

TÜRKÇE İÇİN OKUMA FONKSİYONLU OTOMATİK METİN OLUŞTURMA SİSTEMİ

Şerif ADALI¹Yaşar ERENLER²

^{1,2}İstanbul Teknik Üniversitesi, Elektrik-Elektronik Fakültesi
Bilgisayar Mühendisliği Bölümü, 80626 Maslak, İstanbul

¹e-posta: serifadali@yahoo.com²e-posta: erenler@cs.itu.edu.tr

Anahtar Sözcükler: Yazıdan Konuşma Üretimi, Metin Üretimi, Doğal Dil İşleme.

ÖZET

Bu çalışmada Türkçe için otomatik metin üretim ve konuşma üretim sistemi geliştirilmiştir. Örnek olarak bir diş hekiminin hastalarına ait verileri incelenip bu verilerden Türkçe dil kurallarına göre doğru şekilde cümleler kurarak hasta ile ilgili rapor oluşturmaya ve bu raporun yazıdan konuşmaya dönüştürücü modülü yardımıyla diş hekimine Türkçe konuşma olarak okunmasına çalışılmıştır. Görsel araçlar kullanılarak hasta hakkında bir veri tabanı oluşturulmuştur. İstenilen hasta seçilerek rapor oluşturulmakta ve rapor daha sonra kullanıcıya Türkçe olarak okunmaktadır.

kurallar veritabanı tablolarında tutulmakta ayrıca geçici tablolarda da kelime tiplerine göre hastaya ait veriler tutulmaktadır.

Tablo 1 : Sonlu Durum Makinasında Kullanılan Semboller.

Durum	Sesli	Sessiz	Durum	Sesli	Sessiz
başlangıç	v	c	evc	ev	evcc
c	cv	cc	evcc	ev	evccc
cc	ccv		evccc	ev	evcccc
ccv	v	ccvc	evcccc	ccv	cc
ccvc	cv	ccvcc	v	v	vc
ccvcc	cv	ccvccc	vc	cv	vcc
ccvccc	cv	cc	vcc	cv	vccc
cv	v	cvc	vccc	cv	vcccc
			vcccc	ccv	

1. GİRİŞ

Bu çalışmada doğal dil işlemenin bir kolu olan metin üretme ve yazıdan konuşma üretme konuları üzerine yoğunlaşmıştır. Genel amaç, bir hasta veritabanından alınan verileri Türkçe dilinin kurallarına uygun şekilde dizerek anlamlı ve dilbilgisi kurallarına uygun raporlar oluşturmaktır. Buna ilave olarak, oluşturulan Türkçe metnin sesli olarak kullanıcıya okunması da amaçlanmıştır. Bu yüzden öncelikle Türkçe gramer ile ilgili kurallar üzerinde çalışılmıştır. İkinci olarak Türkçe heceleme kuralları için bir sonlu durum makinesi oluşturulmuş ve fonemleri doğru şekilde heceleme yapacak bir fonksiyon geliştirilmiştir.

Çalışmamızda, E.Gündüz tarafından [1] Perl kullanılarak gerçekleştirilmiş olan Türkçe için heceleme sonlu durum makinasından faydalanılmıştır. Şekil 1'de bu sonlu durum makinası görülmektedir. Tablo 1'de ise, Şekil 1'de kullanılan semboller verilmiş ve sözcüklerin bitebilecekleri durumlar altı çizili olarak gösterilmiştir. Tabloda v sembolü sesli harfi (vowel), c sembolü ise sessiz harfi (consonant) göstermektedir.

Çalışmadaki en önemli kısımlardan biri veri tabanına giriş yapılan verilerin cümle içinde hangi görevi alacağı ve ne şekilde sıralanmaları gerektiğine karar veren mekanizmadır. Bunun için gerekli gramer

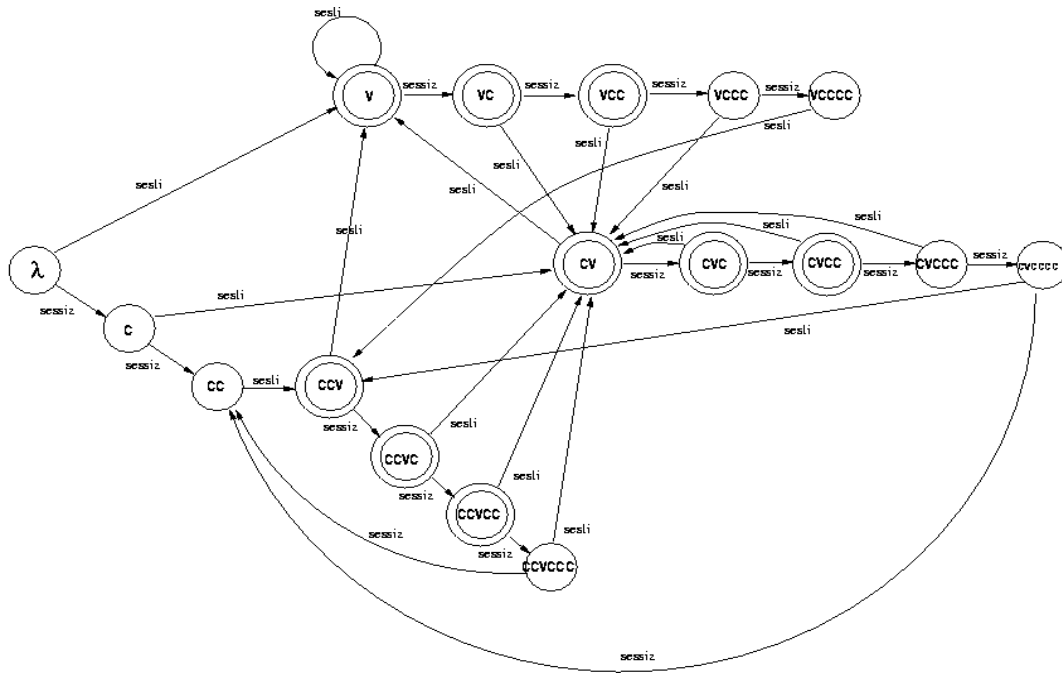
2. SİSTEMİN GENEL YAPISI

Sistem genel olarak iki bölümden oluşur:

- 1) Metin Üretme Alt Sistemi
- 2) Metin Okuma Alt Sistemi

2.1. Metin Üretme Alt Sistemi

Öncelikle sisteme dahil edilen bilgilerin her biri kelime tiplerine göre çeşitli geçici tablolarda tutulmaktadır. Ayrıca veritabanında ilişkisel tablolarda hangi kelime tipinin hangisinden önce gelebileceği konusunda yapısal ilişkiler tutulmaktadır. Böylece kural dışı cümleler kurulmaması için gramer yapısı tüm detayları ile veritabanında tutulmaktadır. Örneğin hastanın adı girildiğinde biz bunu hasta ile ilgili isimler tablosuna atıyoruz. Böylelikle hastanın adının geçtiği bir cümle oluşturduğumuzda hastanın isminin cümlede neresinde olması gerektiğini temel gramer kurallarını belirlediğimiz bu tablolarda yaptığımız sorgulamalar sonucunda verebiliriz. Hastanın şikayetleri ile bu şikayetlere bağlı olarak teşhisler ve tedaviler ilişkisel olarak tutulmaktadır.



Şekil 1 : Türkçe Heceleme Sonlu Durum Makinası.

Şekil 2'deki tablo yardımı ile kullanıcılar gramerde kullanılan cümle öğelerinin sıralamasını dinamik olarak belirleyebilmekte ve böylelikle istenilen sıralama yapılmaktadır. Cümle içinde öğelerin sıralaması için M.Hengirmen'in kitabından [2] faydalanılmıştır. Tüm hasta raporları bu sıralama tablosuna bağlı olarak oluşturulmaktadır.

Cümlelerde geçen kelimelerdeki ad durum eklerini bulmak için her kelimenin sonunda Şekil 3'te gösterilen tablodaki eklerden birinin olup olmadığı kontrol edilir. Buna göre doğru bir cümle kurulup kurulmadığı hakkında sonuç elde edilir.

Sıra No	Öğesi Adı
1	Özne
2	Belirteç
3	İlgeç
4	Tamlayıcı
5	Yüklem

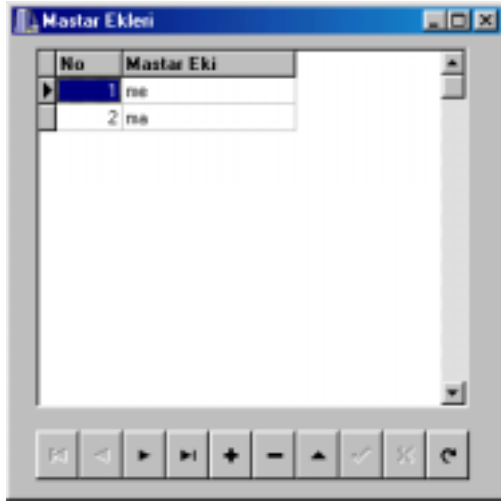
Şekil 2 : Öğesi Sıralama Veri Tablosu.

Tip No	Ad Durum Eki
1	den
2	dan
3	ten
4	tan

Şekil 3 : Durum Ekleri Tablosu.

Şekil 4'deki tablo sayesinde cümlede geçen kelimelerden hangilerinde mastar ekleri bulunduğu saptanır böylelikle cümle dizilişinde doğru kelimelerin kullanılıp kullanılmadığı hakkında fikre varılabilir.

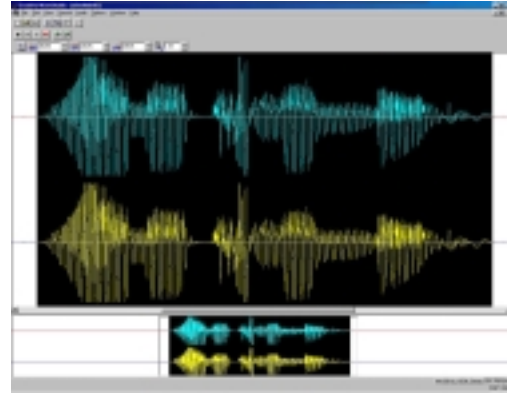
İçinde ek fiil geçiş ekleri bulunan kelimelerin yakalanması için ise Şekil 5'teki tablo kullanılır. Sistemimizde oluşturulan cümleler belli şablona sahip oldukları için bu cümlelerde kullanılan kelimeler ve cümle öğeleri tiplerine ve öğesi sıralarına göre kontrol edilip daha sonra gerekli rapor oluşturulmaktadır.



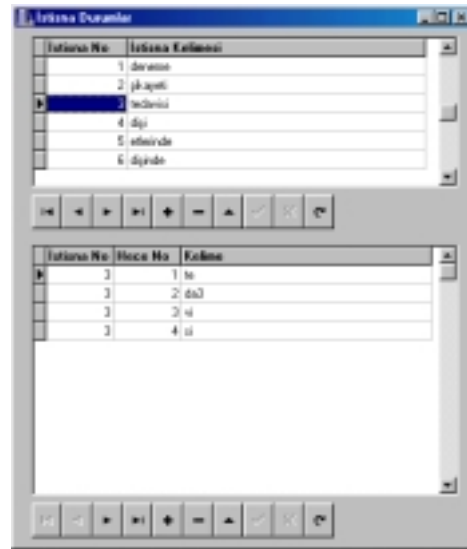
Şekil 4 : Master Ekleri Tablosu.



Şekil 5 : Geçiş Ekleri Tablosu.



Şekil 6 : Seslerin Kesildiği Program.



Şekil 7 : Mükemmel Ses İçin İstisna Tablosu.

2.2. Metin Okuma Alt Sistemi

Çalışmamızda oluşturulan cümlelerde bulunan her kelime heceleme algoritması yardımı ile hecelenmekte ve daha sonra uygun ses dosyaları bulunarak bu dosyalar birleştirilerek arka arkaya okunmaktadır. Böylelikle oluşturulan rapor aynı zamanda sesli olarak kullanıcıya okunmuş olur. Bu yüzden Türkçe’de yer alan 2, 3, 4 harften oluşan tüm temel heceler bir ses örnekleme programı ile oluşturulmuş standart ses dosyası (wav) olarak kaydedilmiştir.

Şekil 6’daki ekran görüntüsünde, hecelerin kaydedildiği program görülmektedir. Bu program sayesinde her türlü hece tek tek okunup en mükemmel haliyle kesilmektedir. Ses tonu ve seviyesi ayarlanıp kaydedilmiştir.

Şekil 7’de görüldüğü gibi, okunan bir kelimede eğer istenilen kalite elde ediliyorsa bu kelimenin daha kaliteli bir şekilde okunması için istisnalar tablosu oluşturulmuş ve her hecede hangi ekin okunacağı belirtilmiştir.

3. UYGULAMA

Bu uygulamada bir diş hekimi tarafından kullanılabilecek olan bir kullanıcı arayüzü oluşturulmuştur. Şekil 8’de görüldüğü gibi önce kayıtlardan raporu üretilmek istenen hasta seçilmekte ve daha sonra rapor oluşturma tuşuna basılarak gramer kuralları çalıştırılmakta ve eldeki verilerle mantıklı cümlelerden oluşan dil kurallarına uygun bir rapor çıktısı oluşturulmaktadır. Daha sonra ise heceleme modülü oluşan raporu ele alarak hecelere ayırır. Bundan sonra hecelere karşılık gelen ses dosyaları birleştirilerek raporu Türkçe olarak kullanıcıya okur.

4. SONUÇLAR

Bu çalışmada kullanılan yöntem diğer konuşma programlarında kullanılan sentezleme metodundan farklı olarak örneklenmiş ses dosyalarının gerçek zamanda birleştirilip ard arda çalınması şeklindedir. Bu nedenle ses çıktılarındaki kalite gayet yüksektir.

Konuşma üretiminin temelinde örneklenmiş seslerin doğru şekilde birleştirilerek okunması yatmaktadır. Eksik ekler de örneklenerek mükemmel şekilde

konuşan bir program haline getirilebilir. Kalıplarda daha esnek bir yapı ile daha çeşitli cümlelere izin verilebilir.

Cümlelerin doğru bir gramerde oluşturulması ve daha sonra oluşturulan cümlelerin en mükemmel hali ile okunması için istisna durumları da kontrol edildikten sonra başarının gerçekten yüksek olduğu sonucu ortaya çıkmıştır.

[illegible]

Şekil 8 : Kullanıcı Arayüzü.

5. KAYNAKLAR

[1] Engin Gündüz, “*Türkçe Hecelime İçin Bir Sonlu Durum Makinesi*”, 2002.

<http://www.ripe.net/home/engin/heceler.html>

[2] Mehmet Hengirmen, "*Türkçe Temel Dilbilgisi*", Engin Yayınevi, Ankara, 1999.

[3] Borland Developers Network, *C++ Builder Newsgroup*.