

# Bilgi Eriřim Sistemleri

***Hayri Sever***

Bařkent Üniversitesi

[sever@baskent.edu.tr](mailto:sever@baskent.edu.tr)

[www.baskent.edu.tr/~sever/](http://www.baskent.edu.tr/~sever/)

***Yařar Tonta***

Hacettepe Üniversitesi

[tonta@hacettepe.edu.tr](mailto:tonta@hacettepe.edu.tr)

[yunus.hacettepe.edu.tr/~tonta/](http://yunus.hacettepe.edu.tr/~tonta/)

# Plan

- Bilgi artışı
- Bilgi erişim sorunu
- Tanım
- Bilgi erişim sistemlerinin mantıksal yapısı
- Erişim kuralları
- Performans ölçümleri
- Bilgi erişim ve Internet



# Bilgi . . .

- Bilgi depolama ortamları
  - beyin, kültürel birikim, “dış” ortamlar
- Bilgi iletim ortamı
  - dağıtık ağlar: 14 saniyede Kongre Kütüphanesi’nin içeriğini bir yerden bir yere iletmek mümkün
- Bilgi işleme ortamları
  - beyin, bilgisayarlar, insanlardan daha “akıllı” makineler, “yapay beyin”

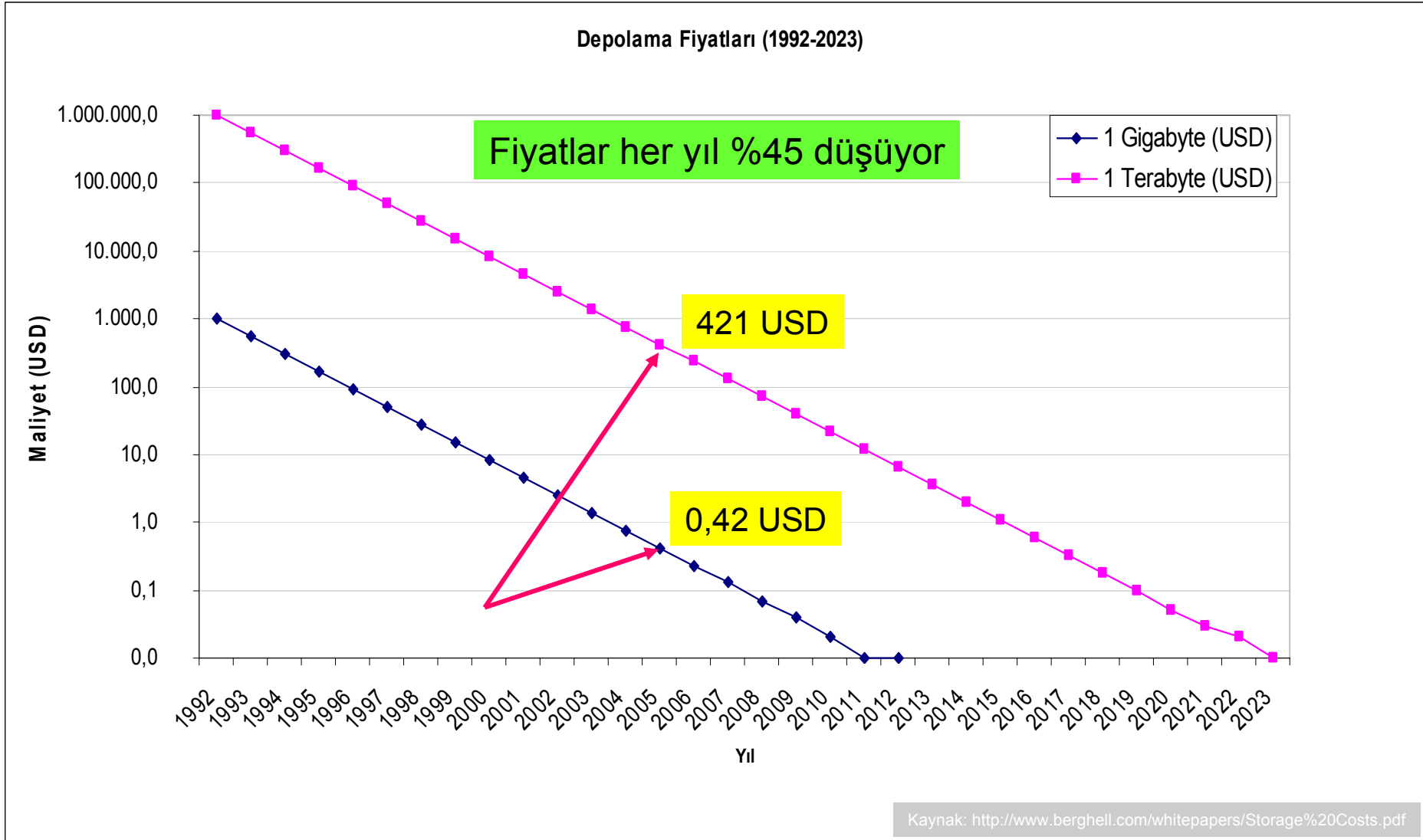


# Depolama Maliyetleri

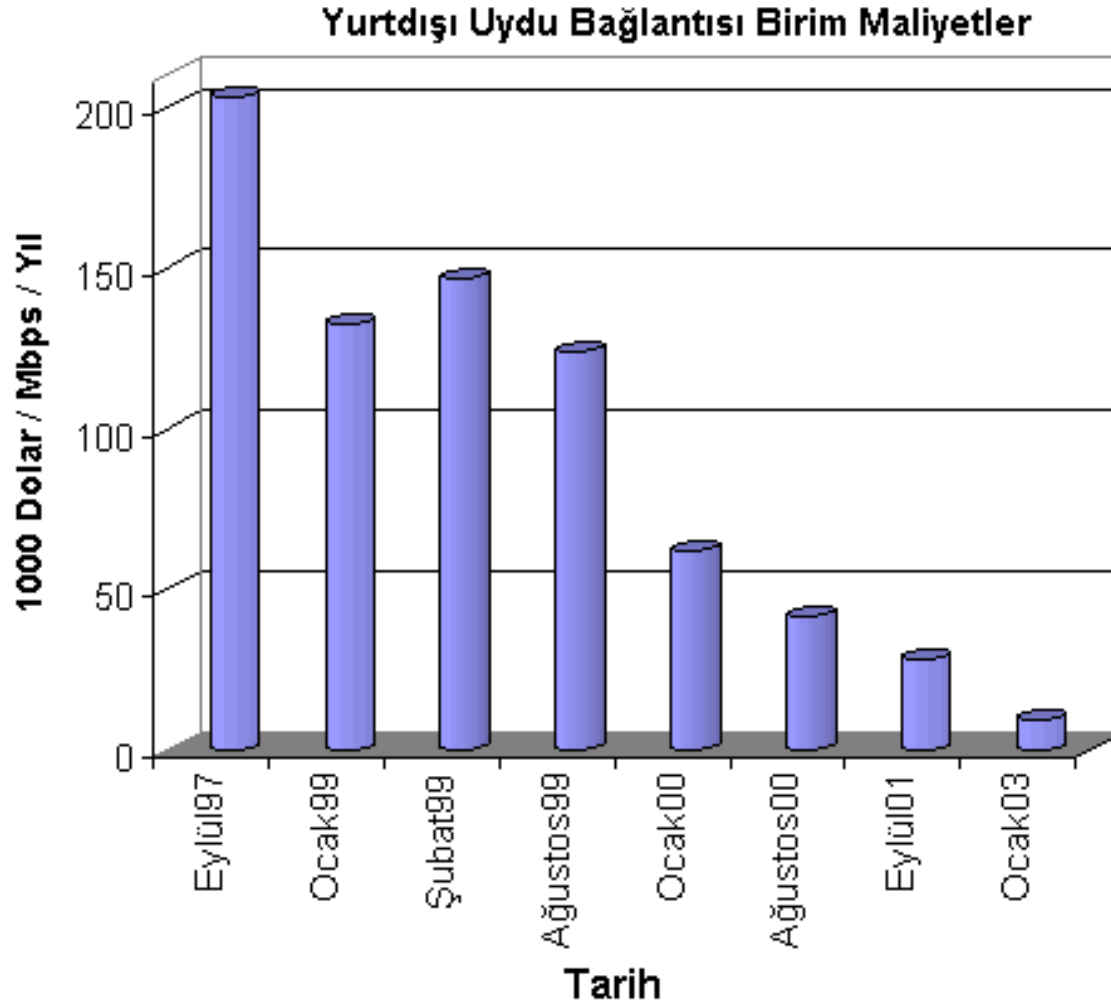
Year	Cost for 1 GigaByte = 1,000 MBytes (Storage for 2 scanned File Cabinets)	Cost for 1 TeraByte = 1,000 GBytes (Storage for 2,000 scanned File Cabinets)
1992	1.000.00	1.000.000.00
1993	550.00	550.000.00
1994	302.50	302.500.00
1995	166.38	166.375.00
1996	91.51	91.506.25
1997	50.33	50.328.44
1998	27.68	27.680.64
1999	15.22	15.224.35
2000	8.37	8.373.39
2001	4.61	4.605.37
2002	2.53	2.532.95
2003	1.39	1.393.12
2004	0.77	766.22
2005	0.42	421.42
2006	0.23	231.78
2007	0.13	127.48
2008	0.07	70.11
2009	0.04	38.56

Kaynak: <http://www.berghell.com/whitepapers/Storage%20Costs.pdf>

# Depolama Maliyetleri



# İletim Maliyetleri



Kaynak: ULAKBİM Faaliyet Raporu, 2003 ([www.ulakbim.gov.tr/hakimizda/faaliyet/faaliyet03.uhtml](http://www.ulakbim.gov.tr/hakimizda/faaliyet/faaliyet03.uhtml))

# Bilgi Patlaması

- 5 Exabyte ( $5 \times 10^{18}$  byte)
  - 2002'de dünyada üretilen yeni bilgi miktarı ( $5 \times 10^{18}$  byte)
- 5 Exabyte'lık bilgi = 37,000 yeni Kongre Kütüphanesi!
- “Yüzey web”de 10 milyar belge var (167 Terabyte) var
- “Derin web”de 550 milyar belge var (91857 Terabyte)

▪ Kaynak: BrightPlanet & Lyman and Varian

# Bir Exabyte (EB) ne kadar büyük?

<b>Kilobyte (KB)</b>	<p>1,000 bytes OR <math>10^3</math> bytes</p> <p>2 Kilobytes: A Typewritten page.</p> <p>100 Kilobytes: A low-resolution photograph.</p>
<b>Megabyte (MB)</b>	<p>1,000,000 bytes OR <math>10^6</math> bytes</p> <p>1 Megabyte: A small novel OR a 3.5 inch floppy disk.</p> <p>2 Megabytes: A high-resolution photograph.</p> <p>5 Megabytes: The complete works of Shakespeare.</p> <p>10 Megabytes: A minute of high-fidelity sound.</p> <p>100 Megabytes: 1 meter of shelved books.</p> <p>500 Megabytes: A CD-ROM.</p>
<b>Gigabyte (GB)</b>	<p>1,000,000,000 bytes OR <math>10^9</math> bytes</p> <p>1 Gigabyte: a pickup truck filled with books.</p> <p>20 Gigabytes: A good collection of the works of Beethoven.</p> <p>100 Gigabytes: A library floor of academic journals.</p>
<b>Terabyte (TB)</b>	<p>1,000,000,000,000 bytes OR <math>10^{12}</math> bytes</p> <p>1 Terabyte: 50000 trees made into paper and printed.</p> <p>2 Terabytes: An academic research library.</p> <p>10 Terabytes: The print collections of the U.S. Library of Congress.</p> <p>400 Terabytes: National Climactic Data Center (NOAA) database.</p>
<b>Petabyte (PB)</b>	<p>1,000,000,000,000,000 bytes OR <math>10^{15}</math> bytes</p> <p>1 Petabyte: 3 years of EOS data (2001).</p> <p>2 Petabytes: All U.S. academic research libraries.</p> <p>20 Petabytes: Production of hard-disk drives in 1995.</p> <p>200 Petabytes: All printed material.</p>
<b>Exabyte (EB)</b>	<p>1,000,000,000,000,000,000 bytes OR <math>10^{18}</math> bytes</p> <p>2 Exabytes: Total volume of information generated in 1999.</p> <p>5 Exabytes: All words ever spoken by human beings.</p>





# Bilgi Artışı

Tablo 1. 2002'de dünyada üretilen özgün dijital bilgi üretimi. İyimser tahminlerde bilginin dijital olarak tarandığı varsayıldı, alt tahminlerde dijital içeriğin sıkıştırıldığı varsayıldı.

Depolama ortamı	2002 Üst tahmin (Terabyte olarak)	2002 Alt tahmin (Terabyte olarak)	1999- 2000 Üst tahmin	1999- 2000 Alt tahmin	% Değişim Üst tahminler
Kâğıt	1,634	327	1,200	240	%36
Film	420,254	76,69	431,690	58,209	%-3
Manyetik	5187130	3,416,230	2,779,760	2,073,760	%87
Optik	103	51	81	29	%28
<b>TOPLAM</b>	<b>5,609,121</b>	<b>3,416,281</b>	<b>3,212,731</b>	<b>2,132,238</b>	<b>%74.5</b>

Kaynak: Lyman ve Varian

# Bilgi Miktarı

- 5,4 Exabyte (milyar x milyar byte: 54 milyar *Economist* dergisinin içeriğine eşit)
- ABD'de her yıl 80 milyar fotoğraf çekiliyor
- 2 milyar röntgen filmi çekiliyor
- Günde 610 milyar e-posta mesajı gönderiliyor
- Her yıl 15 trilyon sayfa yazıcılardan çıktı alınıyor



# Bilgi Eriřim Sorunu

- Wells, “World Encyclopedia” (1936)
- Bush, “As we may think,” *Atlantic Monthly*, (1945)
  - Memex (memory expansion)
- “bilgi eriřim” (IR) teriminin ilk kez kullanımı (Calvin Mooers, 1952)
- Otomatik dizinleme – KWIC/KWOC (Luhn, 1958)
- Boole modeli (Lockheed, 1960’lar)
- Mantıksal model (Mooers, Cooper & Maron, Van Rijsbergen, 1960- )
- Olasılık modeli (Maron-Kuhns, 1960; Robertson-Jones, 1976; Robertson-Marón-Cooper, 1982; Croft, 1979 )
- Vektör uzayı modeli (Gerard Salton, 1961)
- İstatistiksel ağırlıklandırma (tf\*idf, 1970’ler)
- Dil modelleri (Ponte-Croft, 1998)
- Performans ölçümleri
  - Cranfield, Medlars, SMART, STAIRS, TREC, 1960- (Cleverdon, Lancaster, Salton, Blair-Marón, Harman)

# “Memex ve Türkler”

“The owner of the memex, let us say, is interested in the origin and properties of the bow and arrow. Specifically he is studying why the short **Turkish bow** was apparently superior to the English long bow in the skirmishes of the Crusades. He has dozens of possibly pertinent books and articles in his **memex**. First he runs through an encyclopedia, . . . Thus he goes, building a trail of many items. . . Thus he builds a trail of his interest through the maze of materials available to him.

And his trails do not fade. Several years later, his talk with a friend turns to the queer ways in which a people resist innovations, even of vital interest. He has an example, in the fact that the outraged Europeans still failed to adopt the **Turkish bow**. In fact he has a trail on it. A touch brings up the code book. . . . It is an interesting trail, pertinent to the discussion. So he sets a reproducer in action, photographs the whole trail out, and passes it to his friend for insertion in his own **memex**, there to be linked into the more general trail.”



# Bilgi Eriřim

- “bilgi toplama, sınıflama, kataloglama, depolama, büyük miktardaki verilerden arama yapma ve bu verilerden istenen bilgiyi üretme (veya gösterme) tekniđi ve süreci”



- Sistem felsefesi
- Bilim felsefesi
- Dil felsefesi

# Dil Felsefesi ve Bilgi Eriřim

- Bilgi dzenleme ve bilgi eriřim de dil kullanımının zel bir tr
- Sistematik dizinleme (J. Otto Kaiser)
- Dilbilimsel yapılar
  - Szdađarı
  - Anlambilim
  - Szdizimi
- Kataloglama, sınıflama, dizinleme



# Sınıflama Kuramı

- Sınıflama dile dayanıyor
- Dil belirsizlikler içeriyor
- Kavramlar üzerinde anlaşma sağlamak zor
- Domates “meyve” mi “sebze” mi?
- George Lakoff, “Women, Fire and Dangerous Things”
- “Alternatif tıp” hangi konuya girer?
  - Felsefe?
  - Din?
  - Sağlık ve tıp?







American Libraries • October 2000

*"Well, it all started with an unsuccessful subject search"*

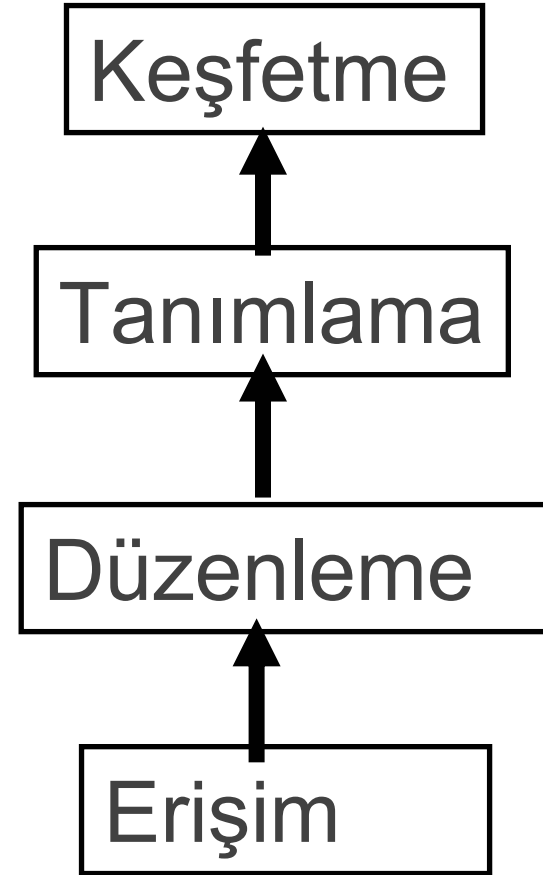
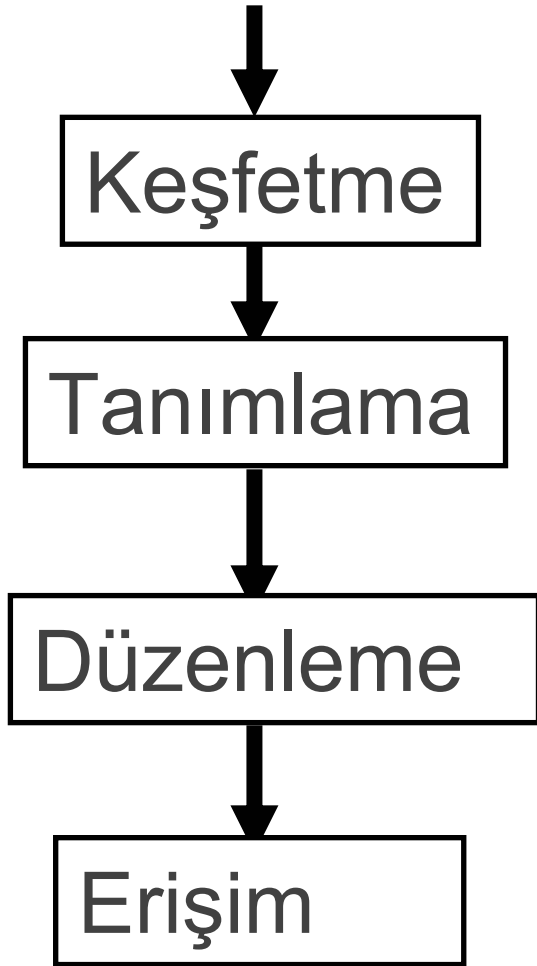


# Bilgi Eriřimin Temel İkilemi

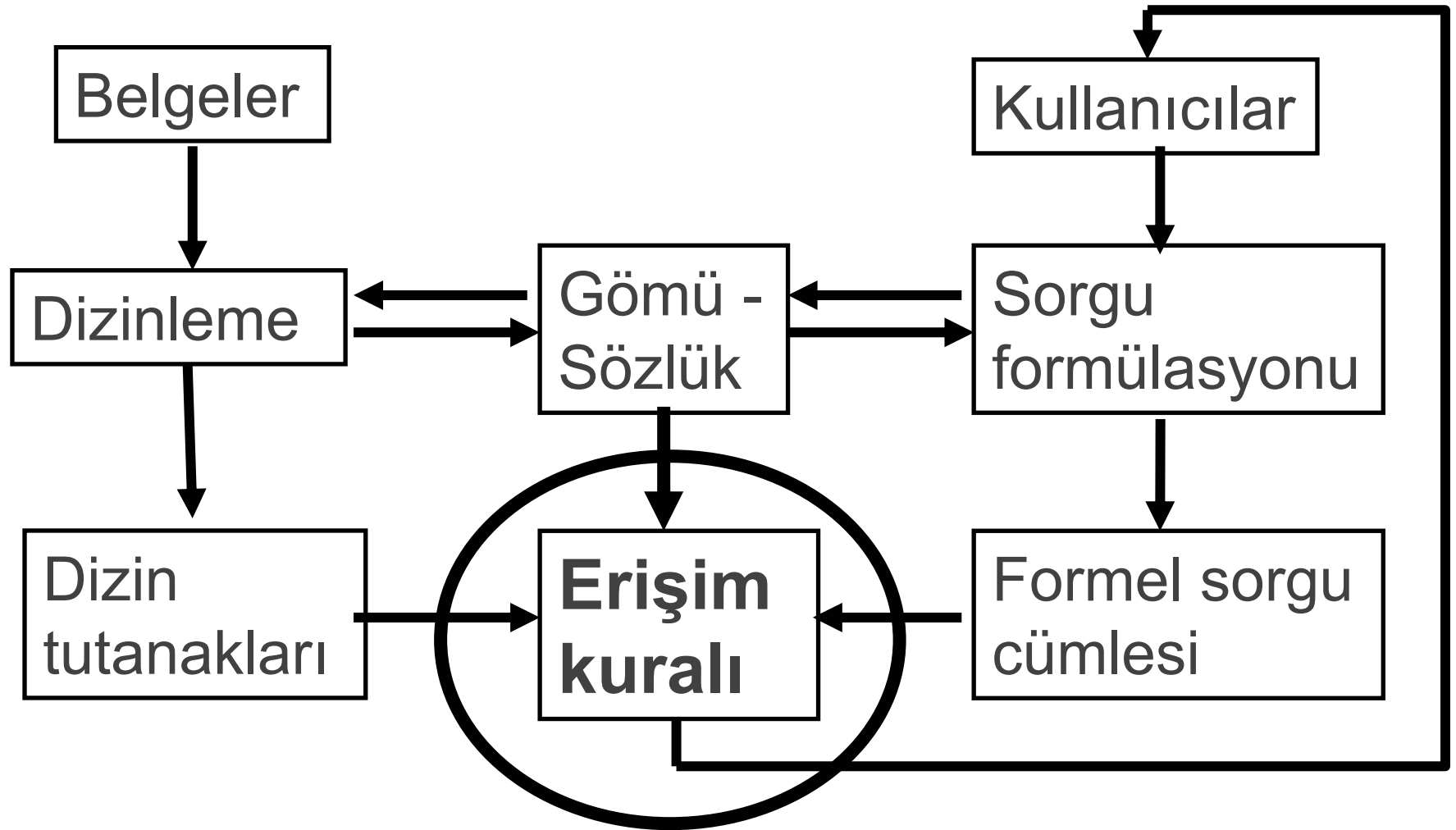
- “Hakkında bilgi bulmak için bilmediđin bir Őeyi tanımlama geređi”  
(Hjerrpe)



# Bilgi Keşfetme, Tanımlama, Düzenleme ve Erişim



# Belge Erişim Sisteminin Mantıksal Düzenlemesi



Kaynak: Maron, 1984

# İdeal Bilgi Erişim Sistemi

- İlgili belgelerin tümüne ve salt ilgili belgelere erişim sağlamalı
- “İlgililik” kavramı
  - Nesnel ilgililik
  - Öznel ilgililik
- Birbirine benzeyen bilgileri bir araya getirmek, benzemeyenleri ayırmak



# Eriřim Kuralları

Boole mantığı

Set kuramına dayanıyor. Boole işleçleri  
–VE, VEYA, DEĞİL- kullanılıyor

Vektör uzayı modeli

$$\sigma(D, Q) = \frac{\sum(t_k \times q_k)}{\sqrt{\sum(t_k)^2} \times \sqrt{\sum(q_k)^2}}$$

$t_k$  = k teriminin belgedeki değeri  
 $q_k$  = k teriminin sorgudaki değeri

Olasılık modeli

$$P(\text{ilgili}) = n / N$$

$$P(\neg \text{ ilgili}) = 1 - P(\text{ ilgili}) = N - n / N$$

n = ilgili belge sayısı  
N = toplam belge sayısı

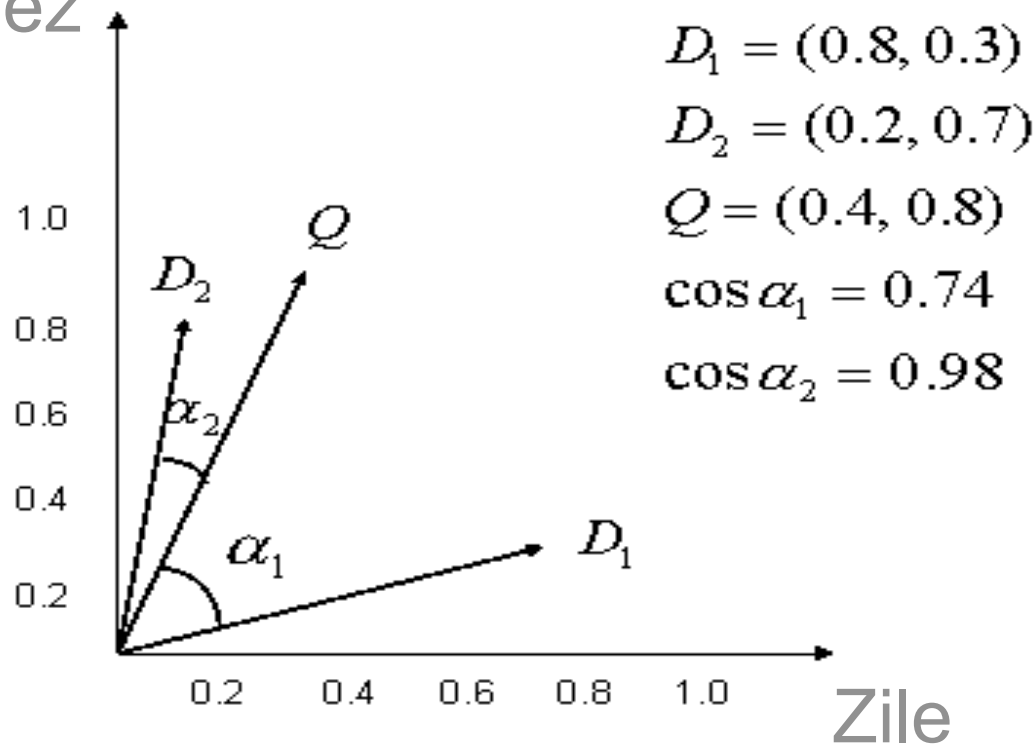
İstatistiksel ağırlıklandırma (tf\*idf)

**Ağırlıklandırma ilkesi:** İlgili belgelerde sık **AMA** derlemin tamamında seyrek geçen terimleri daha yüksek ağırlıklandır

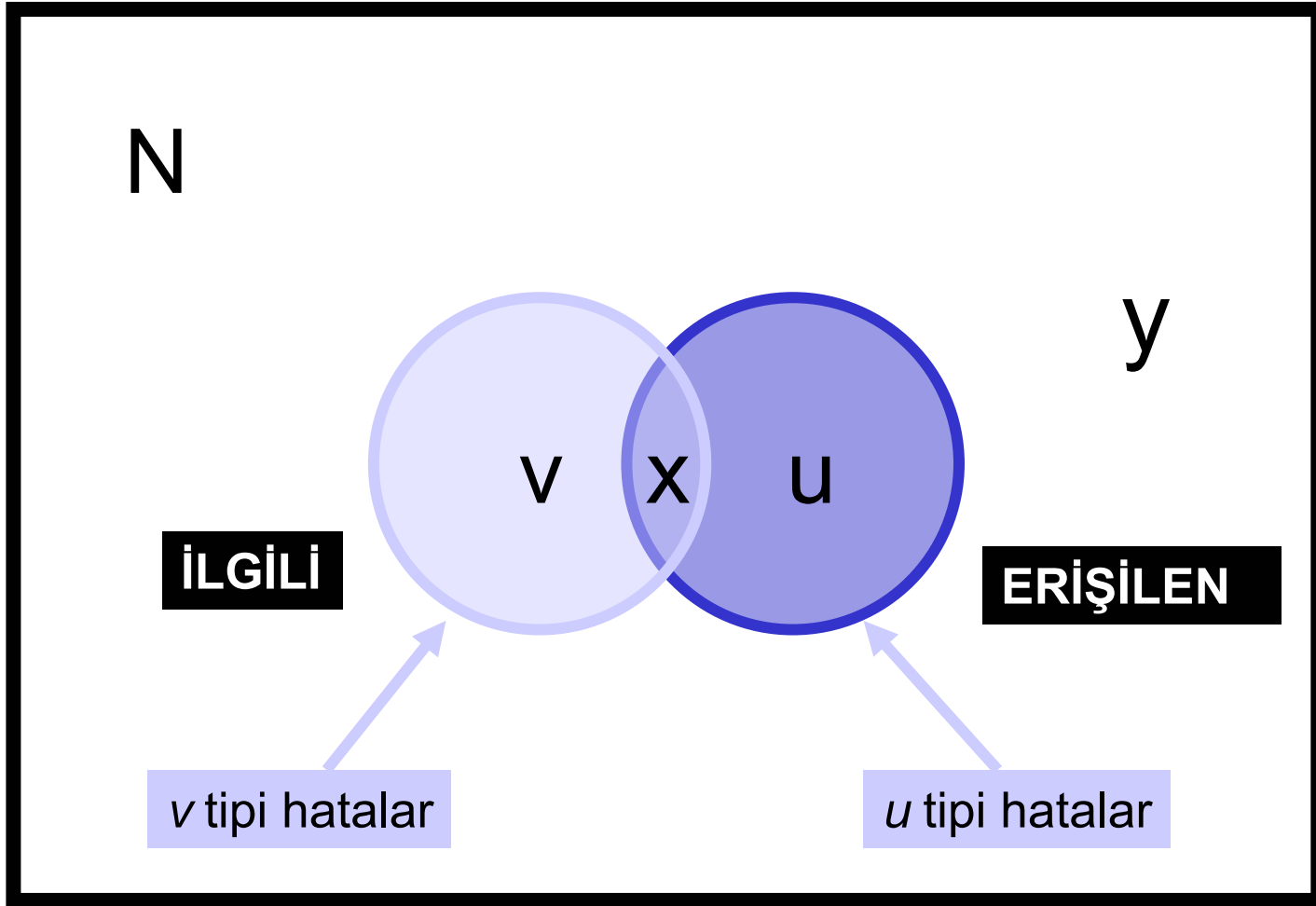


# Benzerlik Skorunun Hesaplanması

Pekmez



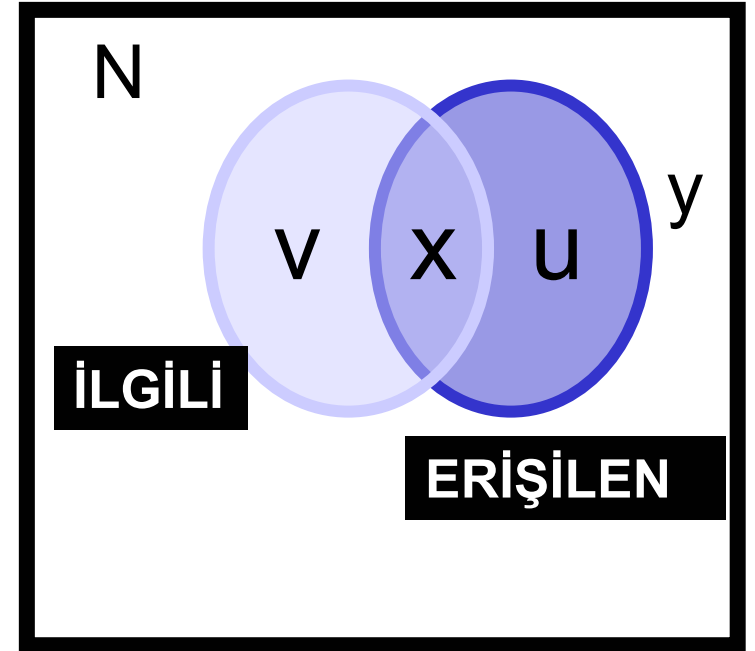
# Bilgi Eriřim Sistemleri Mükemmel Deęil!





# Bilgi Eriřim Performansı

	İLGİLİ	İLGİSİZ	
ERİŐİLEN	x	u	$n_1$
ERİŐİLE -MEYEN	v	y	
	$n_2$		



$$\text{Duyarlık} = x / n_1$$

$$\text{Anma} = x / n_2$$

$$\text{Posa} = u / u + y$$

$$\text{Genellik} = n_2 / N$$

# Yetersizlik Aksiyomları I

- Bir bilgi ihtiyacı bağlamdan bağımsız olarak ifade edilemez.
- Bir makineye bir soruyu uygun arama terimlerine çevirmesini öğretmek olanaksızdır.
- Bir belgenin ilgili olup olmaması görülen diğer belgelere bağlıdır.
- Bütün ilgili belgelerin bulunup bulunmadığını doğrulamak asla mümkün değildir.
- Makineler anlamı tanıyamaz -> entellektüel dizinleme kadar başarılı değildir, vs.



# Yetersizlik Aksiyomları II

- Sözcük sıklığı istatistikleri ne anlamı temsil edebilir, ne de anlam yerine geçebilir.
- Bir bilgi erişim sisteminin bir tekrarlı süreci destekleme yeteneği insanlar tarafından sadece bir kez yapılan ilgililik değerlendirmesiyle değerlendirilemez.
- Ya sağlam ilgililik değerlendirmesi ya da çok etkili mekanik süreçlere sahip olabilirsiniz, ama ikisine birden asla.
- Yani, tutarlı bir şekilde etkin olan tamamen otomatik dizinleme ve erişim mümkün değildir.



# Internet

- Zaman ve mekân engelini ortadan kalkması
- Bilgi kaynaklarına ve hizmetlerine günde 24 saat haftada 7 gün uzaktan erişim
- “Anında memnuniyet”



# Bilgi Eriřim ve Internet...

- “Yangın hortumundan su içmek”
- “Dijital belgeler”: devingen, sınırları belirsiz, kendi kendini deęiřtirebilen belge
- Dizinleme ve bilgi eriřim teknikleri yetersiz
- Eriřim doğrusal ve hantal
- İnsan beyninde ise dizinleme ve eriřim “baęıntılı”
- Acaba yakın gelecekte taklit edilebilir mi?
- Ses, koku, vs. bilgisine eriřim?
- Beyin dıřında kayıtlı bilgiler insanın düşünme ve sorun çözme gücünün bir parçası haline getirilebilir mi?



# Bilgi Eriřim Sistemleri

***Hayri Sever***

Bařkent Üniversitesi

[sever@baskent.edu.tr](mailto:sever@baskent.edu.tr)

[www.baskent.edu.tr/~sever/](http://www.baskent.edu.tr/~sever/)

***Yařar Tonta***

Hacettepe Üniversitesi

[tonta@hacettepe.edu.tr](mailto:tonta@hacettepe.edu.tr)

[yunus.hacettepe.edu.tr/~tonta/](http://yunus.hacettepe.edu.tr/~tonta/)



# Bilgi Eriřim Sistemleri II

Hayri Sever  
Bilgisayar Mühendislięi Bölümü  
Bařkent Üniversitesi  
06530 Baęlıca Ankara  
[sever@baskent.edu.tr](mailto:sever@baskent.edu.tr)

Yařar Tonta  
Bilge ve Belge Yönetimi  
Hacettepe Üniversitesi  
06532 Beytepe, Ankara  
[tonta@hacettepe.edu.tr](mailto:tonta@hacettepe.edu.tr)



# Alt Başlıklar

Metin Teknolojileri

Bilgi Geri-Erişim Sistemi İşlevsel Mimarisi

Dizinleme

Sorgulama

Bilgi Geri-Erişim Modelleri

Arama Motorları

Üst Arama Motorları

Konu Algılama ve İzleme

Bilgi Süzme





# Metin-Tabanlı (Dil) Teknolojiler

- Bilgi Geri-Erişim (BGE)
- Soru Yanıt (SY)
- Bilgi Çıkarma
- Bilgi Süzme
- Umlama
- Özetleme
- Konu Algılama ve İzleme (KAİ)
- Makine Çevrimi
- Ses Tanıma

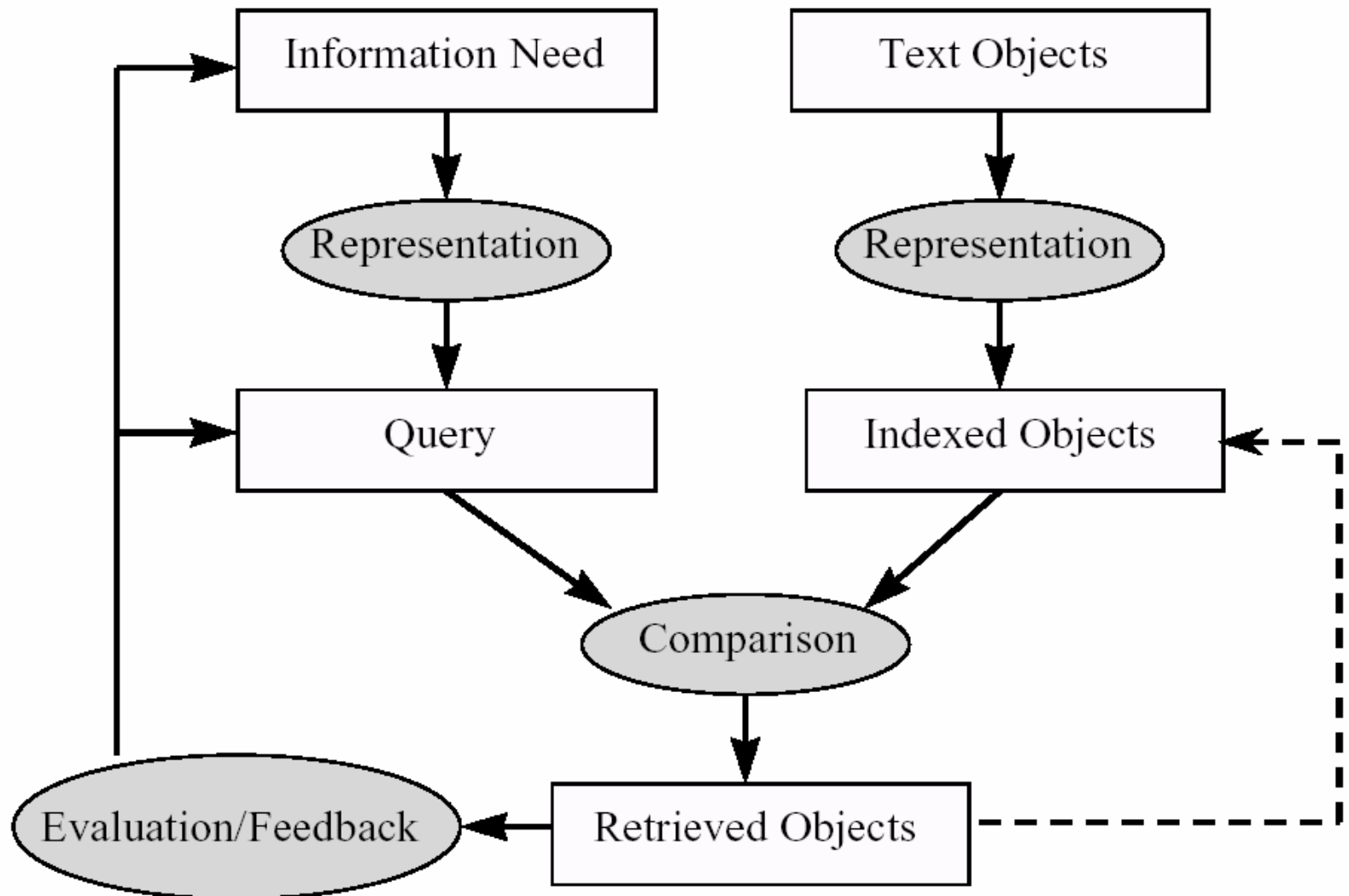


# Bilgi Geri-Eriřim Meseleleri

- Belge nedir ve boyu nasıl hesaplanır?
- Bu belge ne hakkındadır?
- Bu sorgu ne hakkındadır?
- Bu sorgu ve belge aynı Őey hakkında mıdır?
- Bu belge verilen sorgu ile ilgili midir?
- Bu belge sisteme sunulan bilgi ihtiyacı ile ilgili midir?
- Bu belge ne kadar ilgilidir?
- Bu veritabanı verilen sorgu ile ilgili midir?
- Bu resim ne hakkındadır?



# Bilgi Geri-Erişime İşlevsel Bakış



# Belge İşleme ve Gösterimi

- Ön işlem: Noktalama işaretlerinin kaldırılması ve daha sonra durma listesinde bulunan kelimelerin belgeden ayıklanması.
- Gövdeleme: bir kelimededen yapım eklerinin korunup çekim eklerinin atılması.
- Belge Gösterimi için içerik terimleri ve onların göreceli ağırlıkları. Bir terimin ağırlığı onun belge içindeki sıklığı ile doğru, fakat derlem sıklığı ile ters orantılıdır.



# Dizinler

## ➤ Dizin ne içermelidir?

Veritabanı sistemi asıl ve ikincil anahtarları dizinler.

- BGE Problemi: anahtarları kestirebilmek?

- Çözüm: İçerik terimleri.

## ➤ Zipf Kanunu: Terimlerin dağılımı ve sıraları arasındaki ilişki sabit bir değere yakınsar.

## ➤ İçerik terimlerin göreceliği ağırlığı ne olmalıdır?

- Sıklık Modeli: Terim sıklığı? Belge sıklığı?

- Ayrımsama Modeli: belge uzayının yoğunluğunu azaltan terim iyi bir terimdir.

- Dil modeli: Belgenin sözkonusu terimi üretme olasılığı ile derlemin üretme olasılığı arasındaki doğrusal ilişki ağırlığı belirler.



# Zipf Kanunu

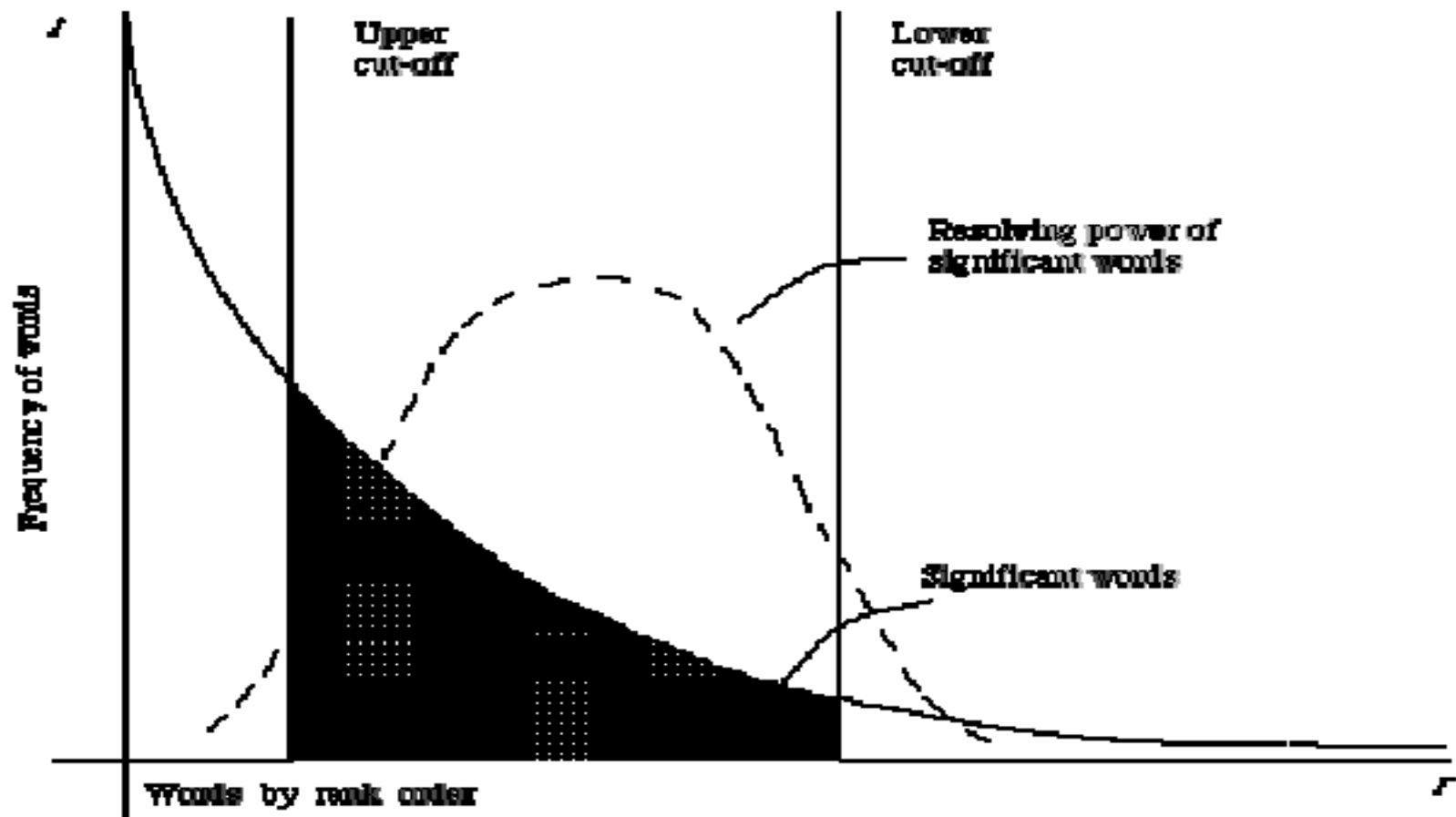
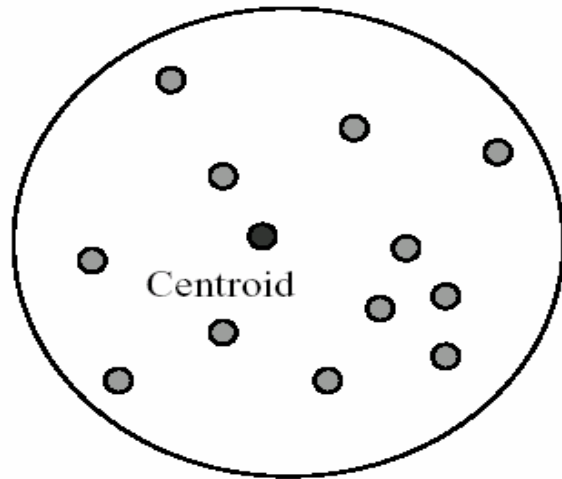
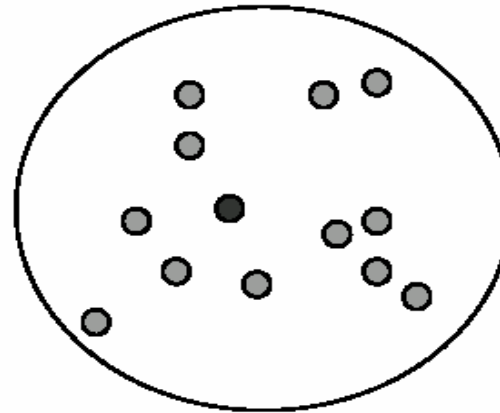


Figure 2.1. A plot of the hyperbolic curve relating  $f$ , the frequency of occurrence and  $r$ , the rank order (Adapted from Schuler<sup>44</sup> page 120)

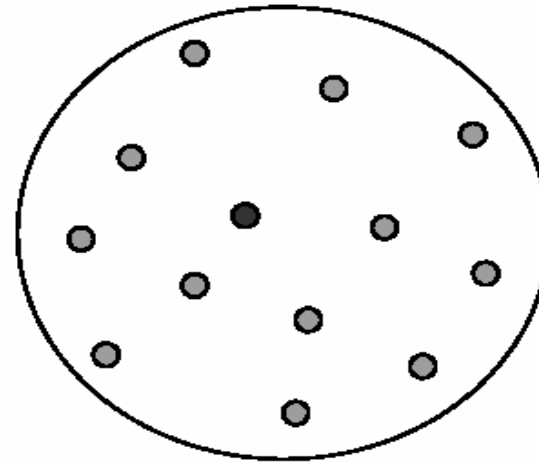
# Ayrımsama Modeli



Document space with all terms



After removal of a good discriminator



After removal of a poor discriminator

# Sorgu İfadesi

- 2 temel sorgu dili türleri
  - Boole, yapılı
  - Serbest metin
- Bir çok sistem birisini ya da her ikisini birden desteklemektedir.
- Sorgu ifadesinin oluşturulmasında kullanıcı arayüzü önemlidir.
- Sorgu ifadesinin oluşturulması için araçlar
  - Sorgu işleme ve ağırlıklandırma
  - Sorgu genişletme
  - Sözlükler ve eş anlamlı sözlük
  - İlgililik geri bildirme





# Sorgu İşleme

- Sorgu işleme adımları otomatik belge dizinlemeninkilerden çok benzerdir.
  - Durma Kelime Listesi farklı olabilir
  - Metin daha az gramatik ve kısa olabilir
- Kullanıcı etkileşimi mümkün ve istenebilir
- Sorgu-tabanlı gövdeleme ve durma kelimeleri
- Diğer olası adımlar
  - Tamlamaların tanınması
  - Negatiflerin tanınması
  - İlgili kelimelerle sorguların genişletmesi

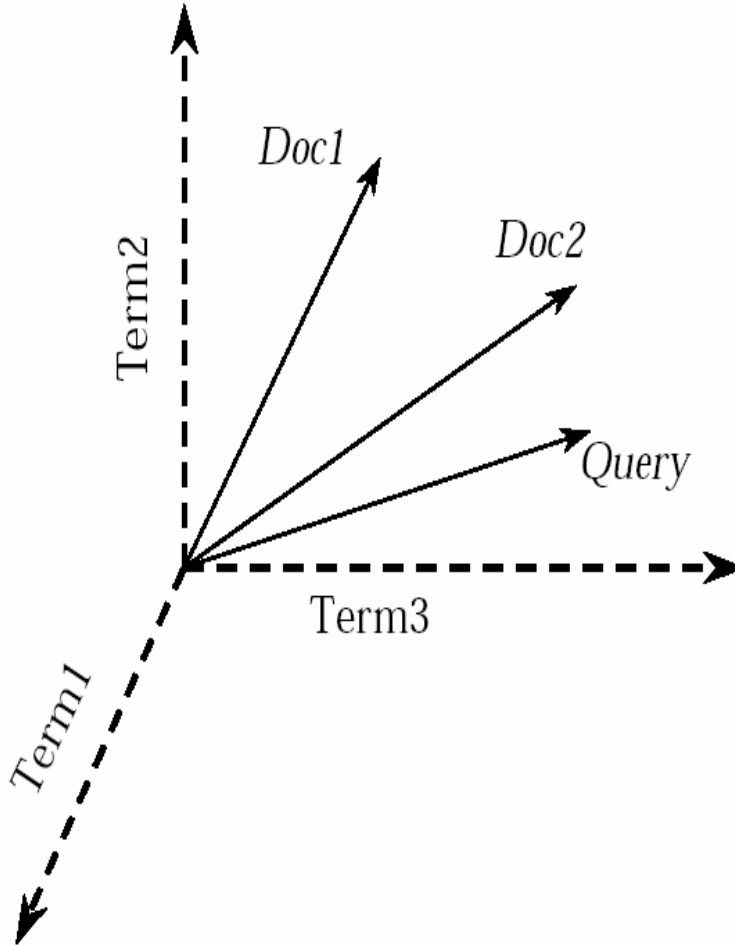


# Geri-Eriřim Modelleri

- Boole model kesin eřleřtirme yaklařımına dayanmaktadır.
- Sorgular belge özelliklerini işlenenler olarak kabul eden mantık ifadeleridir.
  - Geri getirilen belgeler genelde sıralanmaz.
  - Acemi/Tecrübesiz kullanıcılara Boole sorgu ifadesi zor gelebilir.
  - Boole geri-eriřim modeli ile Boole sorguları birbirlerinden ayırma gereksinimi
  - Saf Boole işleçleri: VE, VEYA, VE DEĞİL
  - Bir çok sistem uzaklılık işleçlerine sahiptir
  - Bir çok sistem basit düzenli ifadeleri desteklemektedir



# Vektör Uzayı Bilgi Geri Erişim Modeli



- Belge, terimlerin bir vektörü olarak gösterilir.
- Sorgu, serbest metin veya terimlerin bir vektörü olarak gösterilir.
- İki vektör arasındaki açı benzerlik ile ters orantılıdır.
- Belgeleri sorguya benzerliklerine göre sıralar.

# Vektör Uzayında Benzerlik: Ortak Ölçümler

Sim(X,Y) Binary Term Vectors

Weighted Term Vectors

**Inner product**  $|X \cap Y|$

$$\sum x_i \cdot y_i$$

**Dice coefficient**  $\frac{2|X \cap Y|}{|X| + |Y|}$

$$\frac{2 \sum x_i \cdot y_i}{\sum x_i^2 + \sum y_i^2}$$

**Cosine coefficient**  $\frac{|X \cap Y|}{\sqrt{|X|} \sqrt{|Y|}}$

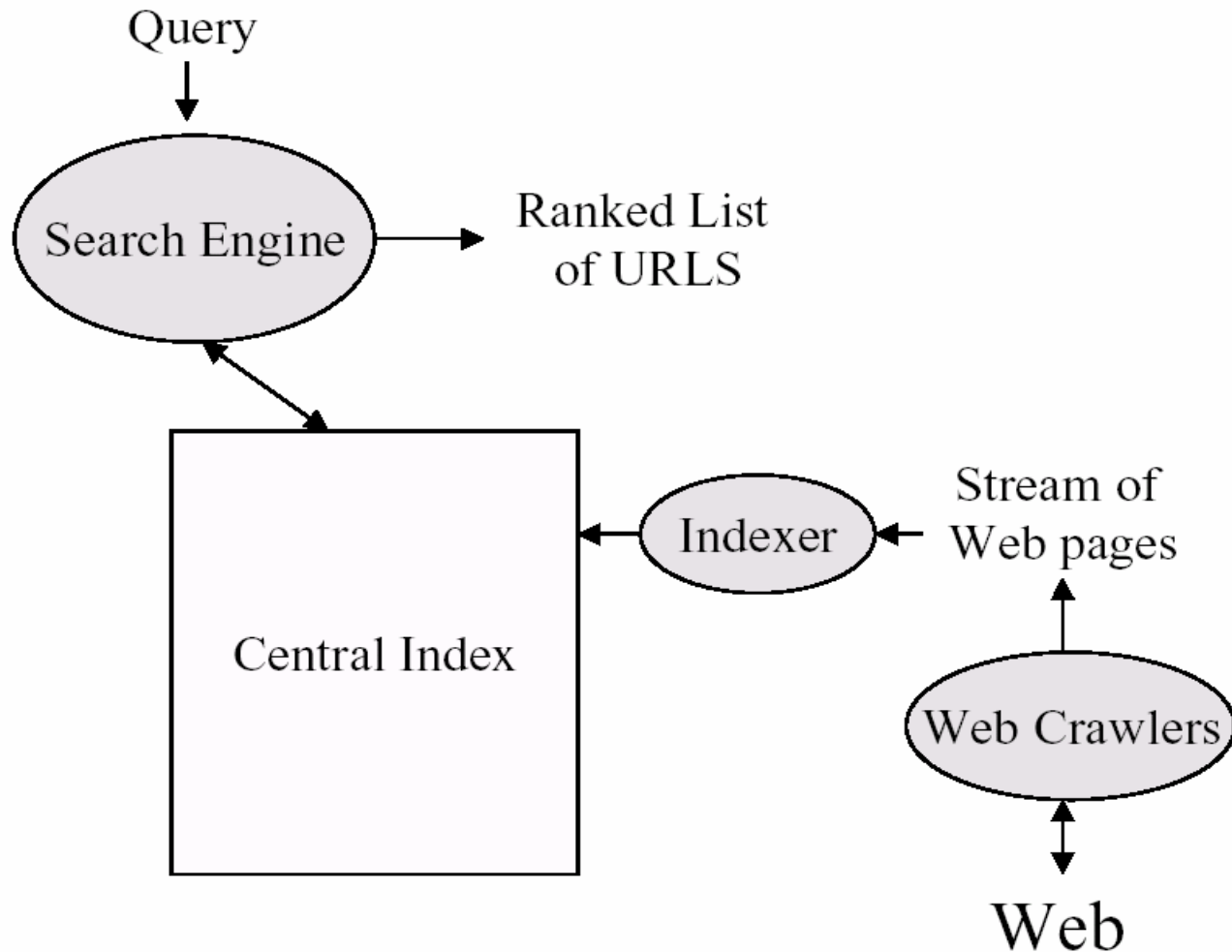
$$\frac{\sum x_i \cdot y_i}{\sqrt{\sum x_i^2} \cdot \sqrt{\sum y_i^2}}$$

**Jaccard coefficient**  $\frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|}$

$$\frac{\sum x_i \cdot y_i}{\sum x_i^2 + \sum y_i^2 - \sum x_i \cdot y_i}$$

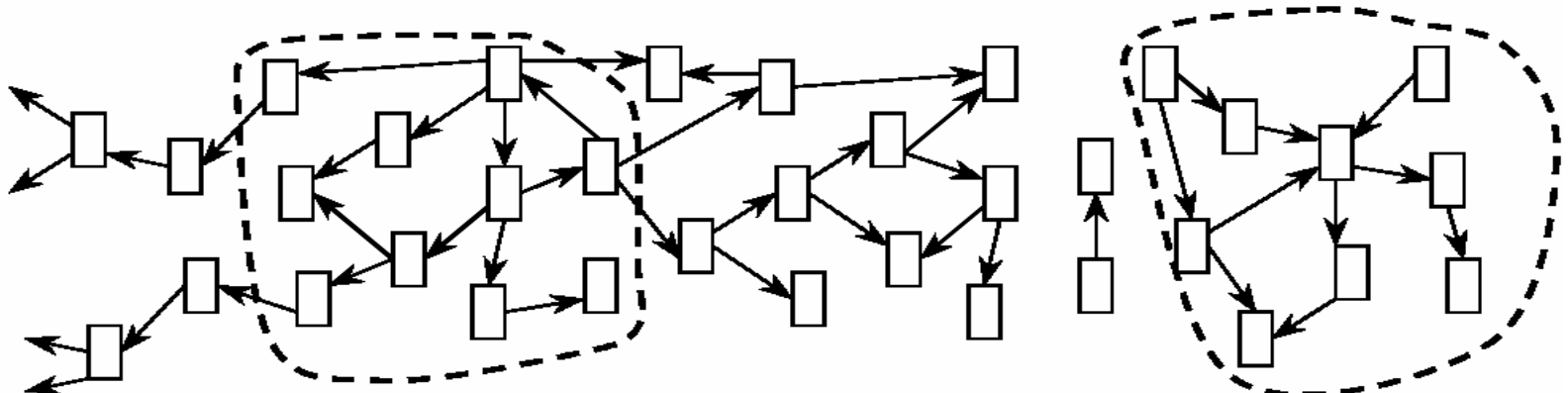


# Arama Motorunun Merkezi Mimarisi

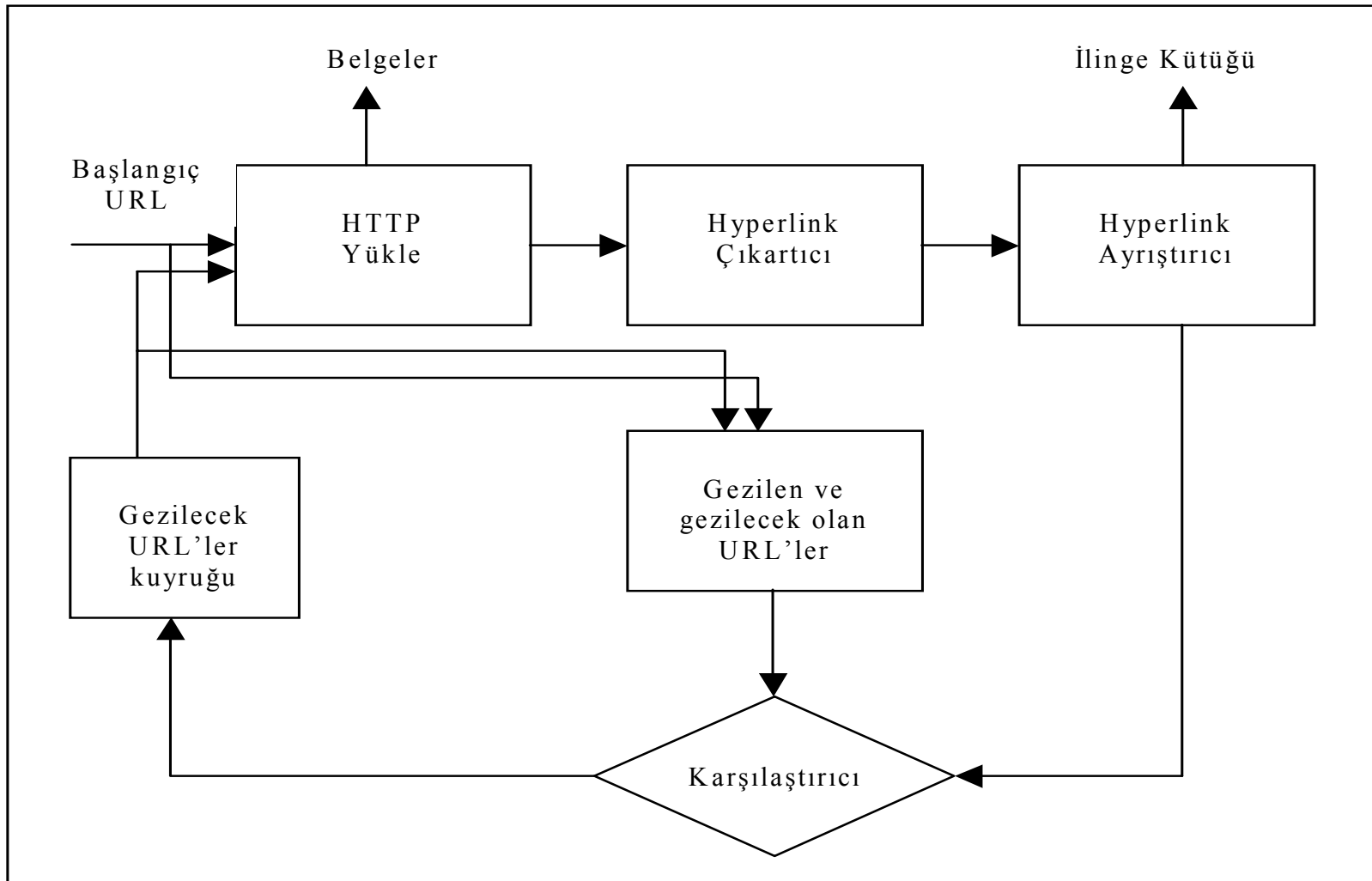


# Web Örümceği ve Veri Toplama

- Hiper-bağlantılı belgeler çizgedeki düğümler olarak görülebilir.
- İlginç altçizgeler: alan isimleri kesişen düğümler
- İzole altçizgeler: Dışardan referans almayan düğümler
- Veri toplama meseleleri:
  - Her bir düğüm nasıl bir kere ziyaret edilecek
  - Düğümlerin temsili örnekleme nasıl elde edilir



# Web Örümceği İşlevsel Mimarisi



# Veri Toplama ve Sorgu Dili: Ortak Sorunlar

- Göreceli yollar: `<A HREF=“../../../bil367/”>Yayınlar</A>`
- Tekrarlı sayfalar (%30): Aynı sayfa, farklı adres.
- Javascript: Dinamik HTML
- Çok büyük sayfalar: 10 MB sayfayı gerçekten tümü ile dizinlemek istiyor musunuz?
- Dinamik içerik: Web kaynakları tahmini olarak ortalama 75 gün değişmeden kalmaktadırlar.
- Kaliteli Web sayfaları: Nasıl ölçülür?
- Meta öznitelikler: description, keywords, title, vs.
- Bir kaç kelimelik sorgular (ortalama 1.5)



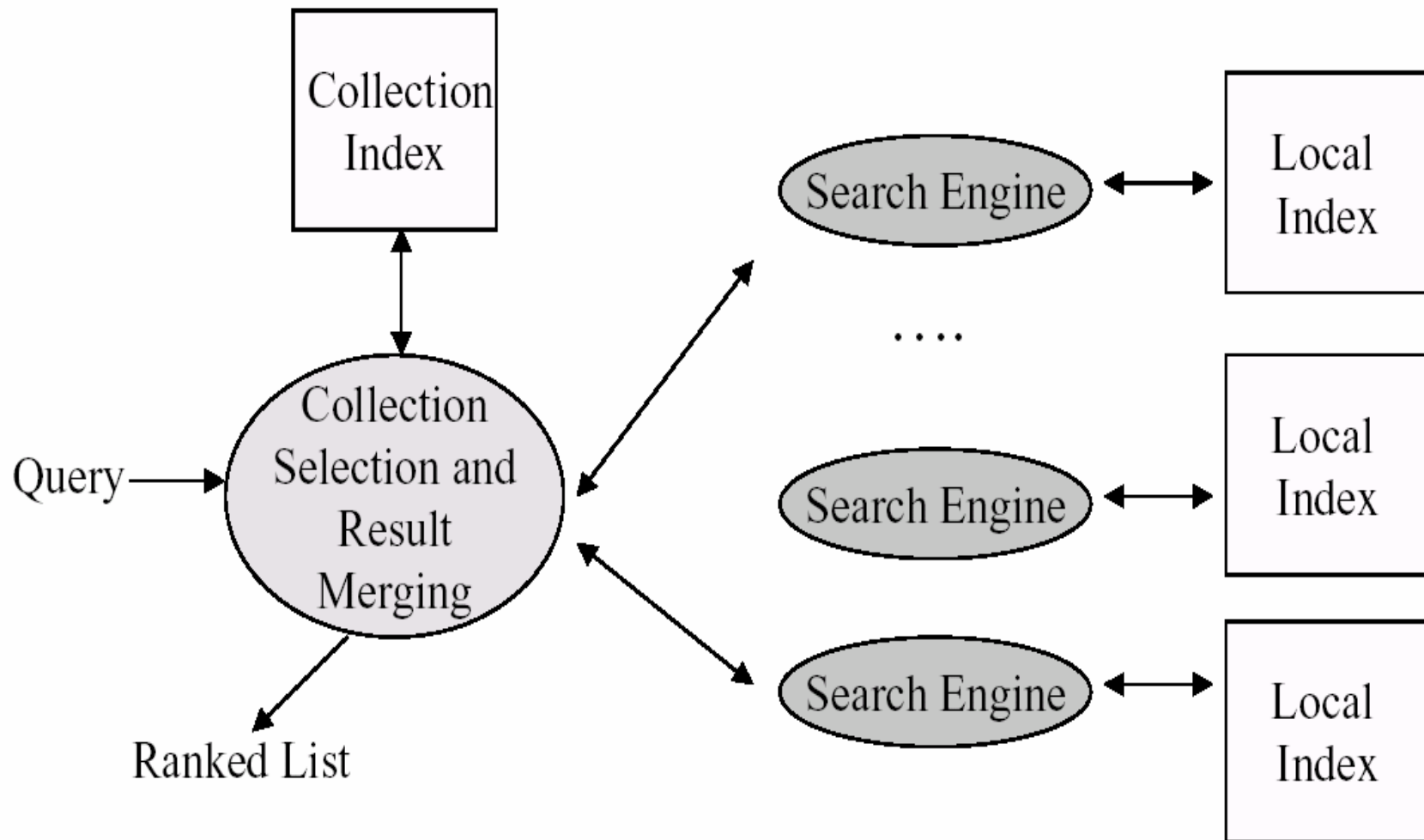


# Üst Arama Motorları

- **Tekli Çerçeve Çalışmaları**
  - Sadece bir sorgu işleme motoru.
  - Belge başlığı, özet, vücut ve kaynakça ağırlıkları.
  
- **Çoklu Çerçeve Çalışmaları**
  - Birden fazla sorgu işleme motoru.
  - Veri Birleştirme ve Koleksiyon



# Üst Arama Motorları



# Üst Arama Problemi

## ■ Skor Normalleştirme Adımı

- ◆ Farklı sistemler farklı erişim çıktılarını sunar.
- ◆ Erişim çıktılarını ortak bir ölçeğe çekilir.

## ■ Birleştirme Adımı

- ◆ CombSUM ve CombMNZ en etkili yöntemler.

Yöntem	İşlev
$CombMI_N$	Bağımsız skor değerlerinin en küçüğü
$CombMA_X$	Bağımsız skor değerlerinin en büyüğü
$CombME_D$	Bağımsız skor değerlerinin ortancası
$CombSU_M$	Bağımsız skor değerlerinin Toplamı
$CombAN_Z$	$CombSUM$ / sıfırdan farklı skor değerlerinin sayısı
$CombMN_Z$	$CombSUM *$ sıfırdan farklı skor değerlerinin sayısı

# Normalleştirme Yöntemleri

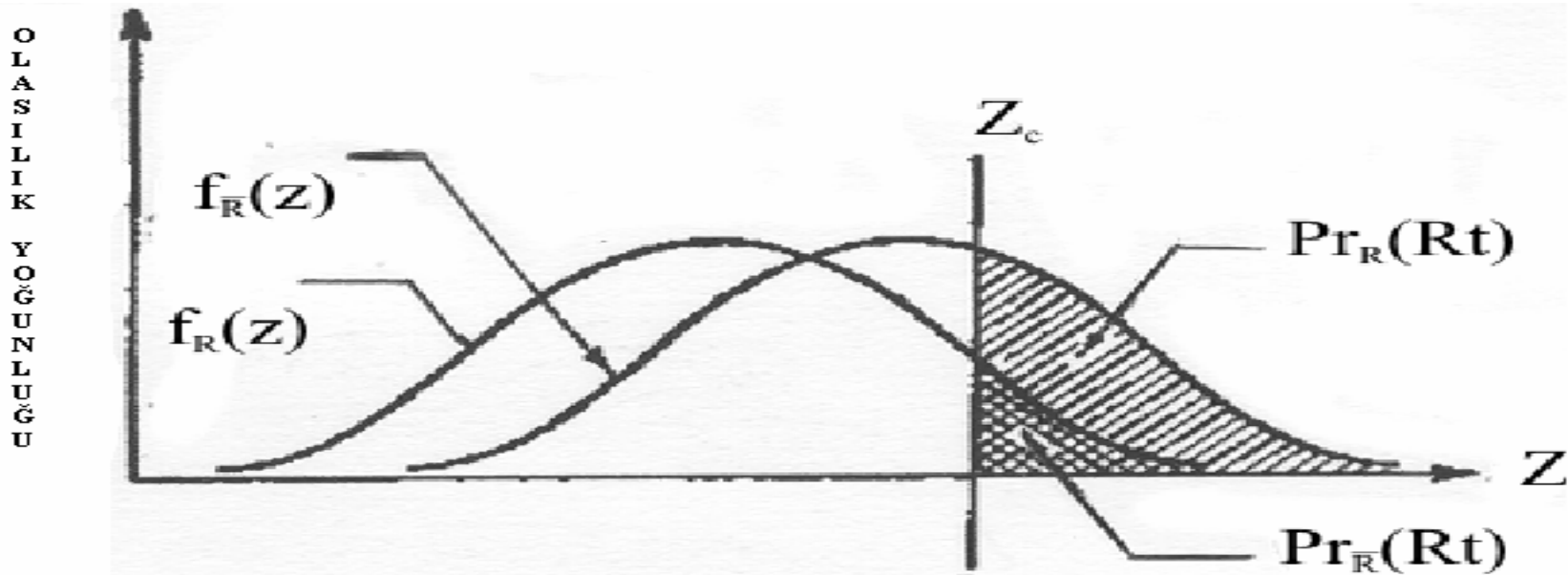
Yöntem Adı	Açıklama
<i>Standart</i>	<i>Minimum değeri 0, Maximum değeri 1 olarak belirle.</i>
<i>Sum</i>	<i>Minimum değeri 0 , ve Sum'ı 1 olarak belirle.</i>
<i>ZMUV</i>	<i>Ortalamayı 0, varyansı 1 olarak belirle.</i>

- **Belge skorlarının doğrusal olarak kaydırılması ve ölçeklenmesi.**
- **Skorlar ilgili belgeler için doğru olasılıkları yansıtmalı.**
- **4 arama sisteminden sonra azalan bir performans oluşmakta.**



# ZMUV Normalleştirme

- Bu teknik iki Gaussian dağılımın varyans ve ortalama değerlerinin normalleştirilmesidir.



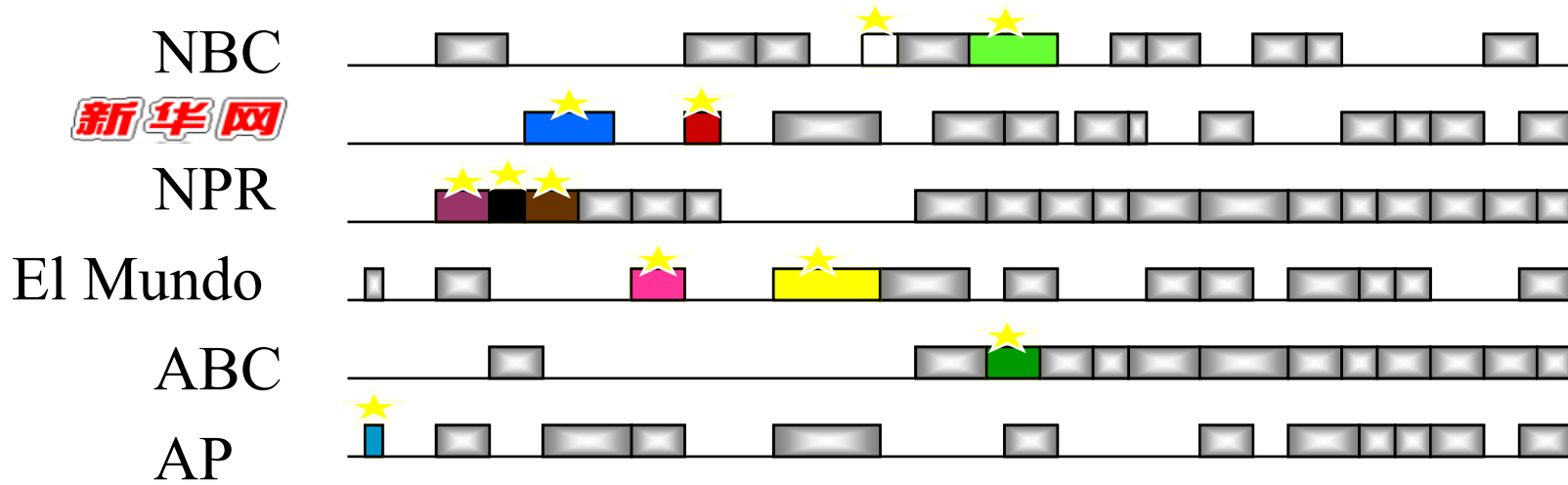
- ✓ İlgili ve ilgisiz belge skorlarının dağılımları iki normal dağılımın birleşimi ile modellenemez.

# Konu Algılama ve İzleme (KAİ)

- Haber yayınlarının izlenerek yeni ve ilginç bir haber olduğunda ilgililerin uyarılmasını sağlayacak teknolojilerin geliştirilmesi.
- Amerika İleri Savunma Araştırma Projeleri (DARPA),
- Ulusal Standartlar ve Teknoloji Enstitüsü(NIST) tarafından desteklenen
- İlk çalışmalar 1997 yılında başlamıştır.

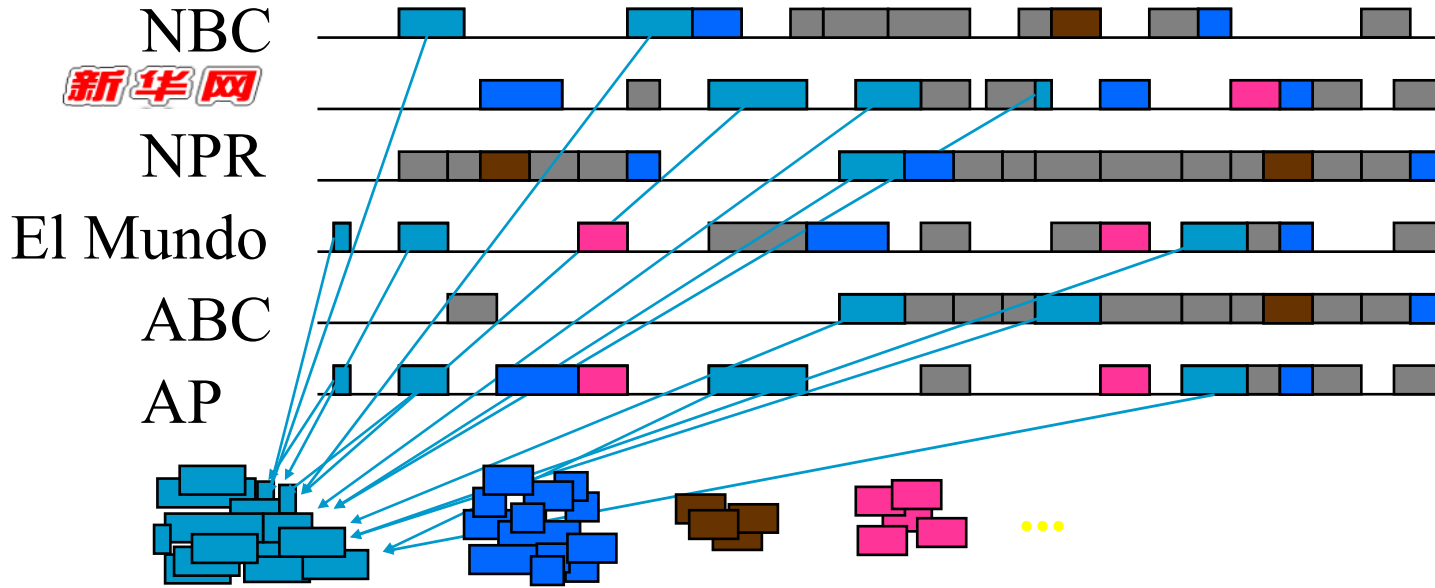


# İlk Hikaye Algılama



Sisteme yeni bir hikaye ulaştığında bunun tartıştığı konunun daha önceden tanımlanmış konularla ilgili değilse, yeni bir konu olduğunun belirlenmesi.

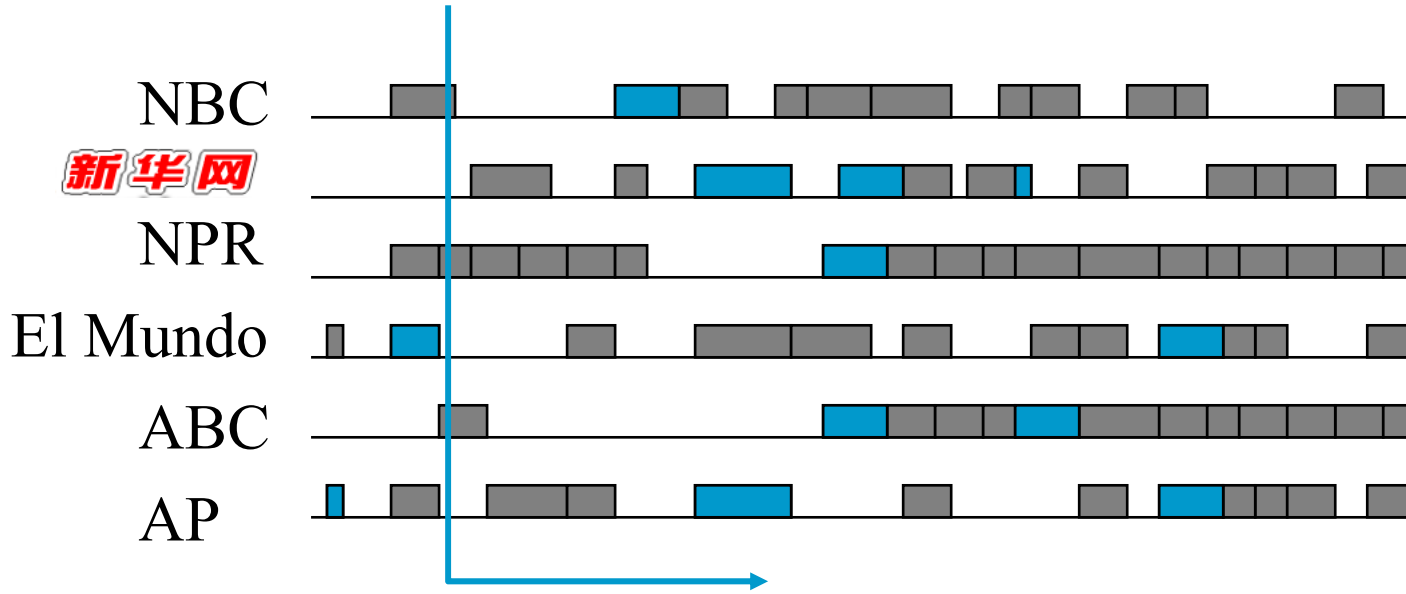
# Küme Algılama



Gelen hikaye bir ilk hikaye değilse ilgili kümeye yerleştirir, eğer ilk hikaye ise bunun için yeni bir küme oluşturur (özellikler vektörünü kullanır).

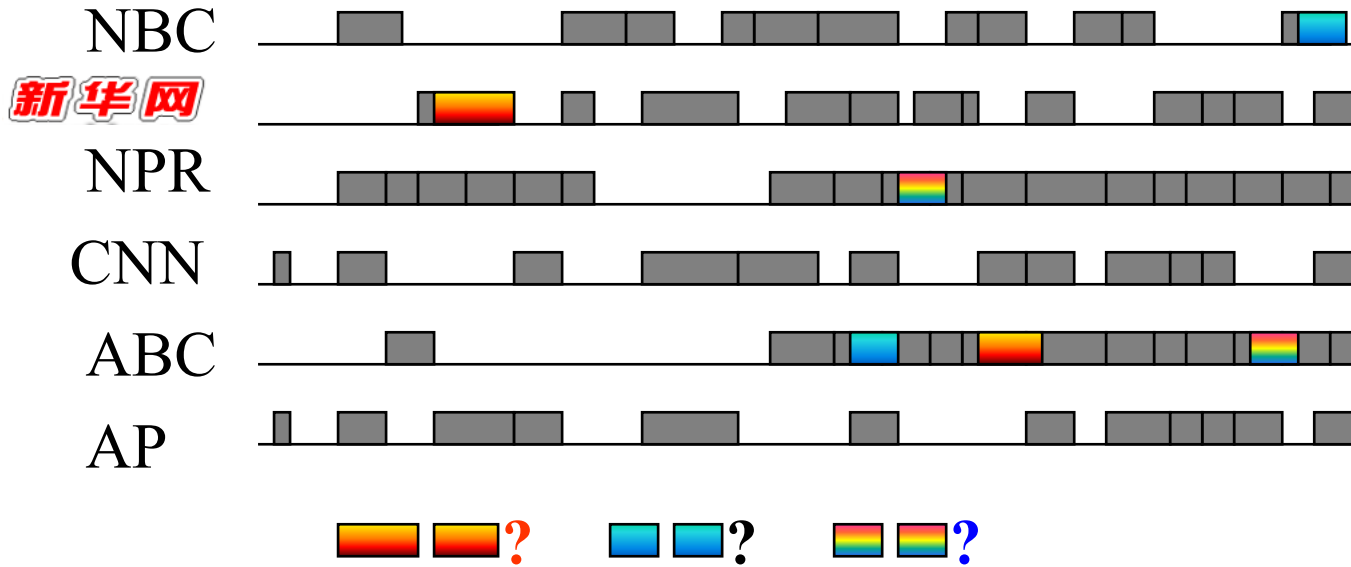


# Haber İzleme



Haber kaynaklarından sisteme gelen haberleri değerlendirilerek, bu haberlerin daha önceden belirlenmiş olan konularla ilgili olup olmadıklarını araştırır.

# Hikaye Bağlantı Algılama



Sisteme ulaşan iki farklı hikayenin aynı konuyu tartışıp tartışmadıklarını anlamayı amaçlar.

# Bilgi Eriřim Sistemleri II

Hayri Sever  
Bilgisayar Mühendisliđi Bölümü  
Bařkent Üniversitesi  
06530 Bađlıca, Ankara  
[sever@baskent.edu.tr](mailto:sever@baskent.edu.tr)

Yařar Tonta  
Bilge ve Belge Yönetimi  
Hacettepe Üniversitesi  
06532 Beytepe, Ankara  
[tonta@hacettepe.edu.tr](mailto:tonta@hacettepe.edu.tr)