

# Bilgi Çıkarımı Information Extraction

Mehmet Fatih AMASYALI

BLM 5212 Doğal Dil İşlemeye Giriş Ders Notları

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Akış

- Bilgi Çıkarımı Nedir?
- Mesaj Anlama Konferansları
- Uygulama Alanları
- Yapılandırılmış, yarı yapılandırılmış dokümanlar
- Basit Çıkarım Şablonları
- NLP'nin Bilgi Çıkarımına Katkısı
  - Öğelerine ayrılmış metinler
  - Özel anlamlı kelimeleri belirleme (Name Entity Recognition)
- Kaynak Seçimi
- Dinamik Web Sayfalarından Bilgi Çıkarımı
  - Alışveriş Robotları (froogle)
- IE performansının ölçümü
- Bilgi Çıkarımında Makine Öğrenmesi
  - Şablonlar metodu için bir deneme

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Bilgi Çıkarımı Nedir?

- Yapılandırılmamış yada yarı yapılandırılmış dokümanlardan önceden tanımlanmış şablonlara uygun bilgileri bulma
- Yapılandırılmamış yada yarı yapılandırılmış dokümanların yapılandırılmış veri tabanlarına dönüştürülmesi

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Mesaj Anlama Konferansları Message Understanding Conference (MUC)

- Amerikan savunma bakanlığı 1990'lerden itibaren bilgi çıkarımı konusuna eğilmiştir.
- MUC her sene yapılan bir bilgi çıkarımı yarışması.
- Haber makalelerinden
  - Terör olayları
  - Şirketler dünyasındaki birleşmeler, yönetim değişikliklerikonularında bilgi çıkarımı

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Uygulama Alanları

- İş ve işçi bulma
- Ürün bulma
- Seminer duyuruları
- Şirket bilgileri
- Üniversite başvuru bilgileri
- Kiralık / satılık daire, araba bilgileri
- *Ortak özellik: ?*

*birden fazla bilgi kaynağının araştırılması gereken durumlar*

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Yarı Yapılandırılmış Doküman Örnek İş İlanı

Subject: **US-TN-SOFTWARE PROGRAMMER**  
Date: **17 Nov 1996** 17:37:29 GMT  
Organization: Reference.Com Posting Service  
Message-ID: <**56nigp\$mrs@bilbo.reference.com**>

### **SOFTWARE PROGRAMMER**

Position available for Software Programmer experienced in generating software for PC-Based **Voice Mail** systems. Experienced in **C** Programming. Must be familiar with communicating with and controlling voice cards; preferable Dialogic, however, experience with others such as Rhetorix and Natural Microsystems is okay. Prefer **5** years or more experience with PC Based **Voice Mail**, but will consider as little as **2** years. Need to find a Senior level person who can come on board and pick up code with very little training. Present Operating System is **DOS**. May go to **OS-2** or **UNIX** in future.

Please reply to:  
Kim Anderson  
AdNET  
(901) 458-2888 fax  
kimander@memphisonline.com

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Elde edilen iş özeti

computer\_science\_job  
 id: 56nigp\$mrs@bilbo.reference.com  
 title: SOFTWARE PROGRAMMER  
 salary:  
 company:  
 recruiter:  
 state: TN  
 city:  
 country: US  
 language: C  
 platform: PC \ DOS \ OS-2 \ UNIX  
 application:  
 area: Voice Mail  
 req\_years\_experience: 2  
 desired\_years\_experience: 5  
 req\_degree:  
 desired\_degree:  
 post\_date: 17 Nov 1996

YILDIZ TEKNİK ÜNİVERSİTESİ  
 BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Yapılandırılmamış Doküman Örnek Haber Metni

- 21 yaşındaki inşaat işçisi Kemal Yaprak, evine dönerken para meselesi yüzünden tartıştığı arkadaşı Hilmi Baker tarafından bıçaklanarak öldürüldü.
- Katil: Hilmi Baker
- Kurban: Kemal Yaprak
- Sebep: Para meselesi
- Suç Aleti: Bıçak

YILDIZ TEKNİK ÜNİVERSİTESİ  
 BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Yapılandırılmış Doküman Amazon Kitap Sayfası

```

.....
</td></tr>
</table>
<b class="sans">The Age of Spiritual Machines : When Computers Exceed Human Intelligence</b><br>
<font face=verdana,arial,helvetica size=-1>
by <a href="/exec/obidos/search-handle-url/index=books&field-author=
    Kurzweil%2C%20Ray/002-6235079-4593641">
Ray Kurzweil</a><br>
</font>
<br>
<a href="http://images.amazon.com/images/P/0140282025.01.LZZZZZZZ.jpg">
</a>
<font face=verdana,arial,helvetica size=-1>
<span class="small">
<span class="small">
<b>List Price:</b> <span class=listprice>$14.95</span><br>
<b>Our Price:</b> <font color=#990000>$11.96</font></b><br>
<b>You Save:</b> <font color=#990000><b>$2.99</b>
(20%)</font><br>
</span>
<p><br>...

```

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Elde edilen kitap bilgileri

Title: **The Age of Spiritual Machines :  
When Computers Exceed Human Intelligence**  
Author: **Ray Kurzweil**  
List-Price: **\$14.95**  
Price: **\$11.96**  
:  
:

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Basit Çıkarım Şablonları

- **regular expression**
  - Amazon liste fiyatı:
    - `<span class="listprice">$43.16</span>`
    - öncül şablon: `"<b>List Price:</b> <span class=listprice>"`
    - şablon: `"\d\\.d{2}"`
    - devam şablonu: `"</span>"`

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## NLP'nin Bilgi Çıkarımına Katkısı

- Bilgiler dinamik web sayfalarından çıkarılacaksa basit regex şablonları yeterli olabilir.
- Bilgiler doğal, insanlar tarafından yazılmış metinlerden çıkarılacaksa NLP metotları yardımcı olabilir.
  - Part-of-speech (POS) tagging
    - Kelimelerin türünü (isim, fiil, sıfat vb.) belirleme
  - Sentaktik çözümleme
    - Kelime gruplarını, ağaçları belirleme, öğeleri bulma: NP, VP, PP
  - Anlamsal Kelime Sınıfları (WordNet'den)
    - KILL: kill, murder, assassinate, strangle, suffocate
  - Name Entity Recognition
- Örnek **Öldürülen** şablonu:
 

Bart killed Rose.

  - Öncül şablon: `[POS: V, synset: KILL]`
  - şablon: `[Phrase: NP]`

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Öğelerine ayrılmış metinler

- “ye” fiilinin nesneleri yiyecek olarak sınıflandırılabilir.

[geyiq.com/forum](http://geyiq.com/forum) - Taksim borsada bira ile yarım döner **yerken**

geyiq.com/forum > geyiq alanı > kıl oluyorum, durum oldum > Taksim borsada bira ile yarım döner **yerken**. Orijinalini görmek için ...

[www.geyiq.com/forum/archive/index.php/t-10219.html](http://www.geyiq.com/forum/archive/index.php/t-10219.html) - 3k - [Önbellek](#) - [Benzer sayfalar](#)

[Yerken Family Grave Search](#)

... It's like we've always known each other! - Pam from CA. Advertisement. Click

Here. Search Page for Surname: **Yerken**. Name: First, Middle, **Yerken** Last. ...

[www.findagrave.com/sumames/y/yerken.html](http://www.findagrave.com/sumames/y/yerken.html) - 13k - [Önbellek](#) - [Benzer sayfalar](#)

[Hürriyetim](#)

... Kelebek, 26.05.2004. Ödülü, Bush kraker **yerken** söylemeyin, ... Umanım kimse ona bu ödülü kazandıgımı, o kraker **yerken** söylemez' diye yanıtladı. ...

[www.hurriyetim.com.tr/haber/0,,sid=436&nvid=416896,00.asp](http://www.hurriyetim.com.tr/haber/0,,sid=436&nvid=416896,00.asp) - 42k - [Önbellek](#) - [Benzer sayfalar](#)

[MILLİYET INTERNET - BUSINESS](#)

... Zeytin **yerken** alzheimer oluyoruz habermiz yok. Zeytini, zehirli tekstil boyası ile veya demir sülfat gübresi ile karatıp satıyorlar. ...

[www.milliyet.com.tr/2003/12/12/business/bus07.html](http://www.milliyet.com.tr/2003/12/12/business/bus07.html) - 30k - [Önbellek](#) - [Benzer sayfalar](#)

[TürkiyeOnline.com - Haber](#)

... sağlık. Mantar **yerken** dikkat! Havalanın ısınmasıyla birlikte doğada ortaya

çıkan mantarların bilinçsiz olarak tüketilmesinin, zehirlenmelere neden ...

[www.turkiyeonline.com/haber/saglik/haber.php?story=2004\\_04\\_02\\_mantar](http://www.turkiyeonline.com/haber/saglik/haber.php?story=2004_04_02_mantar) - 20k - [Önbellek](#) - [Benzer sayfalar](#)

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Özel anlamlı Kelimeleri Belirleme Name Entity Recognition

- *NER sistemleri özel isimleri, tarih, yer, zaman ifade eden kelimeleri vs. belirlerler.*
- *Örnek:*
  - *Jack Brown saw a cat in London.*
  - *[PER Jack Brown] saw a cat in [LOC London] .*
- *PER -- Person*
- *LOC -- Location*
- *ORG -- Organization*

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Kural tabanlı NER sistemleri

- Yüksek performans
- Yüksek maliyet
- Kişi, şirket, yer isimleri listelerine ihtiyaç

```

If FullString = "New York" → LOCATION
If FullString = "California" → LOCATION
If Contains("Mr.") → PERSON
If Contains("Corp.") → COMPANY
If Contains("Inc.") → COMPANY
If Contains("Co.") → COMPANY
If Contains("Michael") → PERSON
If FullString = "Jordan" → LOCATION
  
```

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



Tarih, telefon numarası, email gibi kavramlar özel formatlarından tanımlanabilirler.

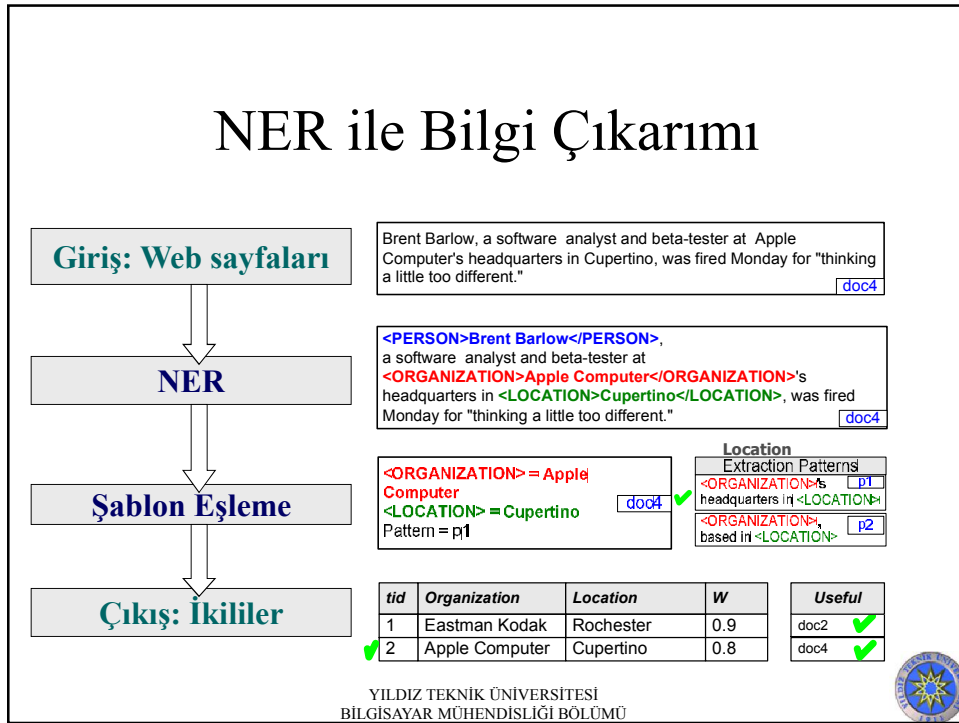
- Phone Number : (###) ###-####
- Date : ####/####
- URL : www.xxxxxx.xxx/xxx/xxxx.html
- Email Address : xxxxx@xxx.xxxxx.xxx

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

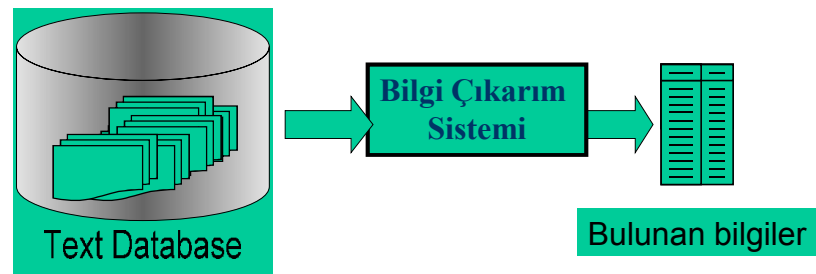




## NER ile Bilgi Çıkarımı



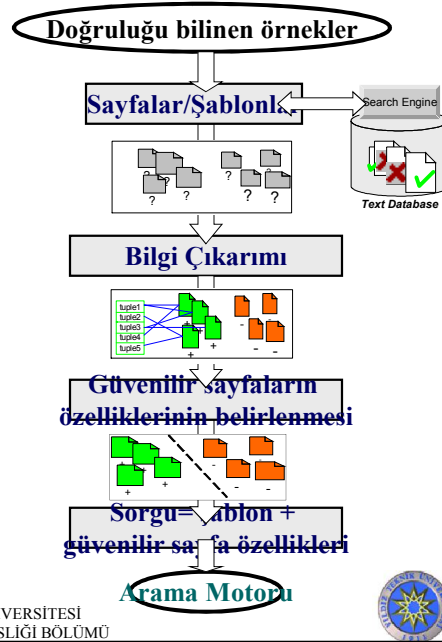
## Kaynak Seçimi



- Problem: Bulunan bilgilerin güvenilirliği.
- Problem: Tüm web sayfalarını işleyemeyiz. Çok zaman alıcı.
- Çözüm: Sadece güvenilir Web sayfalarını işleyelim. Ama nasıl?

## Kaynak Seçimi

- Doğruluğu bilinen örnekler arama motoruna gönderilir.
- Sonuçlardan şablonlar çıkarılır.
- Bu şablonlar arama motoruna gönderilir, uyan sayfalardan bilgiler çıkarılır.
- Çıkarılan bilgilerin doğruluğu bir veritabanından kontrol edilir.
- Doğru ve yanlış bilgi çıkarılan web sayfaları işaretlenir.
- Bu sayfaların özellikleri çıkarılır.
- Bundan sonra şablonlarla birlikte doğru sayfaların özellikleri de aratılır.
- Bu sayede sadece güvenilir sayfalarda arama yapılmış olunur.
- ÖZETLE: Bulunan şablonlara ek olarak, güvenilir sayfaların özellikleri de bulunarak sorguya eklenir.
- SAYFA ÖZELLİKLERİ: içinde geçen kelimeler, url'sinde geçen kelimeler (ör: edu)



YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Dinamik Web Sayfalarından Bilgi Çıkarım Metotları

- Birçok web sayfası veritabanlarından dinamik olarak oluşturuluyor.
- Dinamik web sayfalarında html tag'leri tekrar eder.
- Tekrar eden kalıplar arasında aynı tür bilgiler yer alır.



YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ





## Alişveriş Robotları

- Tekrarlı HTML taglerinin kullanılarak bilgi çıkarılan sistemlere örnek olarak çeşitli web sitelerinde satılan ürünlerin bilgilerini tek bir sayfada toplayan sistemler verilebilir.
- Örnek Siteler:
  - MySimon
  - Cnet
  - BookFinder
  - Froogle

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Alişveriş/Haber Toplama Robotlarının Çalışma Adımları

- 1- Her satıcı/haberci site bilgi çıkarım mekanizmasını kur.
- 2- Kullanıcıdan sorgusunu al. (tür, fiyat vs.)
- 3- Her site için:
  - Kullanıcı sorgusunu siteye gönder
  - Sonuç sayfasını al.
  - Sonuç sayfasını o sayfanın bilgi çıkarım mekanizmasıyla işle. Sonuçlarını kendi veritabanına kaydet.
- 4- Sonuçları (fiyatlara/tarihlere göre) sırala.
- 5- Sonuçları HTML formatına çevir. Kullanıcıya döndür.

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## İçerik sahibi

- İçeriğinin paylaşılmasını istiyorsa API ve RSS'ler ile içeriğini paylaşır.
- İstemiyorsa belirli aralıklarla html sayfa şablonlarını değiştirir, aynı yerden yapılan istekleri bloklar vb.

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## IE performansının ölçümü

- Performans, sistemin eğitimi sırasında kullanılmamış olan elle işeretlenmiş test verisi üzerinde ölçülür.
  - Dokümanlarda yer alan doğru slot sayısı:  $N$
  - Sistem tarafından çıkarılan toplam slot sayısı:  $E$
  - Sistem tarafından çıkarılan toplam doğru slot sayısı:  $C$
- Ölçütler
  - Recall =  $C/N$
  - Precision =  $C/E$
  - F-Measure =  $2 * \text{Recall} * \text{Precision} / (\text{Recall} + \text{Precision})$

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Bilgi Çıkarımında Makine Öğrenmesi

- Çıkarılmak istenen her slot (sütun) için bütün şablonların elle bulunması çok zahmetli, zaman alıcı bir iş.
- Her türlü bilgi çıkarım işi için, her site için bunu ayrıca yapmak gerekli.
- Alternatif ? Makine Öğrenmesi

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Türkçe için bir Şablon Eşleme Denemesi

- “Bütün X’ler Y’dir” ilişkisine uyan ikililerin otomatik olarak bulunması
- içerik
  - Şablonların bulunması
  - Şablonlara uygun ikililerin bulunması
  - İkililerin eklerine ayrılması
  - Verilerden örnekler
  - Sınıflandırma Metodu
  - Sınıflandırma Sonuçları

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Şablonların bulunması

- Keşfetmek istediğimiz ikililerin aralarındaki ilişki türü belirlenir. Ör: “Tüm X’ler Y’dir”.
- Bilinen X,Y ikilileri Google’da aratılır.
- X ve Y arasındaki şablonlar ve frekansları belirlenir.
- En yüksek frekanslı olan şablonlar bu ilişki türünün şablonları olurlar.

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Bulunan Şablonlardan Örnekler Tüm X’ler Y’dir için

- |                 |                      |
|-----------------|----------------------|
| • ve diğer      | • ve her türlü       |
| • ler ve diğer  | • lerden biri olan   |
| • ve benzeri    | • leri ve diğer      |
| • veya diğer    | • larından biri olan |
| • türü olan     | • lerinden biri olan |
| • ları ve diğer | • lardan biri olan   |
| • lar ve diğer  | • adı olan           |
|                 | • ve her tür         |

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Bulunan Şablonlardan Örnekler X'in yeri Y'dir için

- |                                       |                   |
|---------------------------------------|-------------------|
| • y deki x                            | x, y ye zz km     |
| • y de bulunan x                      | x (y ye zz km     |
| • y de x                              | x, y              |
| • x y de                              | x - y             |
| • x y ili sınırları içerisinde        | x bulunduğu yer:y |
| • y ili sınırlarında kalan x          | y-x               |
| • y ili sınırları içinde bulunan x    | x(y               |
| • y ilçesi sınırları içinde bulunan x | x(y               |
| • x y nin sınırları içerisinde        | x(y)              |
| • x/y                                 |                   |
| • x / y                               |                   |
| • x-y                                 |                   |
| • x y ye zz km                        |                   |

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Şablonlar

- Üretken: ürettiği bilgi miktarı
- Güvenilir: ürettiği bilgilerin doğruluk oranı
- İkisi birbiriyle çelişir





## Şablonlara uygun ikililerin bulunması

- Google’da bulunan şablonlar aratılır.
- Sonuç sayfalarındaki şablonların sağ ve sollarındaki kelimeler alınır ve bir dosyaya kaydedilir.

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Şablonlara uygun ikililerden örnekler

- |                      |                  |
|----------------------|------------------|
| • Tüm X’ler Y’dir    | • tür flamingo   |
| • kontrolör personel | • ünite aksesuar |
| • teçhizat malzeme   | • bedel masraf   |
| • kemer teçhizat     | • din azınlık    |
| • protein gıda       | • çelik yapı     |
| • Azerbaycan bölge   | • yem araç       |
| • Ceyda yardımcı     | • kız sıfat      |
| • komünizm ideoloji  | • yapı sorun     |
| • delta Gediz        | • ölçü şart      |
| • kurum Kocaelispor  |                  |
| • fotoğrafçı Robert  |                  |

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## İkililerin elle sınıflandırılması

- Bulunan ikililerden hangilerinin “Tüm X’ler Y’dir” ilişkisine sahip olup olmadığı elle işaretlenir.

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## İkililerin eklerine ayrılması

- Bulunan ikililer geçtikleri cümledeki halleriyle  
*[http://fens.sabanciuniv.edu/TL/cgi-bin/mymorp\\_keyb.cgi](http://fens.sabanciuniv.edu/TL/cgi-bin/mymorp_keyb.cgi)*  
adresindeki araç kullanılarak eklerine ayrılır.
- Araç birden fazla sonuç ürettiği durumlarda ilk çözüm kabul edilir.

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Eklerine Ayrılmış Kelimelerden Örnekler

- **adamlarından biri olan OZAN**
  - Noun+ A3pl+ P3sg+ Abl
  - Noun+ A3sg+ Pnon+ Nom
- **dualar ve her türlü ibadet**
  - Noun+ A3pl+ Pnon+ Nom
  - Noun+ A3sg+ Pnon+ Nom

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Verilerin Modellenmesi

- Farklı ek türü sayısı 53
- Farklı şablon türü sayısı 16
- 2 kelime  $(53 \cdot 2) + \text{şablon türü} = 107$  özellik
- Artık, her bir örnek 107 boyutlu bir uzayda bir noktadır.

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



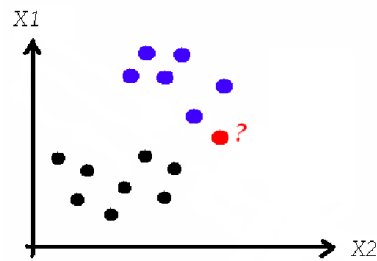
## Eğitim ve Test verileri

- Eğitici sınıflandırma metotlarında eğitim ve test verilerine ihtiyaç var.
- 365 eğitim, 365 test verisi oluşturuldu.
- Her iki veri setinde de 103'er tane istediğimiz, 262'şer tane istemediğimiz türden veri var.

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Sınıflandırma Tekrar



İki sınıf, iki özellik

$X_1$ : tansiyon

$X_2$ : ateş

Hasta kişilerin ölçümleri

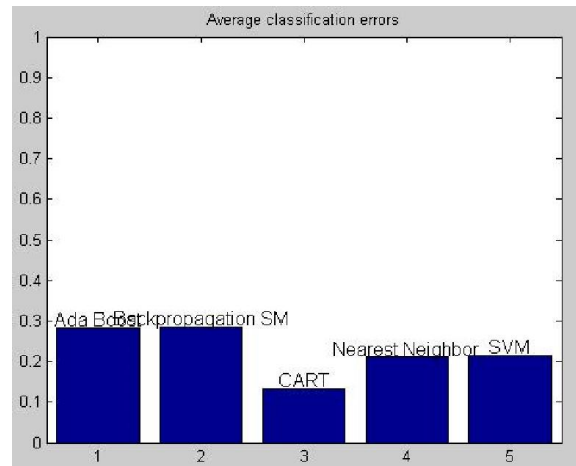
Sağlıklı kişilerin ölçümleri

Yeni ölçümler (Hasta mı? Sağlıklı mı?)

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Farklı sınıflandırıcılar için sonuçlar



YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Frekansların Kullanımı

- Çok daha fazla Internet sayfasından binlerce ikili elde edilir.
- Bulunan ikililerin tekrar sayıları (frekansları) bulunur.
- En çok tekrar eden ikililer veritabanına kaydedilir.

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Metot Hakkında

- Sınıflandırma başarısı %100 olmadığından frekans kullanımı zorunlu. Dolayısıyla çok fazla sayfa işlemek gerekli.
  - Birden fazla kelimedenden oluşan kelime grupları arasındaki ilişkiler bulunamıyor.
  - Her tür ilişkiye uygulanabilir değil.
- Ör: eş anlamlı, zıt anlamlı vs.
- Bulduğu ikililerin “Tüm X’ler Y’dir” ilişkisine uyup uymadığını ~%85’lik bir başarı ile belirleyebildiği görülmüştür.

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



## Kaynaklar

- Rada Mihalcea, “NLP lecture slides”
- [www.ccs.neu.edu/home/futrelle/bionlp/psb2001/Hawaii-Tutorial-Tsujii.ppt](http://www.ccs.neu.edu/home/futrelle/bionlp/psb2001/Hawaii-Tutorial-Tsujii.ppt)
- [www.cs.utexas.edu/users/mooney/ir-course/slides/InformationExtraction.ppt](http://www.cs.utexas.edu/users/mooney/ir-course/slides/InformationExtraction.ppt)
- [www.cs.columbia.edu/~eugene/talks/icde2003.ppt](http://www.cs.columbia.edu/~eugene/talks/icde2003.ppt)
- [www.isi.edu/natural-language/teaching/cs544/cs544-9-apr04.ppt](http://www.isi.edu/natural-language/teaching/cs544/cs544-9-apr04.ppt)
- [www.cs.sfu.ca/~zshi1/personal/projects/Presentation\\_thesis.ppt](http://www.cs.sfu.ca/~zshi1/personal/projects/Presentation_thesis.ppt)

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

