

# TÜRKÇE HASTA HİKAYELERİNDEN ANAHTAR KELİME ANALİZİYLE AKILLI TEŞHİS SİSTEMİ

An Intelligent Diagnostic System from Turkish Medical Narratives by Keyword Analysis Technique

M. Oğuzhan Külekci \*

## Özet:

Bu yazıda Türkçe hasta hikayelerinden anahtar kelime analizi yöntemiyle akıllı bir teşhis sistemi geliştirilmesi öngörülmektedir. Bu sistemin temeli geliştirilen Türkçe biçimbirimsel çözümleyici ile hasta hikayesinde geçen tüm kelimelerin köklerini tespit etmek ve bulunan bu kökleri bilgi ambarındaki anahtar kelimelerle karşılaştırarak hastanın rahatsızlığını sınıflandırmaktır. Bu amaç dahilinde sondan eklemeli dillerde kelimelerin morfolojik analizini yapmak üzere bir metodoloji geliştirilmiş ve bu tekniğe dayalı Türkçe biçimbirimsel çözümleyici oluşturulmuştur. Sistem dört hastalık grubundan (kalp, damar, solunum ve kan rahatsızlıkları) toplam 69 hasta şikayeti üzerinde denenmiş ve % 94 doğruluk oranıyla sonuç vermiştir. Doğal dil işleme tekniklerinin medikal sahadaki uygulamalarına Türkçe bir temel hazırlamak, bu çalışma dahilinde gözetilen bir amaçtır.

Anahtar Sözcükler: Biçimbirimsel çözümleyici, Türkçe, anahtar kelime analizi, akıllı sistem, doğal dil işleme

## Abstract:

In this study, it is proposed to build up an intelligent diagnostic system from the Turkish medical narratives by keyword analysis technique. The basic idea of the work is to recognize the roots of words in the narrative by a Turkish morphological analyzer and to classify the disease of the patient by matching these roots with the keywords in the knowledge database. With this purpose a technique for the morphological analysis of agglutinative languages had been proposed and a Turkish morphological analyzer is programmed with this methodology. 69 Turkish medical narrative texts, which explain the complaints of patients, in four different classes of illnesses as respiratory, hematological, cardiological and vascular problems are fed into the application and the program correctly classifies 94 %. Creation of an infrastructure for the use of natural language processing techniques in the medical area on Turkish is another aim of this study.

Keywords: Morphological analyzer, Turkish, keyword analysis, expert system, natural language processing

\*TÜBİTAK- Ulusal Elektronik ve Kriptoloji Araştırma Enstitüsü, 41470, Gebze, Kocaeli,  
e-posta: kulekci@mam.gov.tr

Bilgisayar tarihinin başlangıcından beri, elektronik bir beyin yaratmak bilim adamlarının bir hayali olagelmıştır. Bu amaçla yapay zeka çalışmaları özellikle son yarım yüzyılda yoğun gelişme kaydetmiş ve medikal sahada da son otuz yıldan beri bu ilerlemeler geniş çaplı kullanım alanları bulmuştur. Tıbbi yapay zeka uygulamaları daha ziyade otomatik teşhis/tanı yapabilen akıllı sistemlerin geliştirilmesine odaklanmıştır [1]. *Dxplain*, *Menelas*, *Iliad*, *Jeremiah*, *Promnet*, *Orthaplanner*, ve *Rapid* rutin kullanımında olan bu tür karar destek sistemlerine örnek olarak verilebilir.

Yapay zekanın birçok teknikleri akıllı sistem tasarımlarında denenmektedir. Coiera herhangi bir kısıtlama olmaksızın elektronik ortama aktarılan tıbbi bilgilerin işlenmesinin önümüzdeki yıllarda gittikçe önem kazanacağını ve bu bağlamda doğal dil işleme teknolojisinin de bu amaç için kullanılabileceğini belirtmiştir[2].

Doğal dil işleme teknolojisi sadece akıllı sistemlerde değil, elektronik hasta kayıtları, hastane bilgi sistemleri, arşivlerin taranması gibi medikal sahanın formatı çok belirli olmayan alanlarında da kullanılabilir. Tıbbi konseptlerin reprezentasyonu ve dil işleme teknikleri karşılıklı olarak birbirlerine destek olabilirler [3]. Medikal kavramlar ve terimler doğal dil uygulamaları için iyi birer bilgi tabanı oluştururken, serbest datanın herhangi bir kısıtlamaya gerek olmadan dil işleme teknikleriyle incelenebilmesi de buna karşılık bir kazanım olabilir.

Bu çalışmada kalp, damar, solunum ve kan rahatsızlıkları olarak belirlenen dört grup hastalık test kümesi olarak alınmıştır. Bu hastalıklarla ilgili olarak hasta şikayetlerinin anahtar kelimeleri belirlenmiştir. Bu sözcüklerin tespiti için '*Semptomdan Teşhise*' isimli kitap kullanılmıştır [4]. Belirtilen hastalık gruplarından oldukları bilinen 69 adet hastanın şikayetleri derlenmiştir. Bu hasta bilgileri daha önce yapılmış olan bir yüksek lisans tezinden alınmıştır

[5]. Geliştirilen uygulama dahilinde test edilen sistem % 94 oranında doğru karar vermiştir. Bulguların değerlendirilmesi ve ileriye dönük çalışmalar Sonuç bölümünde değerlendirilmiştir.

## II. BİÇİMBİRİMSSEL ÇÖZÜMLEYİCİ

Hasta hikayesi verisi içerisindeki kelimelerden anahtar sözcükleri tespit edebilmek için Türkçe biçimbirimsel çözümleyici yapılmıştır. Bu çözümleyici kelimenin kökünü bulmakta kullanılmaktadır. Bir kelimenin özellikle eklemeli dillerde bir çok çekimli hali bulunduğundan, her bir anahtar kelimenin tüm çekimli hallerini veri ambarında saklamak olası değildir. Bunun yerine anahtar sözcüklerin kök halleri bilgi bankasında bulundurulmaktadır. Hasta hikayesi içerisindeki tüm kelimelerin köklerini bulmak ve bu bilgi bankasındaki diğer kelimelerle karşılaştırılmak daha iyi bir çözüm olarak tespit edilmiştir. Bu amaçla Türkçe biçimbirimsel çözümleyici uygulaması yapılmıştır. Çözümleyicinin detayları Külekci'nin yüksek lisans tezinde bulunabilir [6].

Bir dildeki kelimelerin morfolojik analizi için genellikle sonlu durum teknikleri kullanılmaktadır. Türkçe içinde bu tekniklerin uyarlamaları mevcuttur [7,8]. Bu tekniklerin temeli kelimenin kökünün tespiti için olası ekleri tek tek tespit ederek çıkarmaktır. Örneğin 'evdekilerden' kelimesinin analizi 'ev-de-ki-ler-den' şeklinde yapılmaktadır. Bu tekniğin yerine ekleri gruplar halinde tespit edebilmek için yeni bir metodoloji bu çalışma dahilinde geliştirilmiştir. Bu yeni teknikte yukarıdaki örnekteki kelime 'ev-dekilerden' şeklinde tek seferde ayırt edilebilmektedir.

Yukarıda temel fikri verilmeye çalışılan metodoloji Türkçe üzerine aşağıdaki adımlar takip edilerek uygulanmıştır:

### 1. Türkçe Ses Dizgesinin İncelenmesi:

Türkçe kelimelerdeki ses uyumları, eklerin kelimeye eklenirken olan değişimleri, aynı ekin farklı ses olayları incelenmiştir. Sert sessiz benzeşmeleri, ses düşmeleri, yardımcı sesler, büyük ve küçük sesli uyumları, yumuşamalar gibi olaylar sınıflandırılmıştır.

### 2. Türkçe Biçimbirimsel Düzenin İncelenmesi:

Türkçe kelimelerin yapısal özellikleri incelenerek eklerin hangi sıralarla, ne şekilde kaynaşabilirlikleri, yapım ve çekim eklerinin sözcüklere eklenmeleri belirlenmiş ve belli kurallar dahilinde sınırlandırılmaları yapılmıştır.

### 3. Eklerin Sınıflandırılması ve Ek Kümelerinin Oluşturulması:

İlk iki adımın ışığında Türkçe'deki tüm yapım ve çekim ekleri sınıflandırılmıştır. Bu sınıflandırmanın akabinde 'ek kümeleri' oluşturulmuştur. *Ek Kümesi* aynı ses ve morfolojik özellikleri taşıyan eklerin oluşturduğu birlik olarak tanımlanmıştır. Örneğin {im, in, imiz, iniz} ekleri aynı ses karakteristiğini taşıdıkları ve görev olarak aynı iyelik grubunda olduklarından bir *ek kümesi* teşkil ederler. Ek kümelerinin temel prensibi, içerisindeki bir eke kaynaştırılabilen bir öge (ek) diğerlerine de kaynaştırılabilmektedir. Ek kümelerinin birbirleriyle kaynaştırılması sonucunda *ek grupları* ortaya çıkar. Misal olarak {im, in, imiz, iniz} ile {de} kümelerinin birleştirilmesinden {imde, inde, imizde, inizde} *ek grupları* teşkil edilir. Bu gruplar kelimelerin analizinde kullanılacak ana yapı taşlarıdır.

### 4. Ek Grup Oluşturma Tablolarının Oluşturulması:

Oluşturulan ek kümelerinin hangilerinin ne şekilde birleştirileceğini tarif eden tablolar meydana getirilmiştir. Bu tablolar yardımıyla otomatik olarak tüm ek gruplarının oluşturulması amaçlanmaktadır.

### 5. Tüm Olası Ek Gruplarının Oluşturulması :

Dördüncü adımda oluşturulan tablolar vasıtasıyla Türkçe'deki olası tüm ek grupları bir program aracılığıyla üretilerek veri ambarlarında saklanmıştır.

### 6. Kelime Biçimbirimsel Analiz Algoritmasının Yapılması:

Sözcüklerin morfolojik analizi için evvela sağ taraftan kelime kök adaylarının tespiti ve sol taraftan da olası ek gruplarının araştırılması gerekmektedir. Bu işlem için gerekli algoritma tasarlanarak programlanmıştır.

Yukarıda ana hatlarıyla verilen adımların sonucunda Türkçe biçimbirimsel çözümleyici kullanıma hazır hale getirilmiştir.

## III. SİSTEM MİMARİSİ

Uygulamanın temelinde tıbbi anahtar kelime analizi için geliştirilen biçimbirimsel çözümleyici bulunmaktadır. Kelimelerin morfolojik analizi için

gerekli olan Türkçe Sözlük Türk Dil Kurumu web sayfasından alınmıştır [9].

Belirtilen hastalıkların (kan, kalp, solunum, damar) semptomlarına işaret eden anahtar kelimeler bir veri ambarına yüklenerek hastalıklarla ilişkilendirilmiştir. Sistemin mimarisi genel olarak Şekil 1 de gösterildiği gibidir..

Sistem hasta şikayetlerini içeren medikal hikayenin program dahilinde yazılması veya .txt uzantılı bir dosyadan yüklenmesi ile başlar. Metin dahilindeki tüm kelimelerin kökleri biçimbirimsel çözümleyici ile bulunur. Bulunan kökler bilgi bankasındaki anahtar kelimeler arasında aranır. Eğer bir kök anahtar sözcüklerden biriyle eşlenirse o anahtar kelimenin ilişkili olduğu hastalıkların puanları bir artırılır. Böylece tüm arama işlemi tamamlandığında her hastalıktan kaç adet anahtar kelime bulunduğu belirlenmiş olur. Herhangi bir hastalığın metinle alakalı olma oranı o hastalığın puanının tüm hastalıkların toplam puanına bölümüyle elde edilir. Bu şekilde hesaplanan dört hastalığın ihtimal yüzdeleri büyükten küçüğe doğru sıralanarak raporlanır.

#### IV. SONUÇLAR

Uygulamanın test edildiği 69 hasta hikayesi üzerinde alınan sonuçlar Tablo I. de gösterilmiştir.

Tablo I. de görüldüğü üzere genel hata oranı % 6 olarak ortaya çıkmıştır. Bazı hasta hikayelerinde en yüksek ihtimal olarak doğru teşhisin yanı sıra bir diğer hastalık da sistem tarafından eşit derecede olası görülmüştür. Bu tür sonuçlar muğlak (ambiguous) olarak değerlendirilmiş ve tabloda da bu şekilde gösterilmiştir. Bu tür muğlak cevapların oranı hatalara göre çok daha fazladır. Bunun en önemli sebebi bir anahtar kelimenin bir çok hastalığa işaret etmesi ve hasta hikayelerinde yeterince bilgi bulunamaması durumlarında hastalıkların ayırt edilememesi olarak görülmektedir. Bu durumun çözümü ve sistemin daha performanslı olarak kullanılabilmesi için programın kullanıcıyla etkileşimli (interactive) hale getirilmesi önerilebilir. Herhangi bir şüphe veya bilgi yetersizliğinde, uygulama kullanıcıya muğlaklığı giderici sorular yönelterek daha kesin sonuçlara ulaşabilir.

Bir diğer dikkate alınması gereken husus ise her anahtar kelimenin çeşitli hastalıklarda farklı ağırlığının olması ve programın bu şekilde programlanmasıdır. Mesela 'göğüs ağrısı' hem kalp hem de solunum hastalıklarında görülmesine rağmen her hastalıkta tespit edilen diğer öğelere de bağımlı olarak farklı oranlarda etki etmelidir. Bu düşünceler doğrultusunda sistemin 'yapay sinir ağları' kullanılarak tekrar programlanmasının daha performanslı ve güvenilir olacağı

sonucu ortaya çıkmaktadır. Literatürde de bu tür yapay sinir ağı uygulamaları mevcuttur [10].

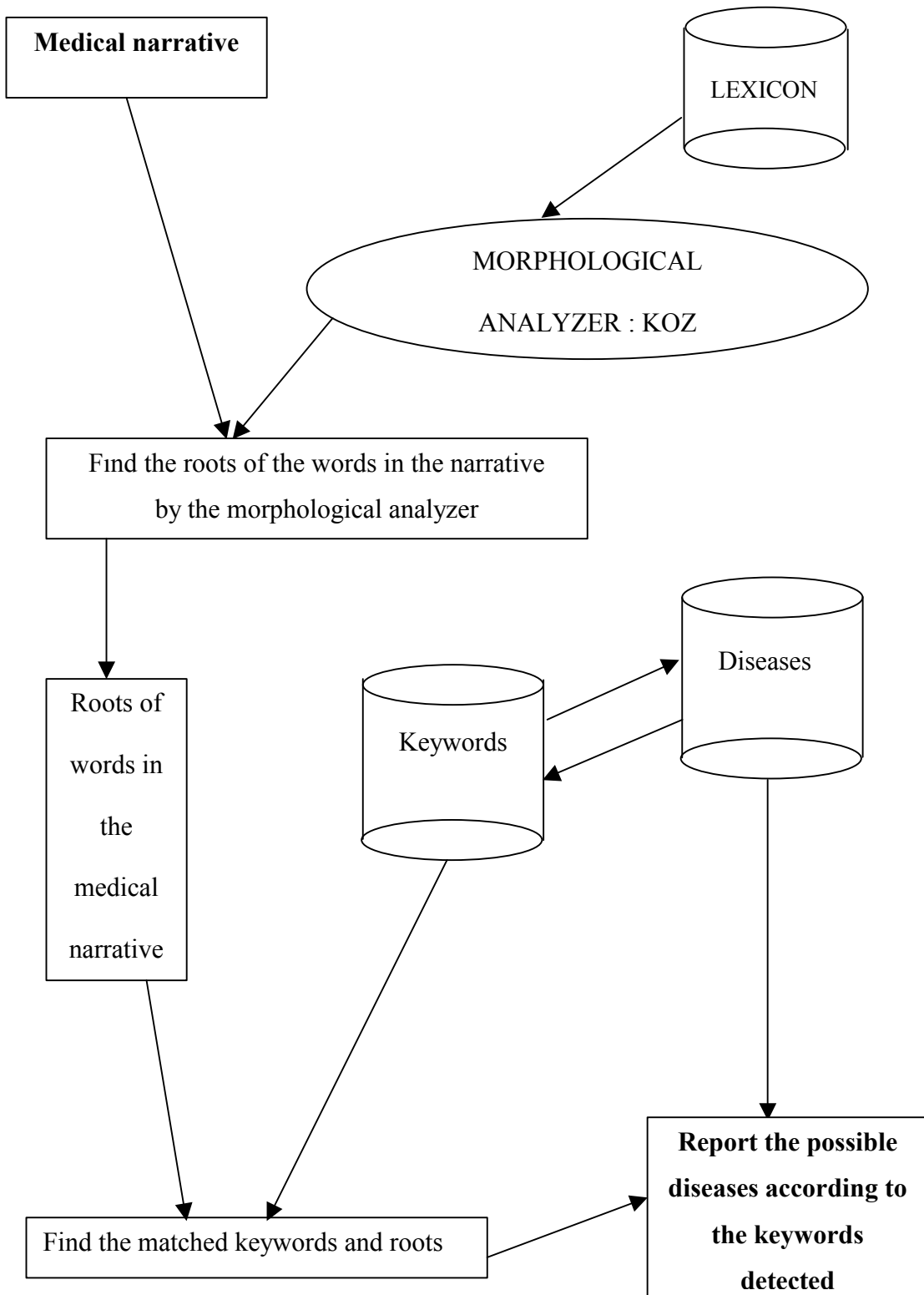
Bu çalışma ile birlikte medikal sahada doğal dil işleme teknolojisinin kullanımı için bir alt yapı oluşturma amacı da güdülmüştür. Yalnızca akıllı sistemlerde değil, bilginin belli düzen dahilinde değil de doğal dilde bulunduğu durumlarda da geliştirilen biçimbirimsel çözücülerden faydalanılabilir.

#### IV. KAYNAKLAR

1. Coiera, E., "Artificial Intelligence in Medicine", in E. Coiera (Eds.), *Guide to Medical Informatics, The Internet and Telemedicine*, Chapter 19, Amsterdam, 1997
2. Coiera, E., "The Role of Knowledge Based Systems in Clinical Practice", in P.Barahona, J.P. Christen (Eds.), *Knowledge and Decision in Health Telematics-The next Decade*, pp. 199-203, Amsterdam, 1994
3. Wagner, J.C., C. Lovis, R.H. Baud, J.R. Scherrer, "Medical Concept Systems, Lexicons and Natural Language Generation", *Lecture Notes in Artificial Intelligence*, Vol. 1211, pp 398-401, 1997
4. Pınar, T., *Semptoma Yönelik Tanı El Kitabı*, Hacettepe Doktorlar Yayınevi, Ankara, 1994
5. Burşuk, E., "A diagnostic expert system for cardiological, respiratory, vascular and haematological diseases", MS Thesis, Boğaziçi University, 1999
6. Külekci, M. O., "An intelligent diagnostic system from the clinical narratives in Turkish", MS Thesis, Boğaziçi University, 2000
7. Güngör, T., "Computer Processing of Turkish Morphological and Lexical Investigation", Ph.D. Dissertation, Boğaziçi University, 1993
8. Oflazer, K., "Two-level Description of Turkish Morphology", *Proceedings of the Second Turkish Symposium on Artificial Intelligence and Neural Networks*, pp. 86-93, İstanbul, 1993.
9. www.tdk.gov.tr, Türk Dil Kurumu
10. Bassoe, C.F., "Automated Diagnoses From Clinical Narratives: A Medical System Based on Computerised Medical Records, Natural Language Processing, and Neural Network Technology", *Neural Networks*, Vol. 8, No. 2, pp. 313-319, 1995

**Tablo I.**  
Sistem sonuç tablosu

	<b>Kalp Hastaları</b>	<b>Damar Hastaları</b>	<b>Kan Hastaları</b>	<b>Solunum Hastaları</b>	<b>TOPLAM</b>
<b>Hasta Sayısı</b>	19	12	18	20	69
<b>Yanlış Teşhis Edilen Hasta Sayısı</b>	0	0	3	1	4
<b>Muğlak (ambiguous) sonuç sayısı</b>	4	0	4	2	10
<b>Hata Yüzdesi</b>	0 %	0 %	16 %	5 %	6 %
<b>Muğlaklık Yüzdesi</b>	21 %	0 %	22 %	10 %	14 %



Şekil 1. Sistem mimarisi