

# Tıpta Metin Madenciliği

Başak OĞUZ<sup>a</sup>, Uğur BİLGE<sup>a</sup>, Osman SAKA<sup>a</sup>

<sup>a</sup> Biyoistatistik ve Tıp Bilişimi AD, Akdeniz Üniversitesi, Antalya

## Text Mining in Medicine

### Abstract

*Our ability to access and store information has grown tremendously since the internet became popular in the mid-90s. We can rapidly download and store large quantities of information on a personal computer (PC), but this huge volume of information on both PC and the internet brought the information retrieval problem. To solve this problem, new methods have been put forward. Text mining saw an increasing interest in the 2000s, as a tool for enabling people to find unknown information and facts from the free-text data. This study is a review, that includes text mining and its common applications in the medical sciences.*

### Key Words

Medicinal science; Text Mining; Information Retrieval

### Özet

*1990'lı yılların ortasından itibaren internetin popüler hale gelmesiyle birlikte bilgiye erişim ve bilgiyi depolama gücü çok büyük oranda artmıştır. Kişisel bilgisayarlara hızlı bir şekilde büyük oranda bilgiler indirilebilmiş ve depolanabilmiştir. Fakat hem kişisel bilgisayarlarda hem de internet üzerinde yaşanan bu bilgi patlamasıyla birlikte istenilen bilgiye erişim problemi ortaya çıkmıştır. Bu sebepten bilgiye erişim probleminin çözülebilmesi için yeni yöntemler geliştirilmeye başlanmıştır. Metin madenciliği, özellikle 2000 yılından sonra daha fazla ilgi gören, serbest formatta bulunan metinlerin içindeki daha önceden bilmediğimiz bilgileri ortaya çıkarmamızı sağlayan işlemler bütünüdür. Bu çalışma metin madenciliğinin ne olduğuna ve tıptaki uygulama alanlarına dair yapılan literatür gözden geçirme çalışmasıdır.*

### Anahtar Kelimeler

Tıp; Metin Madenciliği; Bilgi Erişim

## 1. Giriş

Bilindiği gibi bilgisayarlar, her geçen gün hayatımıza daha çok girmekte ve elimizdeki verileri saklayabileceğimiz, istediğimizde bu verilere ulaşabileceğimiz, arama yapabileceğimiz birer araç olmaktadır. Bununla birlikte, son yıllarda veritabanlarında saklanan verinin hacmi oldukça büyümüş ve buna bağlı olarak verinin çeşitli yöntemlerle organize edilme ihtiyacı doğmuştur.

Dünyadaki verilerin yaklaşık olarak % 90'ı yapılandırılmamış formatta bulunmaktadır [1]. Yapılandırılmamış veri, bilgisayar tarafından kolayca anlaşılmayan ve veri yapısına sahip olmayan müzik, video dosyaları ve serbest formatta bulunan metinlerdir (e-mail, web sayfaları vb.) [2]. Metin madenciliği, belirli bir formatta olmayan, yazı tipindeki veriler içerisinde gizli olan nitelikli bilginin çıkarılması, düzensiz haldeki verinin formatlanması sürecidir [3]. Başka bir deyişle, geniş metin veritabanlarından desenleri kullanarak kullanışlı ve daha önceden bilinmeyen bilgilerin çıkartılması işlemidir [1]. Metin Madenciliği, Metin Veri Madenciliği (Text Data Mining) ve Metin Veritabanlarından Bilgi Keşfi (Knowledge Discovery from Textual Databases ) olarak da adlandırılır [4]. Metin madenciliği yeni bir terim olmasına rağmen, bilgi erişim sistemleri ve NLP ile ilgili yapılan araştırmalara bağlı olarak ortaya çıkmıştır. Bilgi erişim ile ilgili çalışmalar 1960'lı yıllarda başlamış, doğal dili anlamaya ve sayısallaştırmaya yönelik programlar geliştirilmeye çalışılmıştır. 1990'lı yıllara gelindiğinde ise metinlere erişim, metinlerden bilgi çıkarımı, metin kategorizasyonu, metinleri yapısal hale getirmeye yönelik çalışmalar hızla artmaya başlamıştır. Tıptaki verilerin çoğu karmaşık, yapılandırılmamış

formatta, kağıt tabanlı veya elektronik ortamda serbest metin olarak bulunmaktadır. Hekimler karar verme sürecinde veya araştırma yaparken hasta raporları, klinik çalışmalar, araştırma raporları, web sayfaları ve hastane kayıtları gibi serbest metin formatında veya kağıt tabanlı metinleri kullanmaktadır. Fakat yapılandırılmamış formatta bulunan bu veri yığınlarını insan gücüyle analiz etmek ve istenilen bilgiye ulaşmak hem zordur hem de zaman kaybına yol açmaktadır. Hastayla ilgili karar verme süresinin, doğru verilere erişmenin öneminin ve bu verileri kullanarak istenilen bilgilere ulaşmanın zorluğu göz önünde bulundurulduğunda bu tür sistemlere olan ihtiyaç daha belirgin olarak ortaya çıkmaktadır.

Bu çalışmada, ilk olarak metin madenciliği kavramı ve izlenen süreçler ele alınacak, daha sonra da sağlık alanında bu konuyla ilgili yapılan çalışmalardan bahsedilecektir.

## 2. Metin Madenciliği Nedir?

Özellikle İnternet ve kişisel bilgisayarların yaygınlaşmasına bağlı olarak, gittikçe büyüyen hacme sahip doküman yığınları oluşmaktadır. Bu yığınlar içinde önemli bilgiler kaybolup giderken, değerli bilgilere ulaşmak için dokümanların içeriğinin belirlenmesi ve buna uygun sorgulanabilmesi ihtiyacı kendini hissettirmektedir. Metin madenciliği kavramı günümüzde tam olarak tanımlanamamış olsa da son on yılda büyük gelişmeler sağlanmış bir alandır. Bu konuda bir çok tanımlama geliştirilmiş, fakat herkes tarafından benimsenmiş bir tanımlama yapılamamıştır. Yaygın olarak kullanılan tanım, belirli bir formatta olmayan, yazı tipindeki veriler içerisinde gizli olan nitelikli bilginin çıkarılması, düzensiz haldeki verinin formatlanması süreci olduğuna yöneliktir. Kostoff ve arkadaşları bilim ve teknoloji metin madenciliğini “ bilginin teknik literatürden çıkartılması“ olarak tanımlamış ve bilgi erişim, bilgi işleme ve bilgi entegrasyonu olmak üzere üç bileşenden oluştuğunu belirtmişlerdir. Bilgi işlemeyi, erişilen belgelerdeki örüntülerin çıkartılması işlemi, bilgi entegrasyonunu ise erişilen ilgili belgelerin okunarak bilgi işleme aşamasından sonra çıkan sonuçlarla kombinasyonun sağlanması süreci olarak tanımlamışlardır [5]. Losiewicz ve arkadaşları metin veri madenciliğini, metin koleksiyonlarından bilgiye erişmeyi, bireysel metinlerden bilgi çıkarmayı, veritabanlarından bilgi keşfini, organizasyonlarda bilgi yönetimini ve verinin ve bilginin görselleştirilmesi aşamalarını birleştiren bir mimari olarak tanımlamışlardır [5].

Metin Madenciliği, metin içindeki kalıpları tanımlayıp bilinmeyen bilgiyi ortaya çıkararak, var olan yapısı ile metinleri bilgiye dönüştüren anahtar bir süreçtir. Bu süreçte metin ilk başta bir ön işlemden (cümle ve kelime analizleri) geçer, anlamsız kelimeler çıkarılır, metinler kategorilendirilir ve sonuçta geleneksel veri madenciliği yöntemleri kullanılarak (kümeleme, yapay sinir ağları, karar ağaçları, regresyon analizleri vb.) geniş hacimli metinler analiz edilir ve daha sonra elde edilen sonuçlar değerlendirilir [6]. Farklı dillerde binlerce doküman, web sayfa içerikleri, yayınlar ve özetler göz önüne alındığında erişilmek istenen bilgilere ulaşmanın güçlüğü bilinmektedir. Araştırmacılar düzenli haldeki verileri analiz ettikleri gibi (yaş, cinsiyet, kilo, kolesterol, nabız, tansiyon vb); tıbbi raporlardan, internet sayfalarından, makalelerden, fatura bilgilerinden buldukları metin verileri de analiz edebilmektedirler [6]. Bu metinlerin kısa sürede analiz edilmesi ve nitelikli bilgilere çok kısa sürede erişilmesi için metin madenciliği yöntemi kullanılmaktadır.

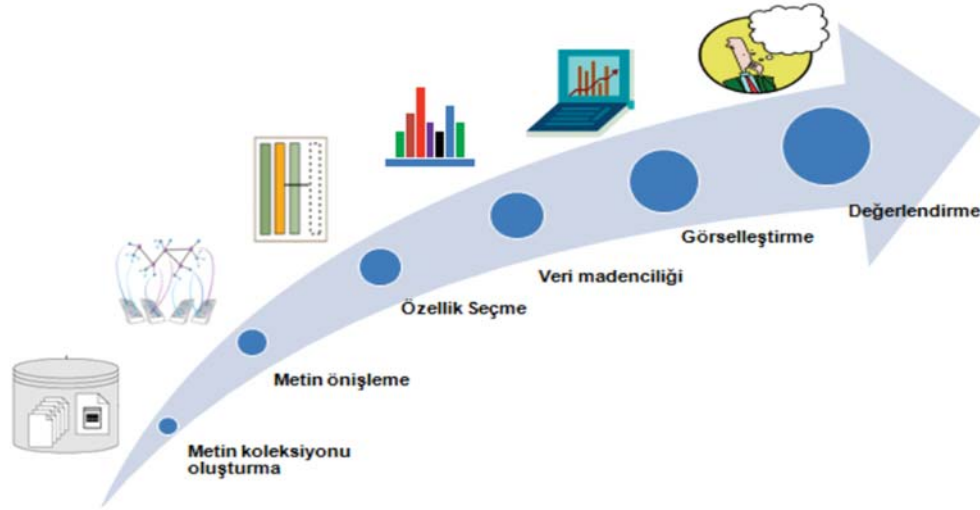
Metin madenciliği genel olarak beş adımdan oluşmaktadır. Bu adımlar;

1. Metin toplama: İlgilenilen konularda bilgi erişim sistemleri kullanılarak metin koleksiyonu oluşturma sürecidir. Bu süreç, günümüzde genel olarak internet üzerinden, özellikle Google arama motoru kullanılarak gerçekleştirilmektedir. Tıp alanında ise metin toplama süreci yaygın olarak PubMed çevrim içi veritabanı üzerinden yapılmaktadır [7].

2. Metin ön işleme: Metni kelimelere ayırma, kelimelerin anlamsal değerlerini bulma (isim, sıfat, fiil, zarf, zamir vb.), kelimeleri köklerine ayırma ve gereksiz kelimeleri ayıklama, dokümanı fazlalıklardan arındırmak, gereksiz bilgileri çıkarmak, yazım kurallarına uygunluğunu tespit etmek, ekleri ve genel

kelimeleri çıkarmak, kelime köklerini belirlemek, önek ve sonekleri ayırmak, kelime anlamlarını belirlemek ( Kelimenin hangi anlamı kullanılıyor?), ilişki analizi (A'nın B ile bir ilişkisi varsa, B'nin de C ile bir ilişkisi varsa, A ile C arasında da potansiyel bir ilişki vardır) yapmak gibi metin belgelerin yapıtaşları olan kelimelerle ilgili işlemleri içeren süreçtir [7].

3. Özellik seçme: Bu aşamada ön işlemde geçen metinlerdeki önemli kelimeleri (varlıkları) belirleme (isimler, tamlamalar, bileşik kelimeler, kısaltmalar, sayılar, tarihler, para birimleri vb.) ve ilişkili olmayan özelliklerin çıkarılması (sadece birkaç dokümanda gözlemlenen özelliklerin çıkarılması, birçok dokümanda gözlemlenen özellikleri azaltma) işlemleri yapılmaktadır.



Şekil-1. Metin Madenciliği Süreci

4. Veri madenciliği: Yapılandırılmış format haline getirilen metinlerin geleneksel veri madenciliği teknikleriyle (karar ağaçları, yapay sinir ağları, kümeleme vb.) analizi sürecidir. Hem veri madenciliğinde hem de metin madenciliğinde gizli bilgilere bakılmakta ve genel yapay zeka, makine öğrenme ve istatistik algoritmaları kullanılmaktadır. Veri madenciliğinde yapılandırılmış numerik veri kullanılırken metin madenciliği yapılandırılmamış metinlerle ilgilidir. Veri madenciliğinde kullanılan veriler veri ambarlarında çıkartılmış, dönüştürülmüş ve yüklenmiş durumda bulunan verileri kullanırken metin madenciliği kesin olmayan verileri modellemeye çalışmaktadır [8].

5. Görselleştirme : Elde edilen sonuçların kullanıcıya sunumunda en etkin ve anlaşılır görselleştirmenin yapılması aşamasıdır.

6. Değerlendirme : Genel olarak sistemlerin değerlendirilmesinde duyarlılık (precision), anma (recall) ya da ikisinin birleşiminden oluşan F-score ölçütü kullanılmaktadır [7]. Duyarlılık, erişim çıktısındaki ilgili belge sayısının erişim çıktısındaki belge sayısına oranıdır. Anma ise, erişim çıktısındaki ilgili belge sayısının belgeler kümesinde ilgili belgeler sayısına oranıdır.

Metin madenciliğinin teknikleri genel olarak bilgi çıkarma, bilgi erişim, doğal dil işleme alanlarında kullanılmaktadır [9]. Bilgi Çıkarma, metin verilerden yapılandırılmış bilginin otomatik olarak çıkartılmasıdır [2]. Bilgi erişim, kullanıcıların anahtar kelimeleri kullanarak bilgi ihtiyaçlarını karşılaması muhtemel derlemedeki ilgili belgelerin tümüne erişmek, ilgili olmayanları da ayıklamaktır [10]. Diğer önemli uygulama alanlarından biri olan Doğal Dil İşleme teknikleri ise, yapılandırılmamış metinsel bilgilerin analizi ve otomatik olarak işlenmesi ile ilgilidir [11]. Konuşma tanıma, yazılı dokümanların otomatik çevrilmesi, otomatik konuşma ve komut anlama, konuşma sentezi, konuşma üretme, otomatik metin özetleme gibi önemli alanlarda kullanılmakta ve doğal dilin makine tarafından anlaşılmasını sağlayan teknikleri içermektedir [2].

Sonuç olarak metin madenciliğinde amaç yapılandırılmamış ya da yarı yapılandırılmış metinlerin formatlanıp yapılandırılmış hale dönüştürülmesi, analiz edilmesi ve nitelikli bilgilere kısa sürede erişilmesini sağlamaktır.

### 3. Tıpta Metin Madenciliği

Metin madenciliği tekniklerinin tıpta kullanımı son birkaç yılda büyük oranda artmıştır. Tıptaki verilerin genel olarak serbest metin formatında bulunması hasta ile ilgili önemli bilgilerin gözden kaçmasına, bilgiye erişimin zorlaşmasına sebep olmaktadır. Özellikle Elektronik Sağlık Kayıtlarının, Sağlık Bilgi Yönetiminin son yıllarda en önemli hedeflerinden birisiyken, böyle bir sistemin başarısının, klinik dokümantasyonun serbest metin formatında yapılmasından dolayı sınırlanmış durumda olması bu tür sistemlere olan ihtiyacı ortaya çıkarmıştır. Yapılan klinik çalışmalar, araştırma raporları, hastane kayıtları, doktor notları, prosedürler ve faturalar tıptaki en önemli veri kaynaklarıdır. Bu verilerin çoğu serbest metin formatında bulunmaktadır [8].

Metin madenciliği, tıp alanında özellikle tıbbi araştırmalarda, semptomlarla hastalıklar ve ilaçlarla kimyasal maddeler arasında nedensel bağları bulmada, hasta kayıtlarının analiz edilmesinde, gen-gen ve protein-protein ilişkilerinin tanımlanmasında, tanı ve tedavileri geliştirmek, servis kalitesini ve faydayı arttırmak, maliyetleri kontrol etmek için kullanılmaktadır [1]. Shatgay'ın çalışmasında, DNA mikroarray deneylerinde genler arasındaki fonksiyonel ilişkilerin keşfedilmesi için biyomedikal yayınlar taranmış ve erişilen makalelerin özetlerinin içeriğine dayanarak ilişkiler bulunmaya çalışılmıştır [12]. Swanson'un çalışmasında ise, metin madenciliği teknikleri hastalıklar ve semptomları arasındaki ilişkilerin ve bağıntıların bulunması için kullanılmıştır. Tıbbi araştırma sayfaları, makaleler, haberler kullanılarak semptomlar, ilaçlar, hastalıklar, kimyasallar arasında ilişki örüntülerine bakılmıştır [1]. Medikal başlık ve özetler incelenerek belirlenen bir problemin (az görülen bir hastalık) nedenleri karşılaştırılmıştır. Başlıklar arasında nedensel ilişkinin bulunması için ARROWSMITH adlı bir yazılım geliştirilmiştir [13]. Swanson'ın sisteminde MEDLINE'da iki başlık kullanılarak arama yaptırılır (magnezyum ve migren) ve sonuçlar (başlıklar ve özetler), yaygın olarak bulunan önemli kelime ve tamlamaları liste şekline getiren ARROWSMITH programına atılır. Migrene bağlı magnezyum eksikliği problem olarak belirlendikten sonra beslenmeyle ve migrenle ilgili literatür taraması yapılmıştır. Sonuç olarak tüm elde edilen veriler incelendikten ve modellendikten sonra magnezyum eksikliğinin migren ağrılarına neden olabileceği bulunmuştur [14]. Lindsay ve arkadaşları Swanson'ın yaklaşımını, metin madenciliği olarak adlandırmadan genişletmişlerdir. Lindsey ve Gordon bu yaklaşıma genel kelimeler ve tamlamaları bulmak için kelime sıklığı istatistiklerini de eklemişlerdir. Fakat Swanson'ın yaklaşımındaki gibi hala birkaç noktada "insan filtreleme" ihtiyaç duyulmaktadır [5].

Johnson ve arkadaşları, serbest formatta bulunan radyoloji raporlarında bulunan bilgileri yapılandıran ve çıkartan RADA adlı bir sistem tasarlamışlardır. Sistem, serbest formatta bulunan metinleri yapılandırmış hale dönüştürürken doğal dil işleme tekniklerini kullanmaktadır. Sistemin değerlendirilmesi için göğüs onkolojisi bölümünden 100 adet radyoloji raporu rastgele seçilmiştir. Raporlardaki bir cümlenin içerdiği ortalama kelime sayısı 14.97 ve sistemin bir raporu analiz etme süresi yaklaşık olarak 1-2 dakika olarak bulunmuştur. Ayrıca sistemden elde edilen sonuçların değerlendirilmesinde altın standart olarak göğüs radyolojisi uzmanından yardım alınmıştır. Sonuçta sistemin anma oranı % 85, duyarlılığı % 89 bulunmuştur [15]. Schadow ve arkadaşları serbest formatta bulunan cerrahi patoloji raporlarından numuneler ve numunelerle ilgili bulgular hakkındaki bilgilere erişmeyi sağlayan bir metod geliştirmişlerdir. Patoloji laboratuvarından elde edilen 622 adet rapor otomatik olarak eleme yapan bir sistem tarafından taranmış ve içinde "tanı" kelimesi geçmeyen raporlar ayıklanmıştır. Sonuç olarak geriye kalan 275 rapor XML formatına dönüştürülmüş ve doku tipi, yeri, toplama methodu ile ilgili bilgiler elde edilmiştir [16].

Biyomedikal alanda metin madenciliği ile ilgili özellikle isimlendirilmiş varlıkları tanıma (named-entity recognition, NER) alanında çalışmalar yapılmıştır. İsimlendirilmiş varlıkları tanıma, metin koleksiyonlarında bulunan tüm isim varlıklarının (ilaç isimleri, hastalık isimleri, simgeler) tanımlanması işlemidir. NER, metin içindeki varlıkların tanımlanması varlıklar arasındaki ilişkilerin bulunmasına, anahtar kavramların tanımlanmasına ve bu kavramların uygun bir şekilde sunulmasına olanak sağlamaktadır [17]. Biyomedikal alanda NER ile ilgili yapılan çalışmalarda serbest metinlerdeki gen



ve protein isimlerini tanımaya odaklanılmıştır [17]. Biyolojik isim varlıkların tüm tiplerinin tam olarak belirtildiği bir sözlük bulunmaması, biyolojik varlıkların çok kelimele olabilmesi, aynı varlığın birden fazla isim alabilmesi vb. yaşanan problemler arasındadır. Bu alanda geliştirilen sistemlerin değerlendirilmesi doğruluk ve duyarlılık ölçümleri ile yapılmaktadır. Bu iki ölçümün birleştirilmesi ile oluşturulan ölçüm ise F-score (harmonik ortalama) ölçümü ise yaygın olarak tercih edilmektedir [17]. Tanabe ve arkadaşları tarafından geliştirilen AbGene sistemi, biyomedikal metinlerdeki gen ve protein isim varlıklarının tanımlanması için oluşturulan en başarılı kural tabanlı yaklaşımlardan birine sahiptir. Sistemde, gen isimleri ile gen olmayan isimlerin ayrımının yapılması için gereken kuralları oluşturmada tümevarımsal mantıksal programlama kullanılmıştır. Sistemin duyarlılık % 85.7, anma oranı % 66.7'dir [17]. Hanisch çalışmasında gen ve protein isimlerinin yer aldığı ve kelimelerin anlamsal olarak sınıflandırıldığı geniş bir sözlük kullanılmıştır. Sistemin doğruluğu % 0.95, duyarlılığı % 0.90'dır [17].

#### 4. Sonuç

Tıpta veriye dayalı bilimsel çalışmalar giderek artmaktadır. Verilerin analizi ile elde edilen bilgi, özellikle tıpta karar verme sürecinde önemli bir role sahiptir. Bilimsel bilginin büyük bir çoğunluğu yayınlarda ve diğer yapılandırılmamış formatlarda kayıt altında tutulmaktadır. Dünyadaki verilerin yaklaşık olarak % 90'ının yapılandırılmamış formatta saklandığı göz önünde bulundurulduğunda bilgiye erişmenin ve bilgiyi en iyi şekilde kullanmanın zorluğu göz ardı edilmemelidir. Özellikle bilginin öneminin artması ve bilgiye erişimde yaşanan problemler bu alanda yeni teknolojilerin geliştirilmesine yol gösterici olmuştur. Metin madenciliği, araştırmacılara literatür gözden geçirmede, işletmelerde serbest formatta bulunan bilgilere erişimde yeni çözümler sunmakta ve harcanılan süreyi büyük oranda azaltmaktadır. Tıptaki verilerin çoğu yarı yapılandırılmış veya yapılandırılmamış formatta bulunduğu için dolayı, hem klinik araştırmalarda hem de karar verme sürecinde hekimlere yardımcı olacak, metin yığınları içerisinde istenilen bilgiye erişmek için harcanan süreyi azaltacak ve hasta bakım kalitesini arttıracak bu tür sistemlerin sağlık alanındaki önemi büyüktür. Özellikle ülkemizde, metin madenciliği alanına henüz yeterli derecede ilgi gösterilmemesi, bu alana daha çok önem verilmesini ve üzerinde daha yoğun çalışmayı gerektirmektedir.

#### 5. Kaynakça

- [1] Bhatt C, Mining the Medical Literature, [http://ai.stanford.edu/~serafim/CS374\\_2004/Lecture%20Notes/lecture6.pdf](http://ai.stanford.edu/~serafim/CS374_2004/Lecture%20Notes/lecture6.pdf) , 11.12.2006.
- [2] Unstructured data, [http://en.wikipedia.org/wiki/Unstructured\\_data](http://en.wikipedia.org/wiki/Unstructured_data) 30.09.2007
- [3] Hwa R, An Overview of Text Mining, <http://www.umiacs.umd.edu/~hwa/textmining.ppt>, 11.12.2006.
- [4] Sehgal AK, Text Mining: The Search for Novelty in Text, <http://www.cs.uiowa.edu/~sehgal/Papers/comp04.pdf> , 12.02.2007.
- [5] Sharp M, Text mining , [http://www.scils.rutgers.edu/~msharp/text\\_mining.htm](http://www.scils.rutgers.edu/~msharp/text_mining.htm), 24.09.2007.
- [6] Cerrito , Inside text mining: text mining provides a powerful diagnosis of hospital quality rankings - Data Warehousing/Mining, [http://findarticles.com/p/articles/mi\\_m0DUD/is\\_3\\_25/ai\\_114167705/pg\\_2](http://findarticles.com/p/articles/mi_m0DUD/is_3_25/ai_114167705/pg_2) , 20.12.2006.
- [7] Mathiak B, Eckstein S, Five Steps to Text Mining in Biomedical Literature, [http://www2.informatik.hu-berlin.de/Forschung\\_Lehre/wm/ws04/7.pdf](http://www2.informatik.hu-berlin.de/Forschung_Lehre/wm/ws04/7.pdf), 17.01.2007.
- [8] Konchady M, *Text Mining Application Programming*. 1st ed. Charles River Media, 2006.
- [9] Karanikas H, Text Mining Software Survey, [www.crim.co.umist.ac.uk/tm/presentations/Karanikas.ppt](http://www.crim.co.umist.ac.uk/tm/presentations/Karanikas.ppt), 11.12.2006.
- [10] Tonta , Bitirim Y, Sever H , Türkçe Arama Motorlarında Performans Değerlendirme, <http://eprints.rclis.org/archive/00009697/01/tonta-bitirim-sever-arama-motorlari.pdf>, 11.06.2007.
- [11] Popowich F, Natural language processing and text mining; Using Text Mining and Natural Language Processing for Health Care Claims Processing, *ACM SIGKDD Explorations Newsletter* 2005: 7: 59 – 66.
- [12] Shatkay H, Edwards S, W. Wilbur J, and Boguski M, Genes, themes and microarrays, using information retrieval for large-scale gene analysis, *In Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology* 2000, 317-328.
- [13] Bekhuis T, Conceptual biology, hypothesis discovery, and text mining: Swanson's legacy, *Biomedical Digital Libraries* 2006, 3:2.
- [14] Gürsakal N, Sözcük ve Sayı, [www.20.uludag.edu.tr/~gursakal/down/say.ppt](http://www.20.uludag.edu.tr/~gursakal/down/say.ppt), 05.11.2006.
- [15] Johnson DB, Taira RK, Cardenas AF, and Aberle DR, Extracting Information from Free Text Radiology Reports, *Int J Digit Libr* 997: 1:297-308.
- [16] Schadow G, McDonald CJ, Extracting Structured Information from Free Text Pathology Reports, *AMIA Annu Symp Proc*. 2003: 584–8.
- [17] Cohen AM, Hersh WR, A Survey of Current Work in Biomedical Text Mining, *Briefings in Bioinformatics* 2005: 6: 1: 57–71.

#### 6. Sorumlu Yazarın Adresi

Başak Oğuz, Biyoistatistik ve Tıp Bilişimi AD, Akdeniz Üniversitesi Tıp Fakültesi, Antalya  
E-posta: basakoguz@akdeniz.edu.tr