15	m	408	%2,672
16	z	297	%1,945
17	y	289	%1,892
18	d	284	%1,859
19	g	265	%1,735
20	x	260	%1,702
21	v	247	%1,617
22	h	246	%1,611

Şekil 3.9 : Sıklık sözlüğü form görünümü

Hazırlanan yazılım, önceki bölümde ayrıntılı olarak açıklanan 14 farklı sıklık sözlüğü barındırmaktadır. İstenilen sıklık sözlükleri seçilip göster düğmesine basıldığında çalışma metnine ait analiz sonuçları kısa süre içerisinde hedef dosya bölümünde tablo halinde oluşmaktadır. Tablolar oluşturulurken RTF formatından yararlanılmıştır.(WEB_13)

3.8.6 Ayarlar

Kullanıcı program ile ilgili düzenlemeleri ayarlar bölümünde bulunan kontrollerle yapabilmektedir. Ana form üzerinde bulunan ayarlar sekmesine basıldığında ilgili kontroller ekrana gelecektir. Şekil 3.10’da görüldüğü gibi hedef dosya görünüm ayarları, renk ayarları ve veritabanı ayarları kullanıcı tarafından gerçekleştirilebilmektedir.



Şekil 3.10 : Program ayarları form görünümü

3.8.6.1 Font ayarları

Program en önemli amaçlarından biri kullanıcıya en az iş yükü yükleyen bir yazılım hazırlamaktır. Mevcut yazılımlarda üretilen sonuç, tam olarak kullanıcıların istediği bir sonuç değildir. Kullanıcılar ortaya çıkan sonuç üzerinde düzenlemeler yapmak zorundadırlar. Hazırladığımız yazılım kullanıcının program tarafından sunulan dizin üzerinde herhangi bir düzenleme yapmasına ihtiyaç bırakmayacak şekilde oluşturulmuştur. Kullanıcılara sonuç aşamasında farklı biçimlerde dizin hazırlama olanağı sağlamak için Şekil 3.11’de görülen hedef dosya font ayarları formu hazırlanmıştır. Bu form sayesinde kullanıcı farklı görünümde dizin hazırlayabilecektir.

HEDEF DOSYA FONT

Özellikler

☒ Madde Başı ☐ Dil ☐ Satır

☐ Anlam ☐ Sayı ☐ Satır No

Seçenekler

Aharoni 12 **B** *I* U

İndeks : Dizin - 1 - [TR]

Ben yeni kitap i+imi oluşturdum 001a/01

Dizin İçin Satırda Bulunacak Satır No Sayısı 10

Font seçeneklerinin ayarlanması oldukça basittir. Özellikler bölümünden ilgili alanın seçilip seçenekler bölümünde istenilen ayarlamalar ile gerekli değişiklikler yapılabilir. Yapılan değişikliklerin görünümü aktif olarak formun alt kısmında görünmektedir.

3.8.6.2 Harf ayarları

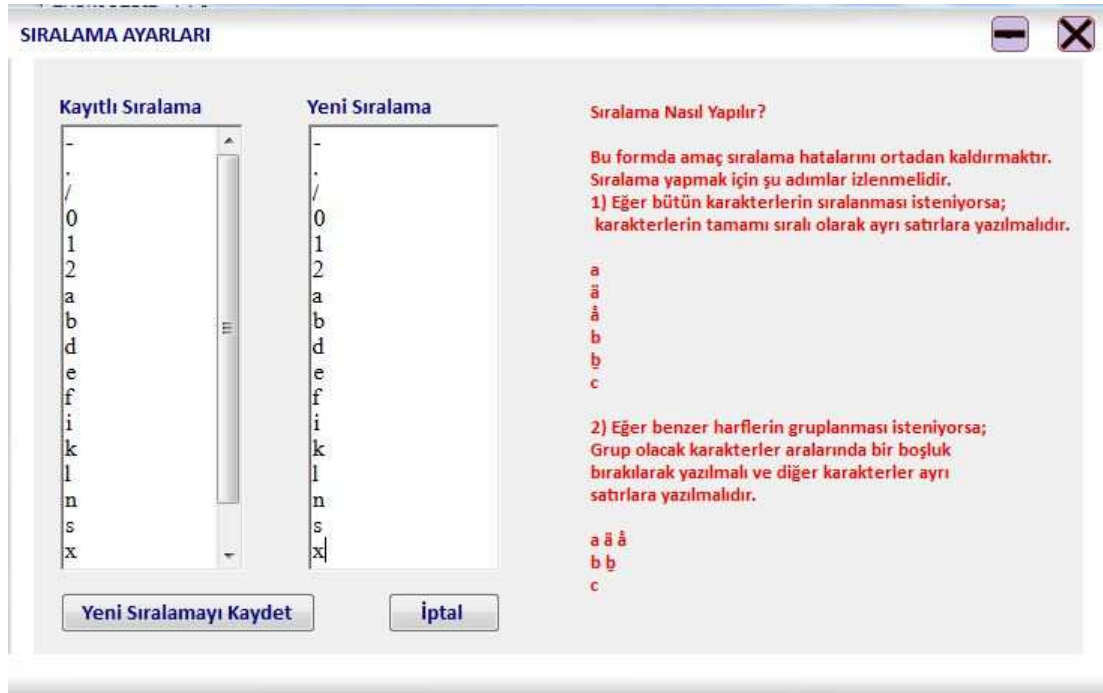
Sıklık sözlüklerinin hazırlanması esnasında metin içerisinde bulunan harfler özelliklerine göre gruplanmaktadır. Bu özellikleri belirtmek için Şekil 3.12’de gösterilen form kullanılmaktadır.

[illegible]

Şekil 3.12 : Harf ayarları form görünümü

3.8.6.3 Sıralama ayarları

Kullanıcıların karakterler sıralaması ile ilgili problemlerini ortadan kaldırmak için Şekil 3.13'te gösterilen form hazırlanmıştır. Bu form ile tüm karakterler istenilen sırada sıralanarak kaydedilebilir. Program sıralama işlemini bu kaydedilen sıraya göre gerçekleştirir.



Şekil 3.13 : Sıralama ayarları form görünümü

3.8.6.4 Renk ayarları

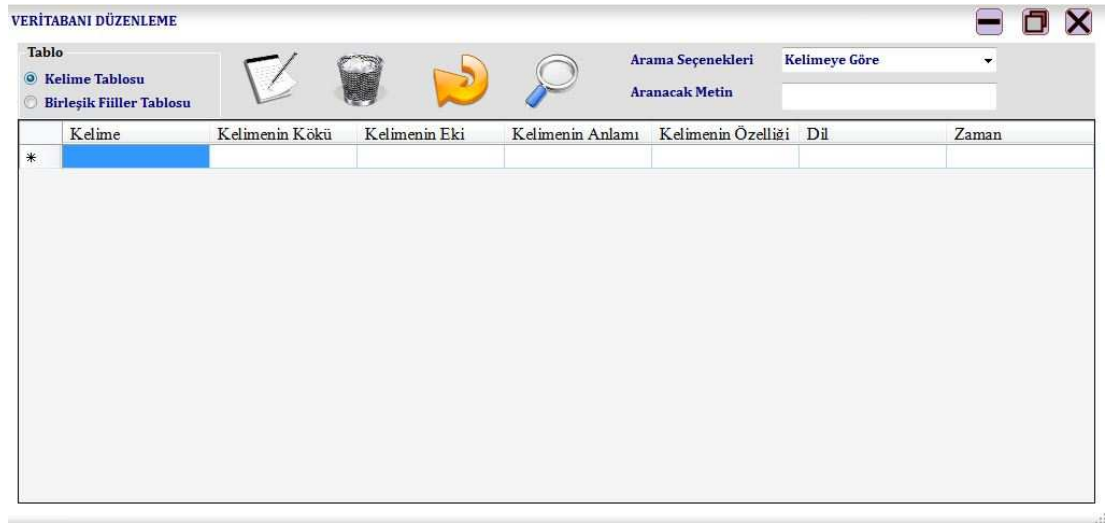
Ayrıştırma ve hata kontrollerinde kelimelerin arkaplan rengini belirlemek amacıyla oluşturulmuştur. Şekil 3.14'te renk çeşitleri gösterilmektedir.



Şekil 3.14 : Renk ayarları form görünümü

3.8.6.5 Veritabanı ayarları

Kullanıcıların ayrıştırma bölümünde hızlı ve doğru sonuçlar elde etmeleri için verilerin kaydedildiği bir veritabanı hazırlanmıştır. Veritabanı kelimelerin kaydedilmesi ve yeni ayrıştırmaların yapılmasıyla genişleyebilmektedir. Herhangi bir hatalı veri girişi durumunda kullanıcının veritabanında bulunan kelimeleri düzenlemesi amacıyla Şekil 3.15'te gösterilen veritabanı düzenleme formu oluşturulmuştur.

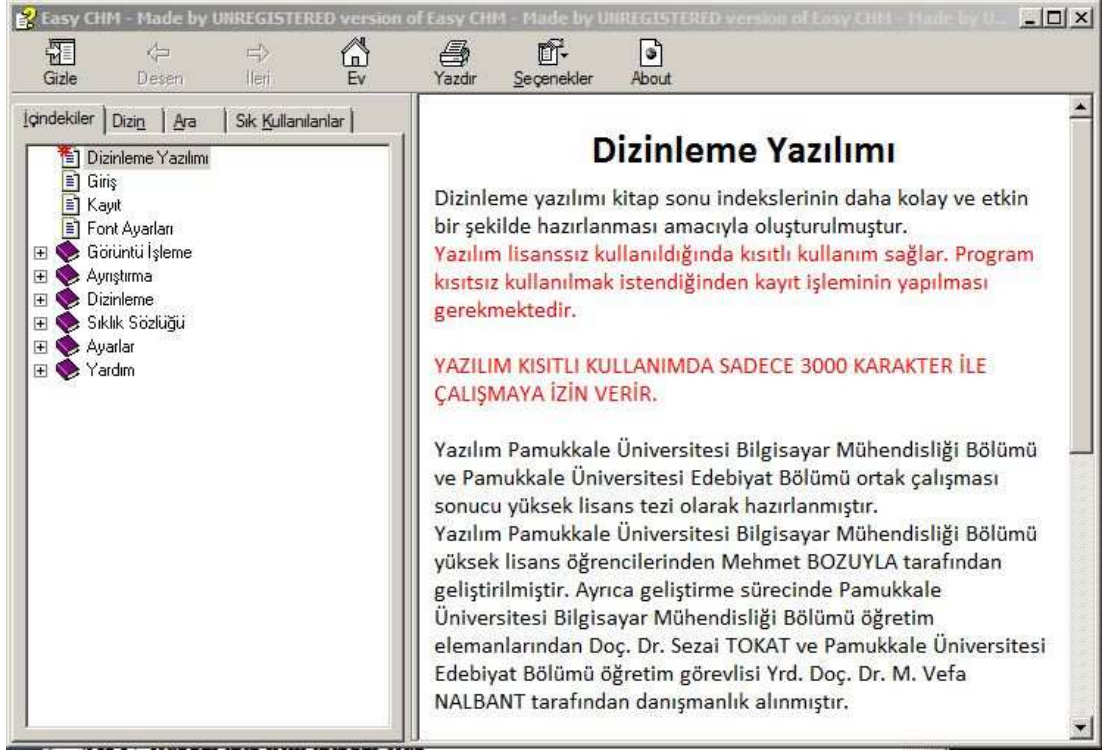


Şekil 3.15 : Veritabanı düzenleme form görünümü

Veritabanı düzenleme formu ile basit bir şekilde kelime ekleme, silme ve düzenleme işlemi yapılabilir. Ayrıca arama seçenekleri sayesinde veritabanında kelime aramada yapılabilir. Değişiklikler yapılırken değişikliğin yapılacağı tablonun seçilmesi ile tablo verileri ekrana getirilmektedir. DataGridView üzerinde gerekli düzenlemeler yapıldıktan sonra form üzerinde bulunan düğmeler yardımıyla değişiklikler kaydedilebilir.

3.8.7 Yardım

Hazırlanan yazılımın başarısında en önemli aşamalardan birisi de kullanıcılara hazırlanan bu programın nasıl kullanılacağını anlatılmasıdır. Programın nasıl kullanıldığı CHM formatında hazırlanan bir yardım dökümanı yardımıyla yapılmıştır. Hazırlanan dökümanın ekran görüntüsü Şekil 3.16'da gösterilmektedir.



Şekil 3.16 : Yardım dosyası

Kullanıcıların rahat anlayabileceği ve örnek çözümlerin sunulduğu bir yardım dosyası hazırlanmıştır. Kullanıcıların herhangi bir problem ile karşılaştıklarında sorunu bildirmek için Şekil 3.17’de gösterilen iletişim formu hazırlanmıştır. Bu form yardımıyla kullanıcıların aktif olarak problemlerini mail yoluyla bildirmeleri sağlanmıştır.

Şekil 3.17 : İletişim formu

3.8.8 Güvenlik

Yazılımlarda güvenlik en önemli aşamalardan birisidir. Programın izinsiz dağıtılması ve değiştirilmesini engelleyerek güvenliğini sağlamak gereklidir. Yazılımlarda güvenliği sağlamak için çeşitli yöntemler kullanılmaktadır. Bu yöntemler yazılım platformu ve kullanıcıların özelliklerine göre farklılıklar gösterir.

İnternet tabanlı yazılımlarda güvenlik kullanıcıya verilen kullanıcı adı ve şifre ile sağlanmaktadır. Her kullanıcı sahip olduğu benzersiz kullanıcı adı ve kendisinin belirlediği bir şifre ile yazılım ortamına erişim sağlar.

Kişisel bilgisayarlarda yazılım güvenliğini sağlamak İnternet tabanlı uygulamalara göre daha zordur. Bu zorluk yazılımın bilgisayara yüklenmesinden sonra üretici tarafından değiştirilmesinin mümkün olmamasından kaynaklanmaktadır.

Kişisel bilgisayarlarda çeşitli yöntemler kullanılarak yazılım güvenliği sağlanmaktadır. Çoğu yazılım üreticisi kurulum işlemini tamamlamak için kullanıcının sahip olduğu aktivasyon kodunun doğruluğunu internet aracılığı ile incelemektedir. Bu inceleme güncelleme esnasında yazılımın kayıtlı olup olmadığını kontrol etmektedir.

Aktivasyon kodları her bilgisayar için farklı olan MAC adresi veya işlemci adresi gibi benzersiz adresler kullanılarak yapılır. Üreticinin hazırladığı özel şifreleme algoritmaları ile bu adresler şifrelenir. Böylelikle aktivasyon kodları her bilgisayar için farklı olacak şekilde oluşturulur. Üretici aktivasyon kodlarının doğruluğunu analiz ederken kurulumun yapılmaya çalışıldığı bilgisayarın özelliklerini kullanarak yeni bir kod üretir. Üretilen kod ile kullanıcının sahip olduğu aktivasyon kodu karşılaştırılır. Kodların uyuşması halinde kurulum tamamlanır. Hata oluşması durumunda kullanıcının kurulumu iptal edilir.(WEB_14)

Hazırladığımız yazılımda mevcut güvenlik yöntemlerinden farklı bir yöntem denenmiştir. İki aşamadan oluşan bu şifreleme sisteminde Şekil 3.18’de görüldüğü gibi program kodu ve üretici kodu olmak üzere iki farklı kod bulunmaktadır.

Program Kodu

RR52 X5CG SZRX GPAT GD5X X5UU

Üretici Kodu

[TürkSözDiz Yazılımının Üretici Kodunu Almak İçin Buraya Tıklayınız.](#)

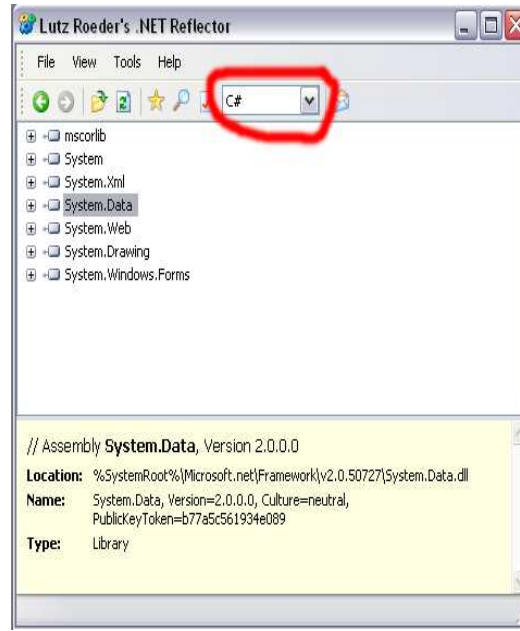
bm.pamukkale.edu.tr/~mbozuyla dizinleme@gmail.com

Şekil 3.18 : Kayıt formu

Program kodu, kurulum aşaması tamamlandıktan sonra program tarafından bilgisayarın benzersiz olan bazı özellikleri kullanılarak oluşturulmuş bir koddur. Üretici kodu ise üretici tarafından program kodunun kullanılması ile üretilen bir koddur.

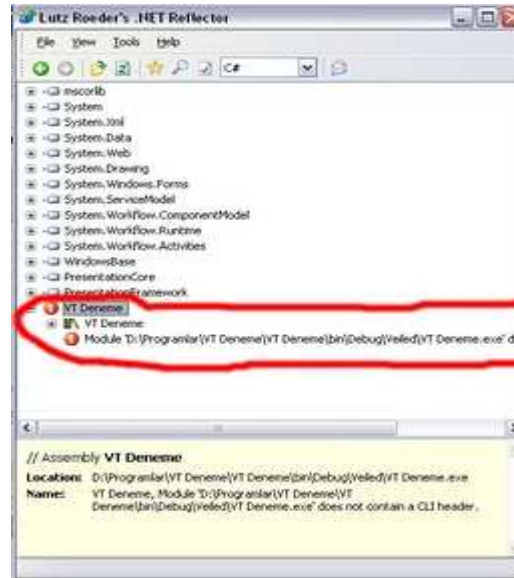
Yazılım özel kullanıcılara hitap ettiği için program kodunun üreticiye gönderilmesi ve üreticinin göndereceği üretici kodunun programda gerekli olan yere yazılması ile aktivasyon işlemi tamamlanmaktadır.

Yazılım geliştirme sürecinde karşılaşılan bir durum program kodlarının çeşitli yazılım araçları ile görüntülenebiliyor olmasıdır. Şekil 3.19'da görüldüğü gibi Reflector yazılımı gibi kod görüntüleyen yazılımlar ile hazırlanan yazılımın kodları görülebilmektedir.(WEB_15)



Şekil 3.19 : Reflector program arayüzü

Yazılım kodların görülmesi üreticinin yazılım üzerindeki hâkimiyetini kaybettirdiğinden kodların gizlenmesi gerekmektedir. Dotfuscator, Postbuild, CodeVeil gibi yazılımlar Şekil 3.20’de görüldüğü gibi yazılımın kodlarını gizlemekte kullanıcılar tarafından müdahale edilmesini engellemektedir.



Şekil 3.20 : Reflector kod gösterimini engelleme

Yazılımın dağıtımı esnasında sürekli kullanım veya belli sayıda kullanım yöntemleri de düşünülmüştür. Fakat izin hazırlama işleminde kullanıcı kaynaklı hataların fazla olabilmesi ve izin hazırlama işleminin yazılım sayesinde 2-3 ay içerisinde tamamlanması bu yöntemlerin uygulanabilirliğini ortadan kaldırmıştır.

3.8.9 Performans analizi

Bir yazılımın en önemli özelliklerinden birisi de performans analizinin iyi yapılmasıdır. Bu çalışmada da yazılımın performansını test etmek amacıyla çeşitli karşılaştırmalar yapılmıştır. Yapılan tüm performans analizlerinin sonucunda mevcut sistemlere göre çok daha yüksek performansta bir yazılım hazırlandığı görülmüştür.(WEB_16)

İlk olarak farklı işlemci ve RAM gücüne sahip bilgisayarların dizin türlerine göre performansı incelenmiştir. Beş farklı bilgisayarda elde edilen sonuçlar Tablo 3.2’de görülmektedir. Yapılan bu analize göre işlemci ve RAM gücüne paralel olarak dizin alma hızının azaldığı görülmektedir.

Tablo 3.2 : Bilgisayar özellikleri ve dizinleme türüne göre performans analizi

	Bench Mark	Gramatiksel Dizin	Bağlam Sözlüğü	Sıkıştırılmış G. Dizin	Ek Dizin
Intel M460 Core i5 2,53GH 3GB	687 puan	80 Saniye	163 Saniye	61 Saniye	5 Saniye
Intel T7500 Core2 Duo 2,2GH 2 GB	639 puan	100 Saniye	192 Saniye	87 Saniye	7 Saniye
Intel P7450 Core2 Duo 2,13GH 3GB	742 puan	115 Saniye	220 Saniye	104 Saniye	9 Saniye
Intel Dual Core T4200 2,0 GH 3GB	639 puan	119 Saniye	240 Saniye	109 Saniye	10 Saniye
Intel Dual Core T2330 1,60 GH 3GB	434 puan	154 Saniye	317 Saniye	140 Saniye	12 Saniye

PerformanceTest yazılımı ile Tablo 3.2’de bulunan bilgisayarların BenchMark değerleri ölçülmüştür. Tabloda da görüldüğü üzere BenchMark değerleri bilgisayar özelliklerine paralel olarak değişmektedir. Yalnız Intel P7450 işlemcili bilgisayarda BenchMark yüksek ölçülmüştür. Bunun sebebi de ekran kartının paylaşımlı olması olabilir.

Tablo 3.3 ile yaptığımız TürkSözDiz yazılımı ve mevcut dizinleme yazılımları dizin türlerine göre karşılaştırılmıştır. Mevcut sistemlerde olmayan dizin türleri “YOK” olarak nitelendirilmiştir.

Tablo 3.3 : Programlara göre dizinleme türleri performans analizi

	Gramatiksel Dizin	Bağlam Sözlüğü	Sıkıştırılmış G. Dizin	Ek Dizin
Cibakaya	YOK	YOK	10* Saniye	YOK
Concordance	YOK	10* Saniye	YOK	YOK
TürkSözDiz	113 Saniye	226 Saniye	100 Saniye	8 Saniye

Tablo 3.3’te ve Tablo 3.5’te yapılan analizlere göre Cibakaya ve Concordance yazılımları daha kısa sürede işlemleri tamamlamaktadırlar. TürkSözDiz yazılımında bulunan font hatalarının giderilmesi, kullanıcı isteğine göre farklı biçimlerde dizin hazırlama, hazırlanan dizinlerin daha kapsamlı olması ve son olarak tüm dizinin kullanıcı tarafından analiz edilebilmesi için ekrana yansıtılması sürelerin farklı olmasına sebep olmaktadır. Cibakaya dizin yazılımının kullanıcılara herhangi bir değiştirme hakkı tanımaması ve Concordance yazılımının ayrıştırma yapmaksızın sadece kelimeleri alfabetik sıralaması dizin hazırlama sürecinde son kontrollerin veya ayrıştırma sürecinin uzun zaman almasına sebep olmaktadır. Bu süreler Tablo 3.5’te görülmektedir.

Yazılımın hazırlanması sürecinde ilk aşamada dizi mantığı kullanılarak hazırlanan algoritma performans olarak yeterli görülmemiştir. Performansı artırmak için bağlı liste yapıları kullanılarak yeni bir algoritma oluşturulmuştur. Algoritmaların performans değerleri Tablo 3.4’te verilmiştir. Sonuç olarak bağlı liste mantığı kullanılarak hazırlanan algoritmanın performans olarak daha verimli sonuçlar verdiği görülmektedir. Ayrıca veritabanına kelimelerin kaydedilmesi ve veritabanından dizinleme işleminin yapılması düşünülmüş yapılan testlerde bu yöntemin zaman olarak verimsiz olduğu görülmüştür.

Tablo 3.4 : Algoritmalarla göre dizinleme türleri performans analizi

	Gramatiksel Dizin	Bağlam Sözlüğü	Sıkıştırılmış G. Dizin	Ek Dizin
Dizi sistemi kullanılarak hazırlanan algoritma	280 Saniye	360 Saniye	230 Saniye	78 Saniye
Bağlı listeler kullanılarak hazırlanan algoritma	100 Saniye	192 Saniye	87 Saniye	7 Saniye

Tablo 3.5’te bütün bir dizini oluşturmak için hazırlanan süreçlerin farklı yazılımlara göre performans analizi yapılmıştır. Tabloda da görüldüğü üzere en kapsamlı ve en yüksek performansı sağlayan yazılımın yeni oluşturulan TürkSözDiz yazılımı olduğu görülmektedir.

Tablo 3.5 : Programlara göre dizinleme aşamalarında gereken toplam süreler

	Ayrıştırma	Dizinleme	Sıklık Analizi	Son Kontrol
Cibakaya	2 ay	10* Saniye	YOK	3 ay
Concordance	YOK	10* Saniye	YOK	5 ay
Metin Sıklık Çözümleyicisi	YOK	YOK	10 Saniye	YOK
TürkSözDiz	5 gün	100 Saniye	3 Saniye	15 gün

Tablo 3.5’te görüldüğü gibi diğer yazılımlarda tek bir tür dizin almak için 5 aylık bir süre gerekirken TürkSözDiz yazılımı 20 gün içerisinde farklı türlerde ve daha detaylı dizin hazırlama olanağı sağlamaktadır. Cibakaya ve Concordance yazılımlarında dizinleme aşaması 10 saniyede tamamlanmasına rağmen dizinleme esnasında yapılamayan işlemlerin kullanıcı tarafından yapılması için son kontrol ve ayrıştırma işleminde aylar süren bir çalışma yapmak gerekmektedir. Buna karşın TürkSözDiz yazılımı kullanıcının tüm yükünü alarak bütün işlemlerin otomatik olarak yapılmasını sağlamaktadır. Her durumda gerekli olan son kontrol kullanıcı bazlı hataların düzeltilmesi amacıyla yapılmaktadır. Bu hatalar kelimelerin yanlış

yazılması veya harflerin yanlış kodlanması gibi hatalardır. Fakat Cibakaya ve Concordance yazılımında yapılan son kontroller kullanıcı hatalarından çok yazılımın eksikliklerinden kaynaklanmaktadır. Font uyumsuzluğu, ayırıştırma işleminin yapılmaması, birleşik fiillerin veya tamlamaların belirtilmesi gibi durumlar buna örnek olarak verilebilir. Bir bütün halinde dizin hazırlama işlemi düşünüldüğünde verimli ve kullanıcı dostu bir yazılımın oluşturulduğu görülmektedir.

4. SONUÇ VE ÖNERİLER

Hazırlanan bu tez çalışmasında bağlı listeler ile hazırlanan dizinleme algoritması dizi yapısı kullanılarak oluşturulan algoritmaya göre performans açısından çok daha verimli sonuçlar verdiği görülmüştür.

Proje içerisinde bulunan görüntü işleme modülü transkripsiyon harflerinin tanınması esnasında hatalı sonuçlar vermektedir. Bu hatalı sonuçların ortadan kaldırılması için yapay zekâ teknikleri kullanılabilir. Tespit ettiğimiz bir sorun, kullanılan transkripsiyon harflerinin şekilsel olarak birbirine çok yakın olmasıdır. Bu da hassas bir algoritmanın tasarlanmasını gerektirmektedir.

Güvenlik açısından bilgisayarın işlemci numarası veya yerel ağ bağlantısına ait MAC adresi kullanılabilir. Bu numaralar her bilgisayar için farklıdır.

Ayrıştırma sürecinde daha iyi sonuçlar alabilmek için yapay zekâ yöntemleri kullanılabilir. Dil çalışmalarında değişken fazla olduğundan yapay zeka yöntemleri uygulamak için uzun süreye ve geniş ekiplere ihtiyaç vardır. Ayrıca dilde belli matematiksel yapı oluşturulmadığından yapay zekâ yöntemleri sonuçsuz kalabilir.

Ayrıştırma sürecinde dilde bulunan bütün kelime ve ses yapılarını kapsayan bir veritabanı oluşturulabilir. Hazırlanacak veritabanında bağlantıların güçlü olması gerekmektedir. Aksi durumda performans kaybı yaşanabilir. Ayrıca veritabanında bulunan veriler hatasız oluşturulmalıdır. Böyle bir veritabanının hazırlanması dil çalışmaları açısından farklı çalışmaların oluşturulmasına zemin hazırlayacaktır.

Tüm kullanıcıların ortak olarak yararlandığı internet ortamında bulunan bir veritabanı oluşturulabilir. Böylelikle farklı kullanıcılar tarafından oluşturulan dizinlerden yararlanılmış olur. Ayrıca geniş bir veritabanı elde edilebilir. Bu işlemin sakıncası yanlış kelime kaydetmeye bağlı olarak veritabanının bozulmasıdır.

Geliřtirdiđimiz ayrıştırma notasyonu yeni karakter ve özelliklerin eklenmesi ile geliştirilebilir.

Reflector gibi kod görüntülemek için kullanılan yazılımların etkisini ortadan kaldırmak için Dotfuscator, Postbuild, CodeVeil gibi yazılımlar kullanılabilir.

KAYNAKLAR

- ALTUN M.** (2003), “İbrahim İbn-İ Bali’nin Hikmet-Name’si (1b-149a) İnceleme-Dizin”, *T. C. Kültür ve Turizm Bakanlığı Kütüphaneler ve Yayımlar Genel Müdürlüğü*, İstanbul
- BAYDUR G.** (1988), “Kitap Sonu İndeksleri”, *Türk Kütüphaneciliği* 2, ss.61-70
- GÖZÜDELİ Y.** (2007), “Yazılımcılar İçin SQL Server 2005 ve Veritabanı Programlama”, Seçkin Yayıncılık
- HENGİRMEN M.** (1999), “Türkçe Temel Dilbilgisi”, Engin Yayınevi, Ankara
- İLERİ C.** (2010), “Türkiye Türkçesi Ağız Araştırmaları İçin Bir Yöntem Denemesi”, *Turkish Studies*
- KAYA C.** (2006), “Windows Altında Dil Çalışmaları İçin Cibakaya 2.2 Kullanım Klavuzu”, İstanbul
- ÖÇALAN M.** (2006), “Ağız Araştırmalarında Bilişim Teknolojilerinin Kullanılması ve Ağız Tezleri İçin Yenilik Önerileri”, *Sosyal Bilimler Enstitüsü Dergisi*, Sayı 20
- RIORDAN M. R.** (2002), “Microsoft ADO .NET Step by Step”, Microsoft Press
- SARI İ.** (2008), “Türkoloji Çalışmalarında Teknolojinin Önemi”, Gazi Üniversitesi,
- UÇAR İ.** (2009), “Haza Kitab-ı Hulasa-i Tıbb Cerrah Mes’ud”, Basılmamış Yüksek Lisans Tezi, Sakarya Üniversitesi, Sosyal Bilimler Enstitüsü
- YANIK M.** (2007), “Microsoft Visual C# .Net”, Seçkin Yayıncılık
- WEB_1,** Kurt A., KARA M., “Metin Sıklık Çözümleyicisi”
<http://nlp.ceng.fatih.edu.tr/~metinanaliz/>
- WEB_2,** ALTUN M., “Dil Bilimi Araştırmaları İçin Bilgisayar Yazılımları”
<http://www.dilbilimi.net/yazilim.htm>

- WEB_3,** İŞLER S., AMASYA M.F., TATLI E. İ., “Kelimenin Cümle İçindeki Durumuna Bakılarak Kök ve Eklerine Ayrılması”,
<http://www.ce.yildiz.edu.tr/mygetfile.php?id=1512>
- WEB_4,** http://tr.wikipedia.org/wiki/Birle%C5%9Fik_fiil
- WEB_5,** “Concordance Manuel For Version 3.3” ,
<http://www.concordancesoftware.co.uk/manual/Concordance33Manual.pdf>
- WEB_6,** KAYA C.,”Cibakaya Dizin Yazılımı”,
http://www.cevalkaya.com/?page_id=15
- WEB_7,** PEKACAR Ç.,SARI İ.,”Dizin Hizmet Programı”, <http://www.isa-sari.com/harici/dizin/>
- WEB_8,** TAŞDELEN T., “UML ile Sistem Analizi – 3 Use Case Modelleme”,
<http://www.gelistir.org/blog/uml-ile-sistem-analizi-%E2%80%93-3-use-case-modelleme/>
- WEB_9,** http://tr.wikipedia.org/wiki/Yazılım_mühendisliği
- WEB_10,** <http://social.msdn.microsoft.com/Forums/en/isv/thread/f7019ee2-61c7-407e-8c45-d90b3582a6c1>
- WEB_11,** YANIK M., “C# Uygulamalarında Word'den Yararlanmak”,
http://www.ceviz.net/c-uygulamalarinda-wordden-yararlanmak_a1221.html
- WEB_12,** NACAROĞLU A. F., “C# ile .NET Ortamında Threading'e Giriş”,
<http://www.csharpnedir.com/articles/read/?id=147>
- WEB_13,** “Rich Text Format (RTF) Specification,version 1.6”,
<http://latex2rtf.sourceforge.net/rtfspec.html>
- WEB_14,** ÖZTÜRK E., “.NET Code Security”,
<http://www.olympus.net/belgeler/dotnet/net-code-security-156213.html>
- WEB_15,** YAĞMUR O., “.NET ve Uygulama Güvenliği”,
<http://www.csharpnedir.com/articles/read/?filter=&author=60&cat=&>

[id=899&title=.NET%20ve%20Uygulama%20G%C3%BCvenli%C4%9Fi%20%20-%201](#)

WEB_16, http://www1.cse.wustl.edu/~jain/cse567-06/sw_monitors1.htm

EKLER

EK A.1 :

Tablo A.1: Transkripsiyon harfleri

ا (ā)	a, ā	ص	ş
ا (ā)	a, e, ı, i, u, ü	ض	ž, đ
ب	b, p	ط	ṭ
پ	p	ظ	ẓ
ت	t	ع	‘
ث	ṯ	غ	ğ
ج	c, ç	ف	f
چ	ç	ق	k
ح	ḥ	ك	k, g, (ñ)
خ	ḫ	ك	ñ
د	d	ل	l
ذ	ẓ, ḏ	م	m
ر	r	ن	n
ز	z	و	v, u, ū, ü, o, ö
ژ	j	ه	h, a, e
س	s	لا	la, lā
ش	ş	ی	y, ı, i, ī
		ء	’

Orijinal Metin

بوردن كوردن حقت وجودى		سرآب فى سرآب فى سرآب	
﴿ نيازى جسم وقلب وروح كه دینور ﴾ ﴿ جناب فى جناب فى جناب ﴾			
استر ایسك معرفده اولهسن عالجناب	اهل عرفان ایسكندره یوزی ایله تراب	چوقده ویرمه کندی دنیا به برده چک ایاک	دو بند برده مزین بکم قانی آغز درو طولا ب
یو خرافی نیجیل چالشی معمور ایتک	ریان تمیر ایدرکن برانی اولدی خراب	چوق سکرندی غشیله بوسرای صوسانوب	بولدیله هیچ بری بومهراده برقطره آب
کوز قولا ق دل قابولرین باغله حکم بر زمان	اوله کیم حقدن بکا کولکلن اوله فتح باب	کراولدن قورتلیم دیرسن یوری عاشق اول	دو نه دونه عشق اودله جسم و جان قیل کاب
کیر بودرد میخانه سین قومه الدن کاسه بی	هیچ بودرک تاندن اوز که عاشقه بو قدر شراب	هتک دایم بواولسونکیم حق اکلایسن	حق بیلمکدن یک اولز ایکی عالده صواب
کر عذاب آخرتدن بولق استرک خلاص	عارف اول جهل اودندن قویسر جله عذاب	﴿ بو نیازی کندی دیر بو سوزی ای پسر ﴾ ﴿ هب آتی شو بولر طوبیاسین کو کدن این ددت کتاب ﴾	
آج کوزک دلداره باق آچلیدی یوزندن قباب	طلسمی سوردی چقاردی آرم بردن آتاب	شول ستمهم دیم خرین لبدن ایجه کور	قطره سین نوش ایلین عشاق ابد کور من عذاب
اوتوز ایکی حرف بیلک ددت کتابک اسلیدر	سفته و جهنمه بالمش قوسوی ارتاب	مکتب مرهاته کبر اوقی بو علسک اسلای	کورد نیجه درج اولش و علمده ددت کتاب
هر نه اوقورسن اوتوز ایکی دن طشره دکل	طوت قولاغک کسکادر جله دلدرن خطاب	﴿ هر نه کیم کور کوزک آندن جمال یارماق ﴾ ﴿ چونکه کندی ای نیازی قالدی اسلا حجاب ﴾	

Transkripsiyon

İster işen ma'rifetde olasin 'âlîcenâb
Eli-i 'irfân iğinde yüzini eyile turâb

Çok da virme kendini dünyâya bir dem çek elin
Döndiremezsin begim katı ağırdır bu tolâb

Bu hârâbı nicele çalırdı ma'mûr imcege
Bir yanın ta'mîr iderken bir yanı oldı hârâb

Çok seğırdi şafetleyle bu sarâyı şû şâmûb
Bulmadılar hiç biri bu şârlâda bir katre âb

Bir zamân yüz virme dünyâ eliline 'uzletde ol
Akl-fikrin bir yere cem' it yüzüne çek nikâb

Göz kulâk dil kâpûların bağla muhkem bir zamân
Ola kim hâkdan yana göhûlden ola fetvâ-i bâb

Ger ölümden kurtulam dîrsen yürü 'âşık ol
Dâne dâne 'aşk odıyla ile cîm ü câmî kıl kebâb

Gir bu derd meyhlânezine koma elden kâzeyi
Hiç yürek kânnıdan eğer 'âşkâ yokdur çarâb

Himmetin dâ'im bu olsun kim hâkkı anlayasın
Hâkkı bilmekten yeg olmaz iki 'âlemde şavâb

Ger 'âğâb-ı âhîretiden bulmak istersen halâş
'Ârif el cehl odından kopıser cümle 'âğâb

Bu Niyâzî kendiden dîmez bu sözü ey püser
Hep anı söyler tıyarsın gökden inen dört kitâb

Aç gözün dil-dâre bak açıldı yüzünden nikâb
Zulmeti sürdi çıkardı ara yerden âfritâb

Şol âlkahum rabbuhum hânnın lehinden içegör
Kâtresin nûş eyleyen 'uşşâk ebed görmez 'âğâb

Otuz iki harf bildin dört kitâbın aşıldır
Şafla-i vecinde yazılmış kânnûs bi-irtiyâb

Mekteb-i 'irfân gir oku bu 'ilmîn salın
Gör nice dere olmuş bu 'ilme dört kitâb

Her ne okursan otuz ikiden taşra degil
Yüzünün metnini şerh ider o kurbân faş u bâb

Her ne söz ki söylenür 'âlemde Türkî yâ 'Arab
Tür kulağın ki sâhâdır cümle dillerden iştirâb

Divân-ı Niyâzî, Dersaadet: Emniyet Kütüphanesi, 1325, s. 11

Şekil A.1 : Osmanlıca divan ve Transkrip edilmiş metin

ÖZGEÇMİŞ



Ad Soyad: Mehmet BOZUYLA
Doğum Tarihi ve Yeri: 26. 08. 1988 ELAZIĞ
Adres: Değirmenönü Mah. 1334. Sok. No:6 DENİZLİ
Lisans Eğitimi: Pamukkale Üniversitesi Bilgisayar Mühendisliği Bölümü