

Web Madenciliği

22 Şubat 2017

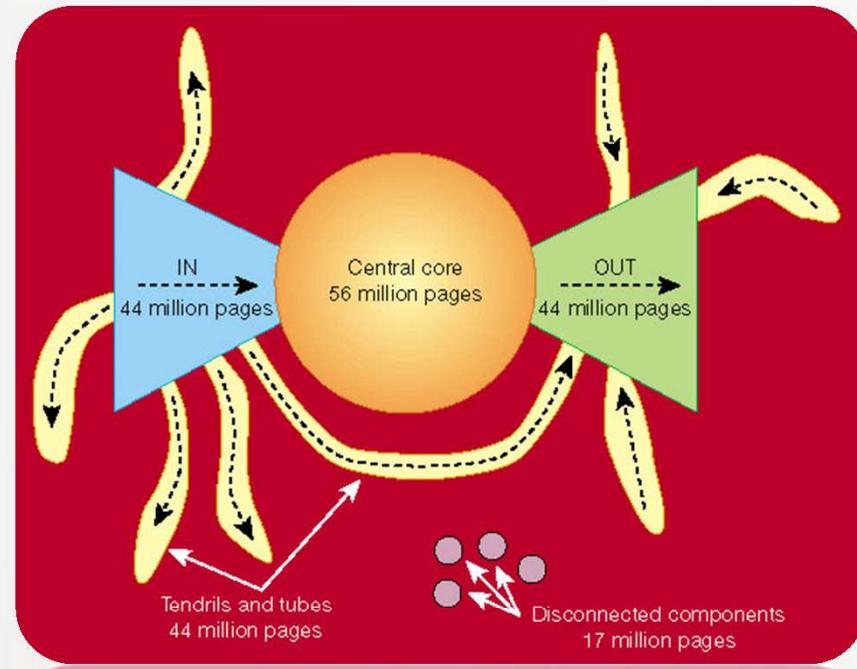
Dr. Mehmet Siddık Aktaş

Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü

World Wide Web (Web, WWW, W3)



- ❑ CERN (the European Organization for Nuclear Research)
 - Tim Berners-Lee
- ❑ GUIs
 - Berners-Lee (1990) (HTML, HyperText Transfer Protocol – HTTP, Web Browser), HTTP Web Server
 - Erwise and Viola(1992), Midas (1993) (Initial GUI based Browsers)
- ❑ Mosaic (1993)
 - National Center for Supercomputing Applications (NCSA)
 - a hypertext GUI for the X-window system
 - HTML: markup language for rendering hypertext
 - HTTP: hypertext transport protocol for sending HTML and other data over the Internet

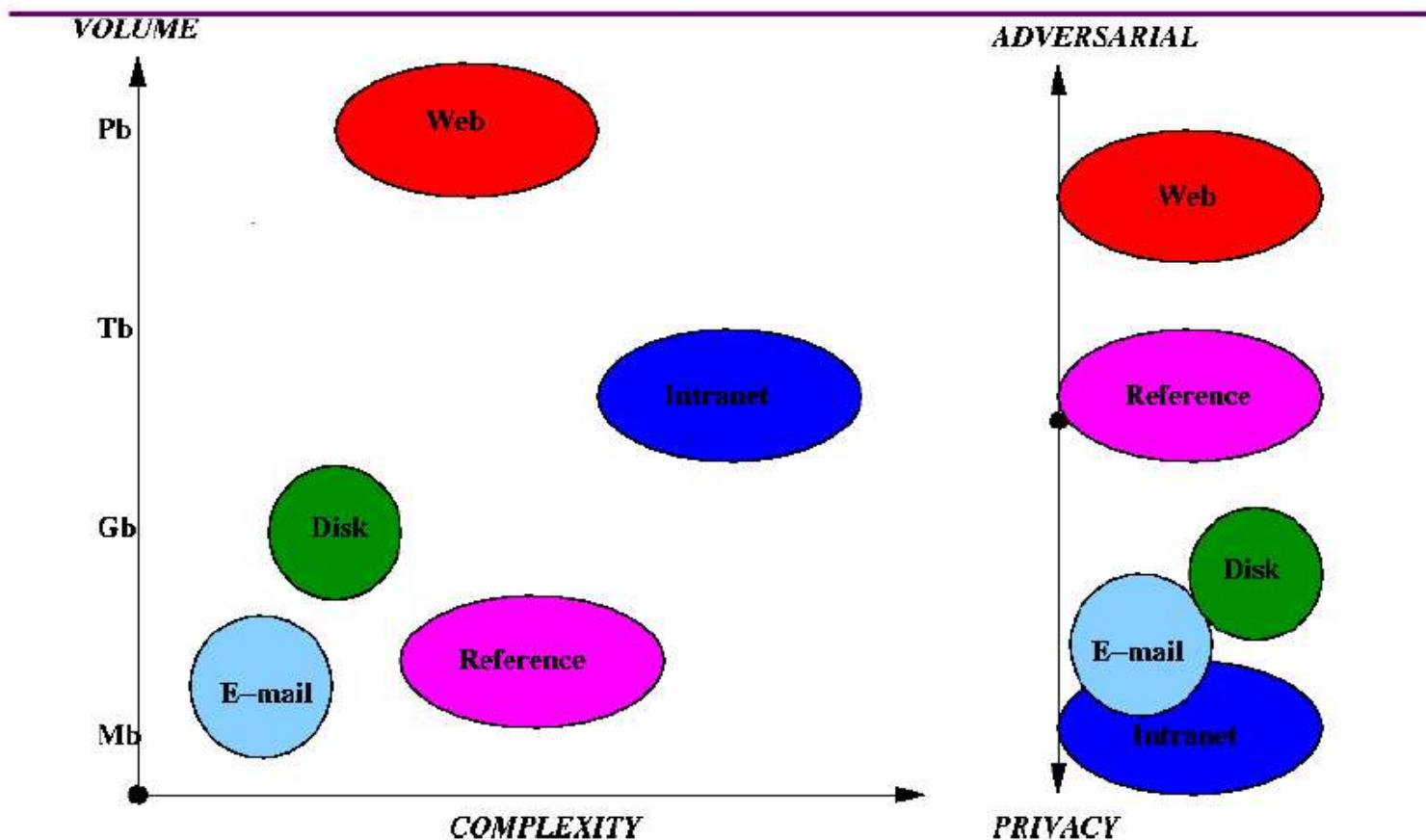


The Web is a bow tie

- Hypertext dökümanları
 - Text, Multimedia
 - Links
- Farklı doküman tipleri
 - PDF, Word, Excel ...
- Web
 - Milyarlar mertebesinde döküman,
 - Milyonlar mertebesinde farklı yazar var.
 - Herhangi bir merkezi editor yok.
 - Milyonlarca dağıtık bilgisayar üzerinde çalışıyor.
 - Farklı iletişim araçları kullanılarak ulaşılabiliriyor.
 - Internet üzerinde çalışıyor.

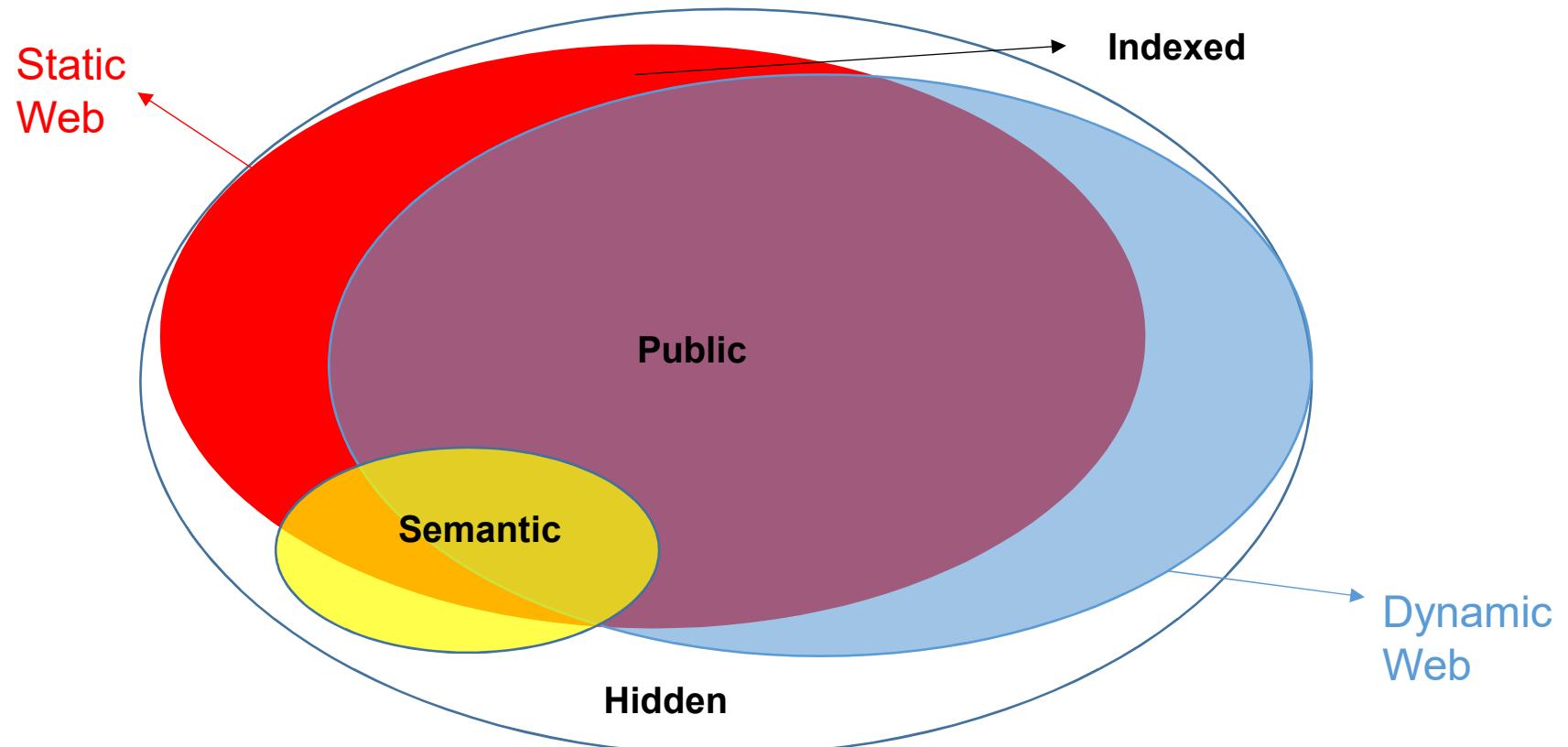


Different Views on Data





The Different Facets of the Web



<http://ijcai-11.iiia.csic.es/files/proceedings/T16-Web-Mining.pdf>

Jean-Loup Guillaume and Matthieu Latapy

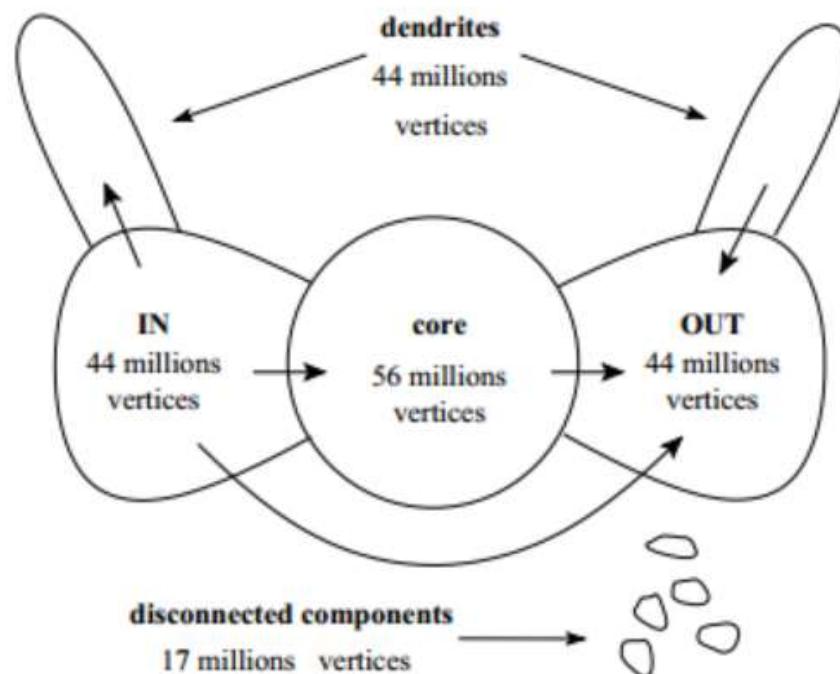
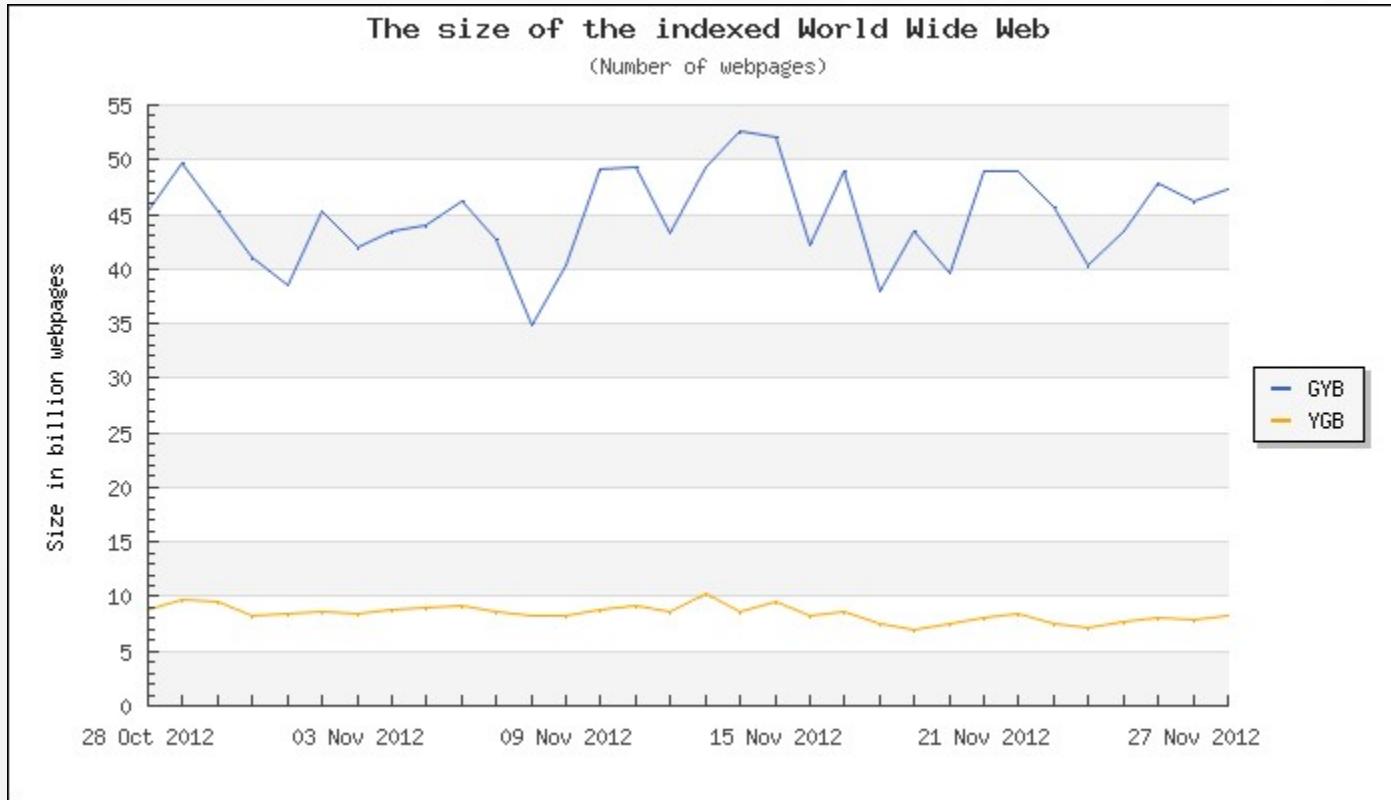


Fig. 3: The bow-tie macroscopic structure of the Web graph [BKM⁺ 00]: the core, the IN component, the OUT component and the dendrites. Each of these parts contains around one quarter of the pages, the disconnected part being reduced to less than 10% of the whole.

Web Ne Kadar Büyüк?

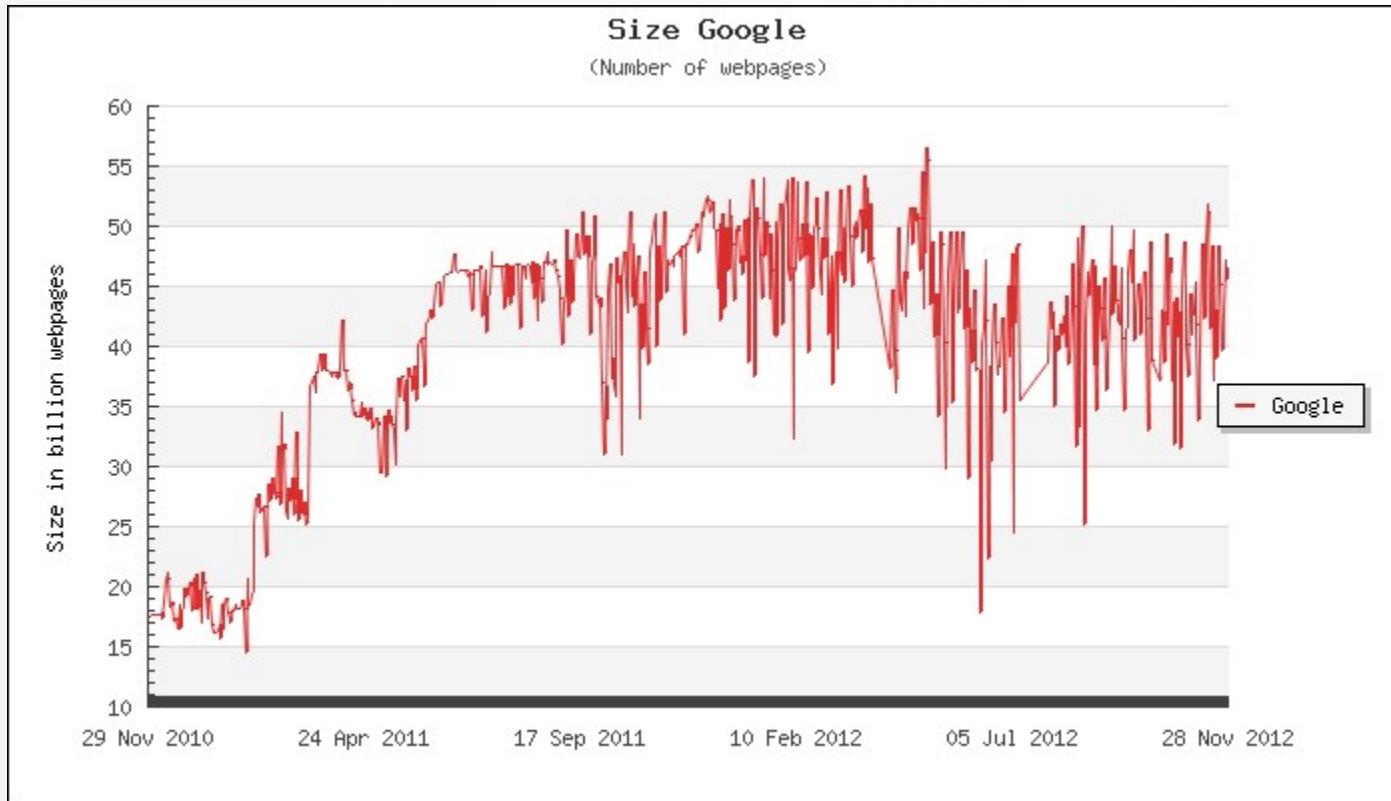


Web Ne Kadar Büyüк?

Google, Yahoo ve Bing indexlerinin tahmini büyüklüğüne göre Web en azından 8.34 milyar sayfadan oluşuyor.

www.worldwidewebsize.com 28 Kasım 2012

Web Ne Hızla Büyüyor?



Web Ne Hızla Büyüyor?

Google indexlerinin tahmini büyüklüğüne göre %30
büyümüş. 18-48 milyar Web sayfası artışı.

Kaynak: www.worldwidewebsize.com 28 Kasım 2012

Web Madenciliği

- Web sayfaları
- Erişim Log dosyaları
- Kullanıcı kayıt bilgileri
- Oturum ve hareket bilgileri
- Site yapısı ve içeriği

□ Web Madenciliği

- Çeşitli yapıdaki web sayfası dokümanlarını, içeriklerini, link yapılarını ve kullanım bilgilerini incelemek, bunlardaki anlamlı bilgileri keşfetmek için veri madenciliği tekniklerinin kullanılmasıdır.

- ❑ Web madenciliği kullanılan verilerin yapısına göre 3 gruba ayrılır.
 - ❑ Web içerik madenciliği
 - ❑ Text ve multimedia madenciliği
 - ❑ Web yapı madenciliği
 - ❑ Link analizi, grafik madenciliği
 - ❑ Web kullanım madenciliği
 - ❑ Log analizi, sorgu madenciliği
- ❑ Web madenciliği teknikleri beraber kullanılabilir.
 - ❑ Web' in karakteristik özellikleri
 - ❑ Uygulamalar (Ör: Arama motoru)

Dinamik



What for?

- The Web as an object
- User-driven Web design
- Improving Web applications
- Social mining
-



The Wisdom of Crowds

- James Surowiecki, a *New Yorker* columnist, published this book in 2004
 - “Under the **right** circumstances, groups are remarkably intelligent”
- Importance of diversity, independence and decentralization

Aggregating data

“large groups of people are smarter than an elite few, no matter how brilliant—they are better at solving problems, fostering innovation, coming to wise decisions, even predicting the future”.

Photos: [Explore Flickr](#) • [Learn More](#)

flickr^{BETA}

Tags / jaguar / clusters

jaguar

[SEARCH](#)

(Or, try an [advanced search](#).)



[car](#), [cars](#), [auto](#), [etype](#), [automobile](#), [classic](#), [vintage](#), [autoshow](#), [red](#), [show](#)

→ [See more in this cluster...](#)



[zoo](#), [animal](#), [cat](#), [animals](#), [bigcat](#), [seattle](#), [woodlandparkzoo](#), [sleep](#), [edinburgh](#), [caged](#)

→ [See more in this cluster...](#)



[guitar](#), [fender](#)

→ [See more in this cluster...](#)



[aircraft](#), [raf](#)

→ [See more in this cluster...](#)

These are the most recent photos tagged with jaguar. [See more...](#)



Flickr: Geo-tagged pictures

flickr®

No has iniciado sesión

Iniciar sesión

Ayuda

Inicio

La visita

Crear cuenta

Explorar

Buscar un lugar

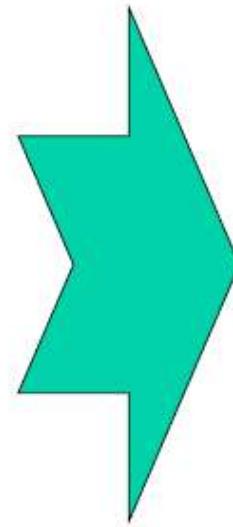
Buscar





The Wisdom of Crowds

- Popularity
- Diversity
- Quality
- Coverage



Long tail



The Long Tail

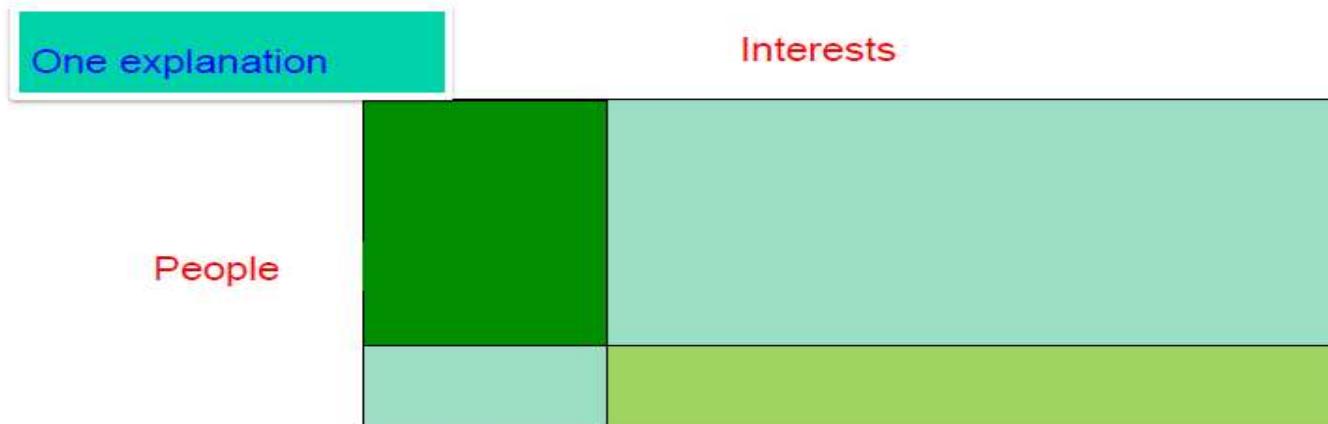
Explore Flickr through tags

architecture art australia beach birthday blue bw **california** canada
canon china christmas city concert england europe **family** festival flower
flowers food **france** friends fun germany green italy **japan** london
music **nature** new newyork night **nikon** nyc paris park **party**
people portrait red sanfrancisco sky snow spain street **summer** sunset taiwan
travel trip uk usa vacation water **wedding** white winter



Heavy tail of user interests

- Many queries, each asked very few times, make up a large fraction of all queries
 - Movies watched, blogs read, words used ...



<http://ijcai-11.iiia.csic.es/files/proceedings/T16-Web-Mining.pdf>



The Wisdom of Crowds

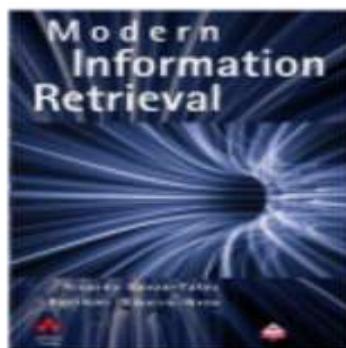
- Crucial for Search Ranking
- Text: Web Writers & Editors
 - not only for the Web!
- Links: Web Publishers
- Tags: Web Taggers
- Queries: All Web Users!
 - Queries and actions (or no action!)

Web İçerik Madenciliği nedir?

- Web kaynaklarının içeriklerinden yararlı bilgiyi elde etmek olarak tanımlanabilir.
- Web sitelerinin dokümanlarındaki terimleri ve link'leri bularak sayfanın indekslenmesi ve web sitesinin yapısı ortaya çıkarılmaya çalışılır.



What is Information Retrieval (IR)?



IR: Part of computer science which studies the **retrieval of information (not data)** from a collection of **written documents**. The retrieved documents aim at satisfying a **user information need** usually expressed in **natural language**.

- **Documents, unstructured, text, large**
- **Information need**
- **Store, search, find**
- **The World Wide Web?**
- **Relational databases?**

DIKW

- **Data:** Raw web pages
 - **Information:** Result of query
 - **Knowledge:** Result of processing query result by user
 - **Wisdom:** Synthesis of many such actions by a set of users
-
- One possible classification of steps in process



Information Retrieval vs. Databases

Information retrieval	Data retrieval
Retrieve all objects relevant to some information need	Retrieve all objects satisfying some clearly defined conditions
Find all documents about the topic "semantic web"!	SELECT id FROM document WHERE title LIKE "%semantic web%"
Result list	Well-defined result set

The screenshot shows a Google search results page for the query "semantic web". The results include links to the Semantic Web Wikipedia page, the W3C Semantic Web Activity statement, and other related resources.

```
[selke@tbdb ~]$ db2 "SELECT id FROM document WHERE title LIKE "%semantic web%" FETCH FIRST 3 ROWS ONLY"
```

```
ID
```

```
-----  
45489  
9635899  
98556
```

```
3 record(s) selected.
```

4

<http://www.ifis.cs.tu-bs.de/teaching/ss-11/irws>



Web Search

- Very similar to information retrieval
- Main differences:
 - Links between Web pages can be exploited
 - Collecting, storing, and updating documents is more difficult
 - Usually, the number of users is very large
 - Spam is a problem



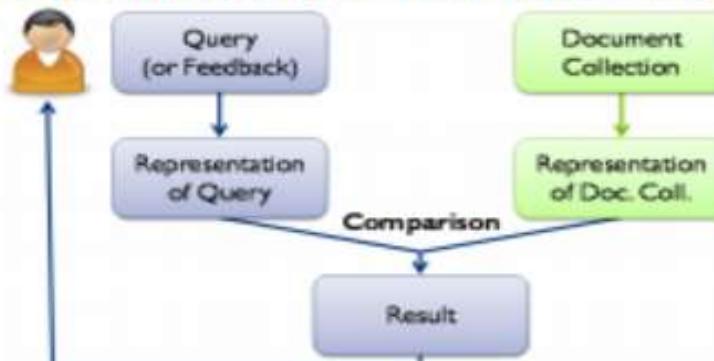
5

<http://www.ifis.cs.tu-bs.de/teaching/ss-11/irws>



IR Models

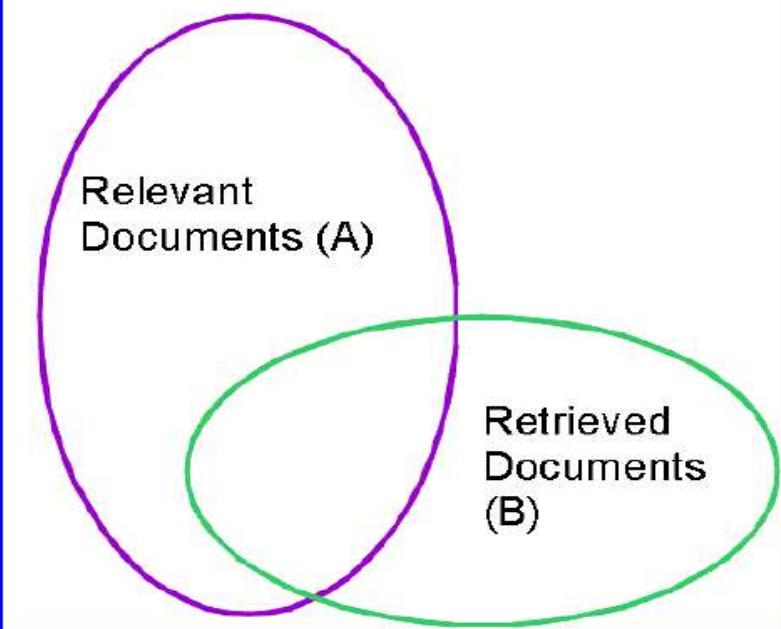
- Any IR system is based on an **IR model**
- The model defines ...
 - ... a **query language**,
 - ... an internal **representation of queries**,
 - ... an internal **representation of documents**,
 - ... a **ranking function** which associates a real number with each query–document pair.
- Optional: A mechanism for **relevance feedback**



38

“Classic” Information Retrieval

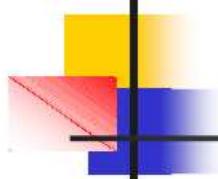
The Collection



- Given a query, the system retrieves a set B of documents
- Every retrieved document is either relevant or irrelevant to the query

Quality metrics:

- Recall : $(A \cap B) / A$
- Precision : $(A \cap B) / B$



Recall and Precision on the Web

- Relevance of document to queries is not binary – there are many shades of gray
- Broad-topic queries:
 - abundance problem
 - Precision is the dominating factor: users mostly satisfied with a few good results (a few *authoritative* pages)
- Narrow-topic queries:
 - Find a needle in an enormous haystack
 - Recall demands engines cover significant portions of the Web
- Common measure: precision@10
- Nowadays larger emphasis on *diversity*
 - Positive recall for many aspects of the query

New Google Update "Improves Diversity" Of Domains In Results

Sep 14, 2012 at 5:03pm ET by [Danny Sullivan](#)



There's been [chatter](#) about Google having done some type of algorithm update this week that's having an impact on rankings. Now it's finally confirmed, a new change meant to allow more domains to appear in search results.

The head of Google's spam fighting team Matt Cutts [tweeted](#) this about it:

Just fyi, we rolled out a small algo change this week that improves the diversity of search results in terms of different domains returned.

Lately, more and more people have been noting that Google's search results can sometimes be dominated by pages that all come from the same domain. In other words, do a search, and all the listings seem to come from the same web site.

Here's an example of this from last month, [when we wrote](#) about the problem:



christopher jagmin plates

About 5,570 results (0.20 seconds)

All from this domain

[home * christopher jagmin](#)
christopherjagmin.com/
Christoper Jagmin Design. Living well isn't about excess. A peanut butter and jelly sandwich tastes better on a beautiful **plate**. An escapist novel is more ...
SHOP - Stores - Press - Custom

[shop * christopher jagmin](#)
christopherjagmin.com/shop/
Odd Number Dinner **Plates** Four 10.5" **plates**. Even Number Dinner Plates Four 10.5" **plates**. Odd Little Number **Plates** Four 7.5" **plates**. Even Little Number ...

[buy > Little Silhouette Plates * christopher jagmin](#)
www.christopherjagmin.com/shop/buy.phtml?id=31
These four 7 1/2 inch porcelain **plates** are very friendly, and can also be used for salads, for bread, for snacks, or just for fun. They would be very happy to hang ...

[buy > Odd Poco Number Plates * christopher jagmin](#)
www.christopherjagmin.com/shop/buy.phtml?id=17
Odd Poco Number **Plates**. ... No need for a passport with these little 7.5" Spanish **plates**. Perfect for tapas, tacos, with a little guacamole on top! 4 **plates** to a set ...

Diversity of Search Results

Slide 30

MA1

Mehmet Aktas, 12/15/2013



The Bag of Words Representation

- A very popular **representation of documents** is the **bag of words model**
 - Each document is represented by a bag (= multiset) of **terms** from a predefined **vocabulary**
 - Standard case:
 - Vocabulary
= set of all the words occurring in the collection's documents
 - Each document is represented by the words it contains



That's one small step for a man,
a giant leap for mankind

→ { that's, one, small, step,
for (2), a (2), man, giant,
leap, mankind }

<http://www.ifis.cs.tu-bs.de/teaching/ss-11/irws>



The Bag of Words Model (2)

- Cons:
 - Word order gets lost
 - Very different documents could have similar representations
 - Document structure (e.g. headings) and metadata is ignored
- Pros:
 - Simple set-theoretic representation of documents
 - Efficient storage and retrieval of individual terms
 - IR models using the bag of words representation work well!





The Bag of Words Model (3)

- Any document can be represented by an **incidence vector**:

vocabulary (aka index terms) →

that's one small step for a man giant leap mankind taikonaut Zhai's is China

That's one small step for a man,
a giant leap for mankind

Taikonaut Zhai's small step is a
giant leap for China

1 1 1 1 2 2 1 1 1 0 0 0 0

incidence matrix
(aka term-document matrix)

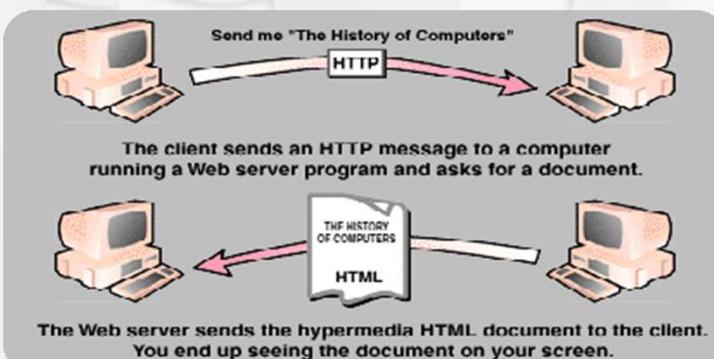
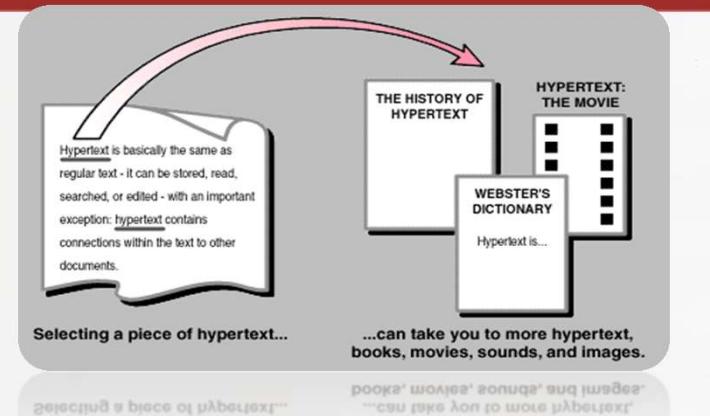
0 0 1 1 0 1 0 1 1 0 1 1 1

❑ Web Yapı Madenciliği nedir?

- ❑ Web yapı madenciliğin amacı web sayfaları arasındaki linkleri takip ederek bilgi üretmektir.
- ❑ Web yapı madenciliği, yapısal veri tiplerine göre 2 ye ayrılır.
 - ❑ Hyperlink bir web sayfasını farklı bir lokasyona yönlendiren yapısal eleman olduğu için webdeki hyperlinklerin modelinin çıkarılmasıdır,
 - ❑ Web sayfası dokümanlarındaki HTML ya da XML etiketleri analiz ve tanımlarında ağaç (tree) benzeri yapıların kullanılmasıdır.

- Internet – Web
- Link - Hyperlink
- Hypertext - Hypermedia
- HTTP - HTML
- Anchor Text - URL

Karıştırılan Web Kavramları

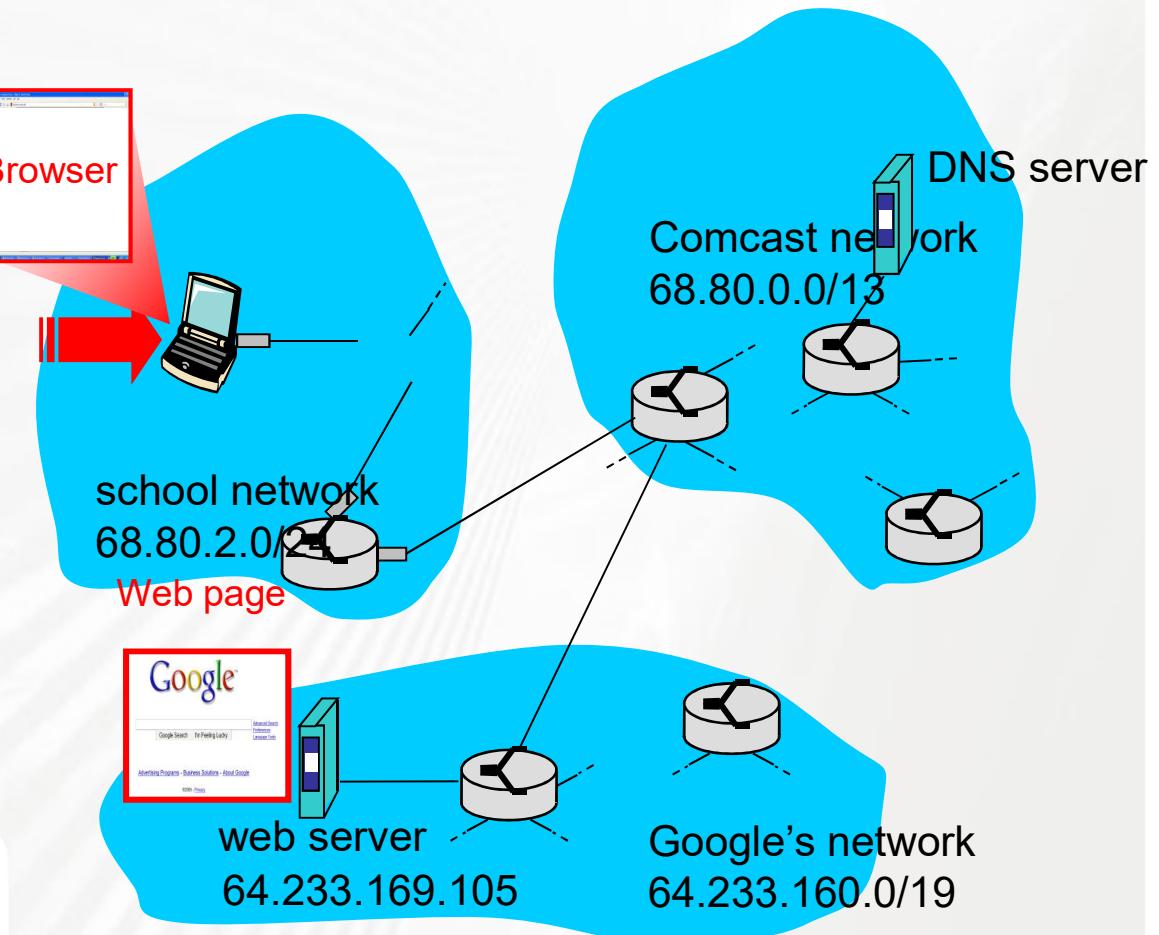


A **hyperlink** is a link you can click on or activate with the keyboard or other device in order to go somewhere else. The hyperlinks are used on the pages and they navigate from one page to another between same or different Sites/servers.

A **URL/Link** can be thought of as the "address" of a web page and is sometimes referred to informally as a "web address."

Like if we write "<http://google.com>" in the address bar, it is a link and if we apply it on the page so that, anyone can navigate from the site to Google, it is called **hyperlink**.

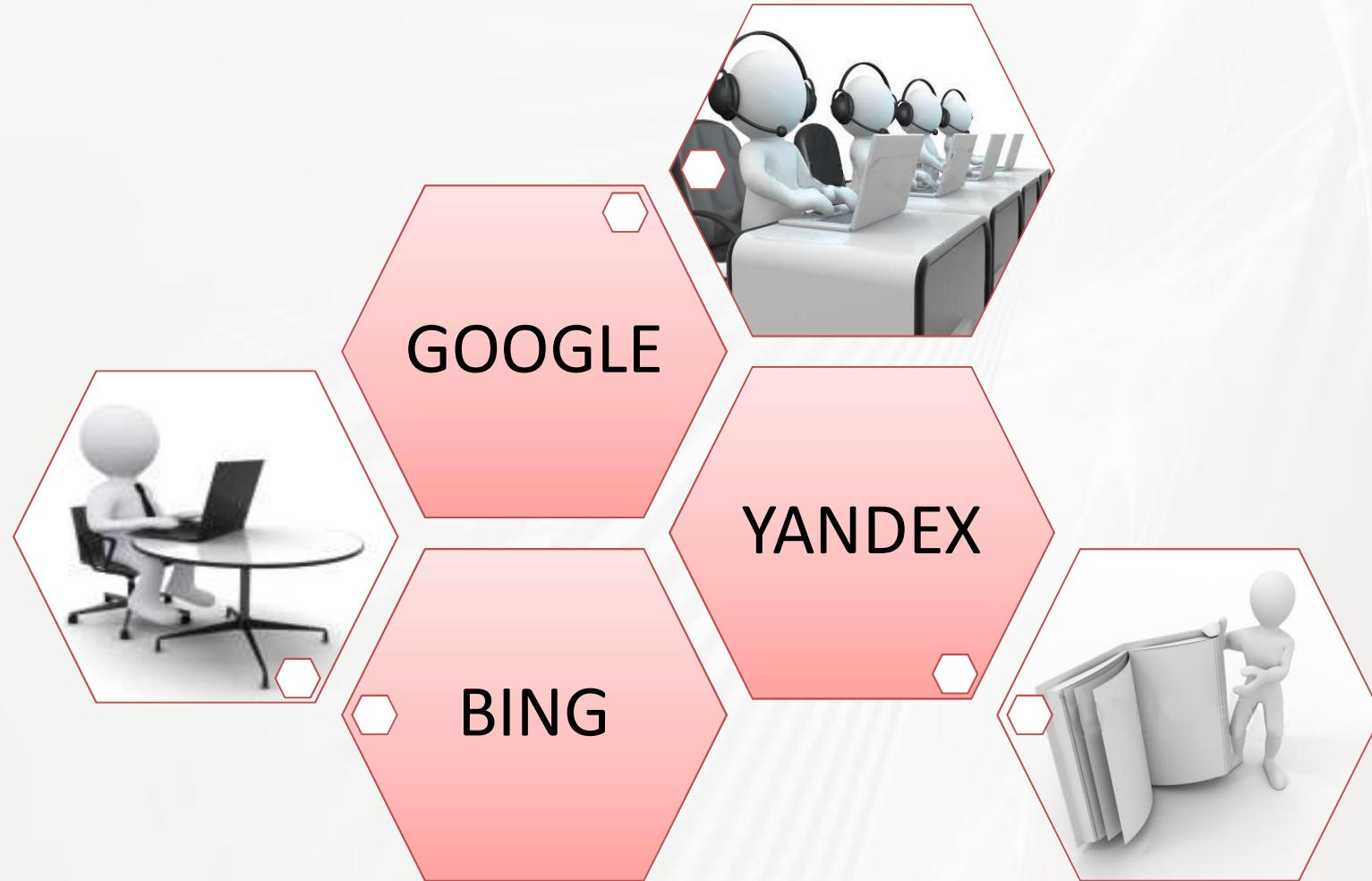
(.



Anchor text usually gives the user relevant descriptive or contextual information about the content of the link's destination. The anchor text may or may not be related to the actual text of the URL of the link. the English-language Wikipedia's homepage might take this form:

```
<a href="http://en.wikipedia.org/wiki/Main_Page">Wikipedia</a>
```

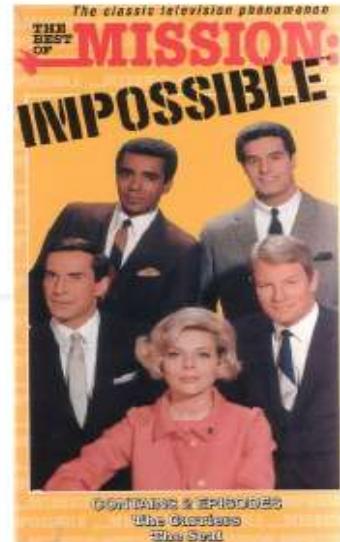
The anchor text in this example is "Wikipedia", the unwieldy URL http://en.wikipedia.org/wiki/Main_Page displays on the web page as Wikipedia, contributing to clean, easy-to-read text.



Mission Impossible?

Search engines:

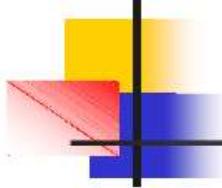
- Crawl and index **tens of billions** of documents
- Answer **hundreds of millions** of queries per day
- Devote less than **1 second** of processing time per query execution



Users:

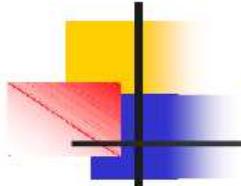
- Submit **very short** queries (averaging about 2.6 terms)
- Expect to receive **the most relevant** results on the Web
- In a **blink of an eye**

In terms of 1990 Information Retrieval research - almost unimaginable!



The Static Web “Corpus”

- **Bulk** 10s of billions of “static” pages; doubling every 8-24 months
- **Lack of stability**..... Estimates vary widely (half life of page is several months)
- **Lexicon Size** 10s-100s of millions of words
- **Heterogeneity**
 - Type of documents Text, pictures, audio, video, scripts,...
 - Quality From junk to the best content on earth
 - Authorship From masterfully edited content to user-generated content
 - Languages 100+
- **Duplication**
 - Syntactic..... 30%-40% (near) duplicates
 - Semantic..... ??
- **Complex graph topology**..... ≥ 8 links/page in the average; bow-tie structure
- **Spam**..... 100s of millions of pages



Web Searchers - Observations

- Make ill defined queries
 - Short (2.54 terms average, 80% contain less than 3 words)
 - Use imprecise (and often misspelled) terms
 - Unfamiliar with query syntax (80% queries without operator)
- Wide variance in information needs, expectations, education/knowledge, screen sizes, IP bandwidth, patience
 - Different modalities (mobile, desktop) = different needs and expectations, even with same person
- Specific behavior (desktop and laptop)
 - 85% look **over one result screen only** (mostly "above the fold")
 - 78% of queries are not modified
- Overall, we as users are investing low cognitive effort per query (formulating and looking at results)



Millî arama motoru geliyor

Türkiye, kendi arama motorunu yapacak. TÜBİTAK mühendislerinin geliştireceği arama motoru, Google ve Yandex'e rakip olacak. İsmi konulmayan arama motorunda Türkçe karakterlerle arama gerçekleştirilebilecek.

12 Ekim 2012 Cuma 12:26

Print A A

Like 0

Tweetie 0

+1 0

Share

Türk mühendisler, Amerikalı Google ve Rus Yandex'e karşı yeni bir arama motoru için düğmeye bastı. Ulaştırma Bakanlığı'nın bütçesini ayırdığı yerli arama motoruna Sanayi Bakanlığı da destek verecek. İki yıl içinde faaliyete geçmesi planlanan proje ile memurlar ve tablet ile eğitim gören öğrenciler bilgilere anında ulaşabilecek. Teknolojiye yönelik yatırımlar artarken Türkiye, kendi arama motorunu yapmak için harekete geçti. ABD'li Google ve Rus Yandex'e Türk mühendisler tarafından geliştirilen arama motoru ile rakip olunacak. TÜBİTAK mühendislerinin yüzde yüz yerli yazılımla hizmete sunacakları arama motoru için beş yıllık bir eylem planı hazırlandı. Jlaştırma, Denizcilik ve Haberleşme Bakanlığı'nın bütçesini ayırdığı yerli arama motoru için Bilim, Sanayi ve Teknoloji Bakanlığı da destek verecek. TÜBİTAK uzmanlarının üzerinde çalıştığı projenin ilk aşaması iki yıl içerisinde faaliyete geçirilecek. e-devlet çerçevesinde kamu kurumları yerli arama motorunu kullanabilecek.

Kamunun ihtiyaç duyduğu tüm aramalar yüzde 100 Türkçe karakterler ile gerçekleştirilebilecek. Projenin ikinci ayağında ise TÜBİTAK'ın arama motoru Fatih Projesi kapsamında öğrenciler tarafından da kullanılabilecek. Öğrenciler dağıtılacek tablet bilgisayarlar üzerinden TÜBİTAK tarafından geliştirilen yerli arama motoru ile istedikleri bilgiye erişebilecek. Özellikle eğitim ve sağlık alanlarındaki aramalar yapılabilecek. Üçüncü aşamada ise hem Türkiye genelinde hem de dünya genelinde hızla kullanımının yaygınlaşması planlanıyor.

Google Search: spears

Web Images Groups News Froogle more » spears Search Advanced Search Preferences

Web Results 1 - 10 of 1,000,000

News results for spears - View today's top stories

[Knee Injury Closes Spears' Onyx Hotel](#) - Billboard - 1 hour ago
[Britney Spears' tour is canceled](#) - San Diego Union Tribune - 7 hours ago
As fall approaches, Spears may start to smell Curious - Houston Chronicle - Jun 14, 2004

Britney Spears :: The Official Web Site
The Official Web Site of Britney Spears. Your official source for all things Britney. ...
Remember, proceeds benefit the Britney Spears Foundation. ...
www.britneyspears.com/ - 41k - Jun 14, 2004 - [Cached](#) - [Similar pages](#)

Britney Spears - britney.com - Jive Records
iTunes. Real/Rhapsody. Napster. Under 11.
www.britney.com/ - 10k - [Cached](#) - [Similar pages](#)

Britney Spears Portal - pics, lyrics, MP3s and more!
Britney Spears pics, lyrics, MP3s, news, gossip, fan sites, forums, and much more!
Britney Spears Portal, ...); Britney Spears Portal, ...
www.britney-spears-portal.com/ - 25k - [Cached](#) - [Similar pages](#)

Britney Spears guide to Semiconductor Physics: semiconductor ...
Britney Spears lectures on semiconductor physics, radiative and non-radiative transitions,
edge emitting lasers and VCSELs. ...
britneyspears.ac/lasers.htm - 13k - [Cached](#) - [Similar pages](#)

BritneySpears.org: Your online guide to Britney!
A comprehensive Britney Spears fansite which pays tribute to Britney with the most active
message board, daily news, many pictures, desktop media and more. ...
www.britneyspears.org/ - 78k - Jun 14, 2004 - [Cached](#) - [Similar pages](#)

Britney-Spears.To You! - The Britney Spears Community
Britney Spears : biography, discography, musics, real, mp3, videos, pictures, clips,
guestbook, www board, free page, search engine, links and more. ...
www.britney-spears.to/ - 9k - [Cached](#) - [Similar pages](#)

The Mystery of Britney's Breasts
www.liquidgeneration.com/poptoons/britneys_breasts.asp - 2k - [Cached](#) - [Similar pages](#)

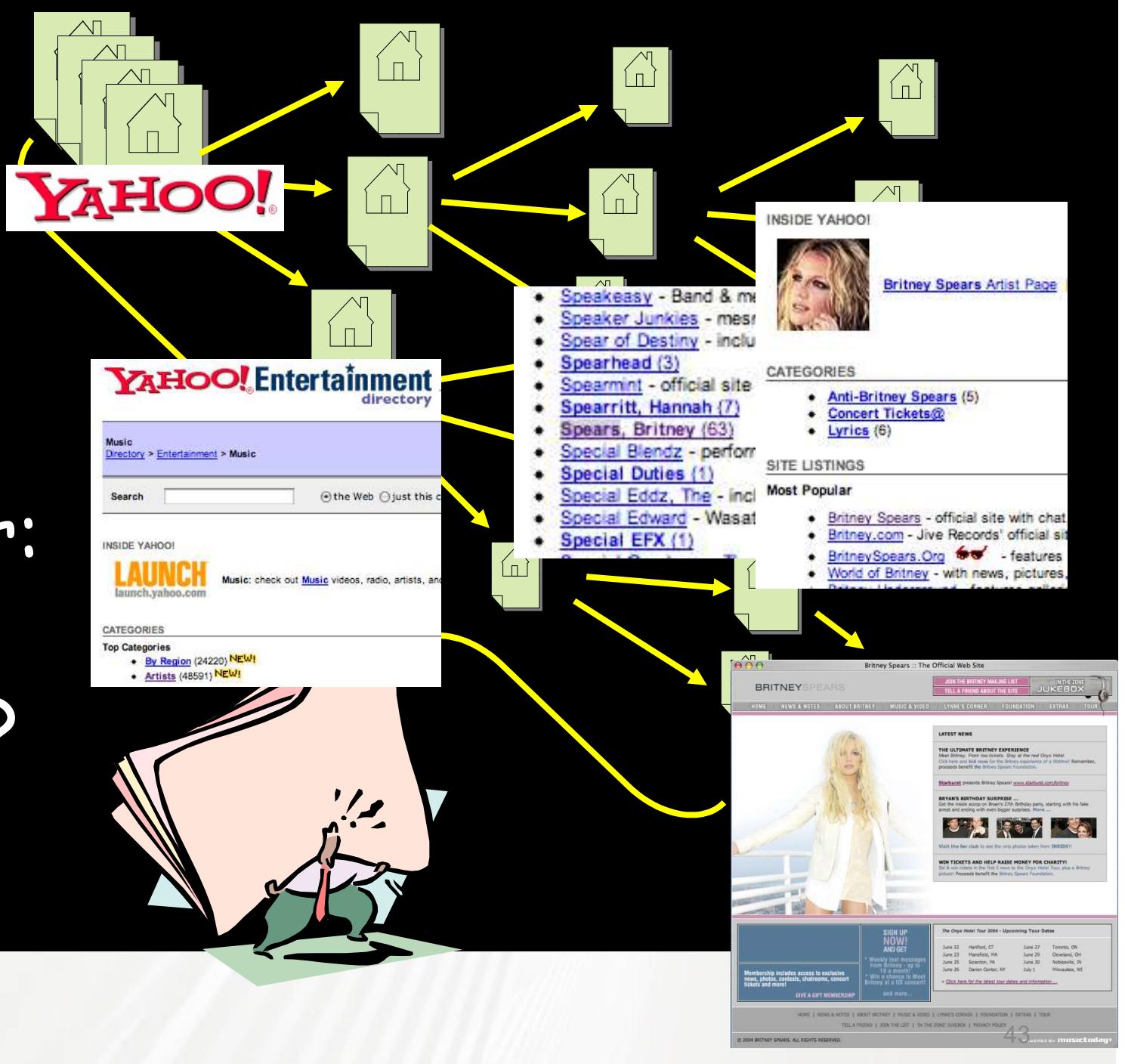
Britney Spears spelling correction
The data below shows some of the misspellings detected by our spelling correction system
for the query [britney spears], and the count of how many different ...
www.google.com/jobs/britney.html - 40k - [Cached](#) - [Similar pages](#)

Britney Spears pictures news music Britney Spears lyrics
Britney Spears pictures mp3 sites gallery photos images music fun games chat lyrics. ...
Britney Spears Forum Come see what is inside the Britney Spears forum! ...
www.britney-spears.com/ - 42k - Jun 14, 2004 - [Cached](#) - [Similar pages](#)

Britney Spears Zone - Your Guide to Britney Pictures and News
Britney Spears, Britney Spears, Britney Spears, ...
www.britneyzone.com/ - 101k - Jun 14, 2004 - [Cached](#) - [Similar pages](#)

Soru: Arama motoru, sonuçlarda listelenen tüm sayfalarda, bu kelimenin olduğunu nasıl biliyor?
Cevap: Tüm bu sayfalar daha önceden indirilip, indeksleniyor.

Crawler: Nasıl çalışır?

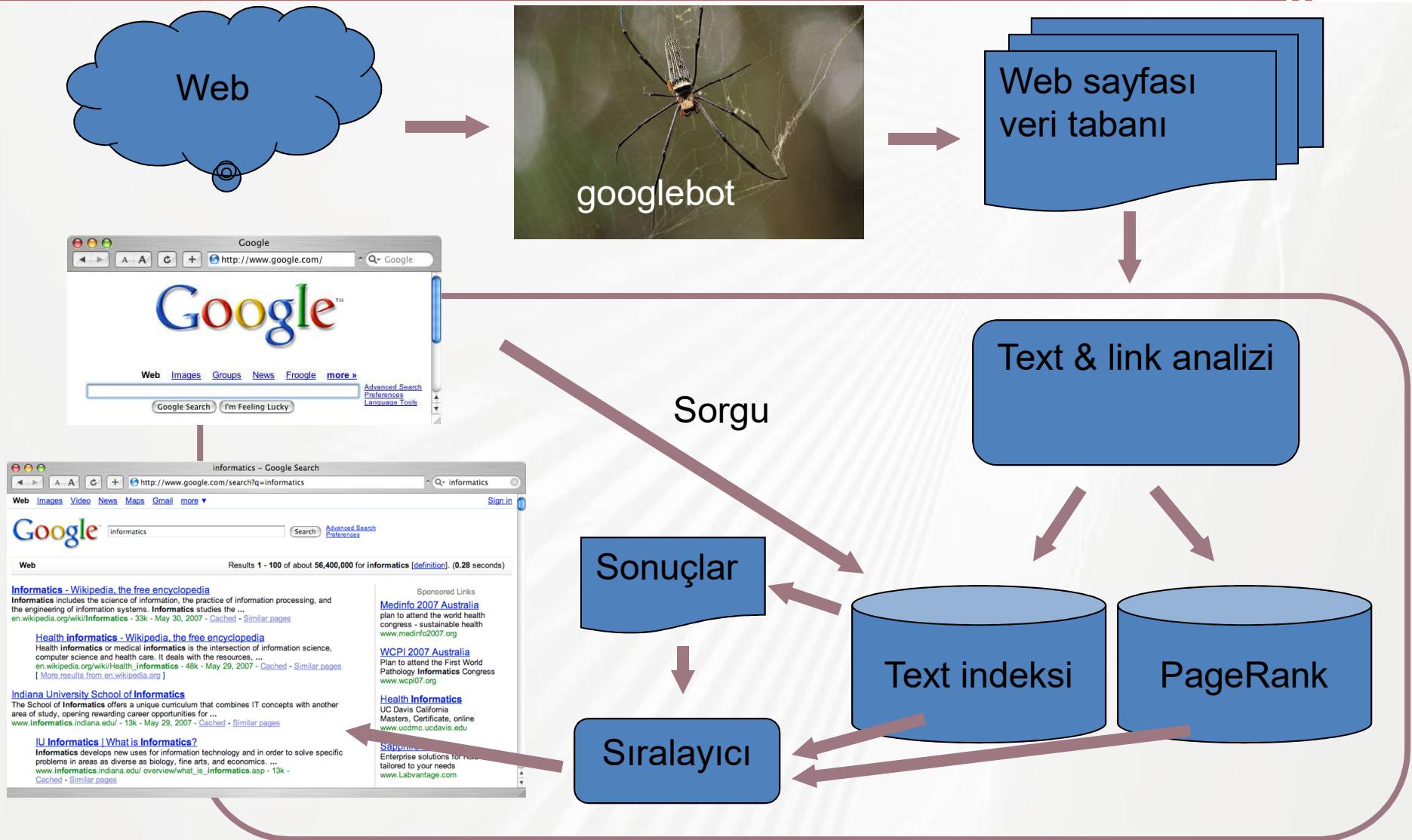


Crawler neden kullanılır?

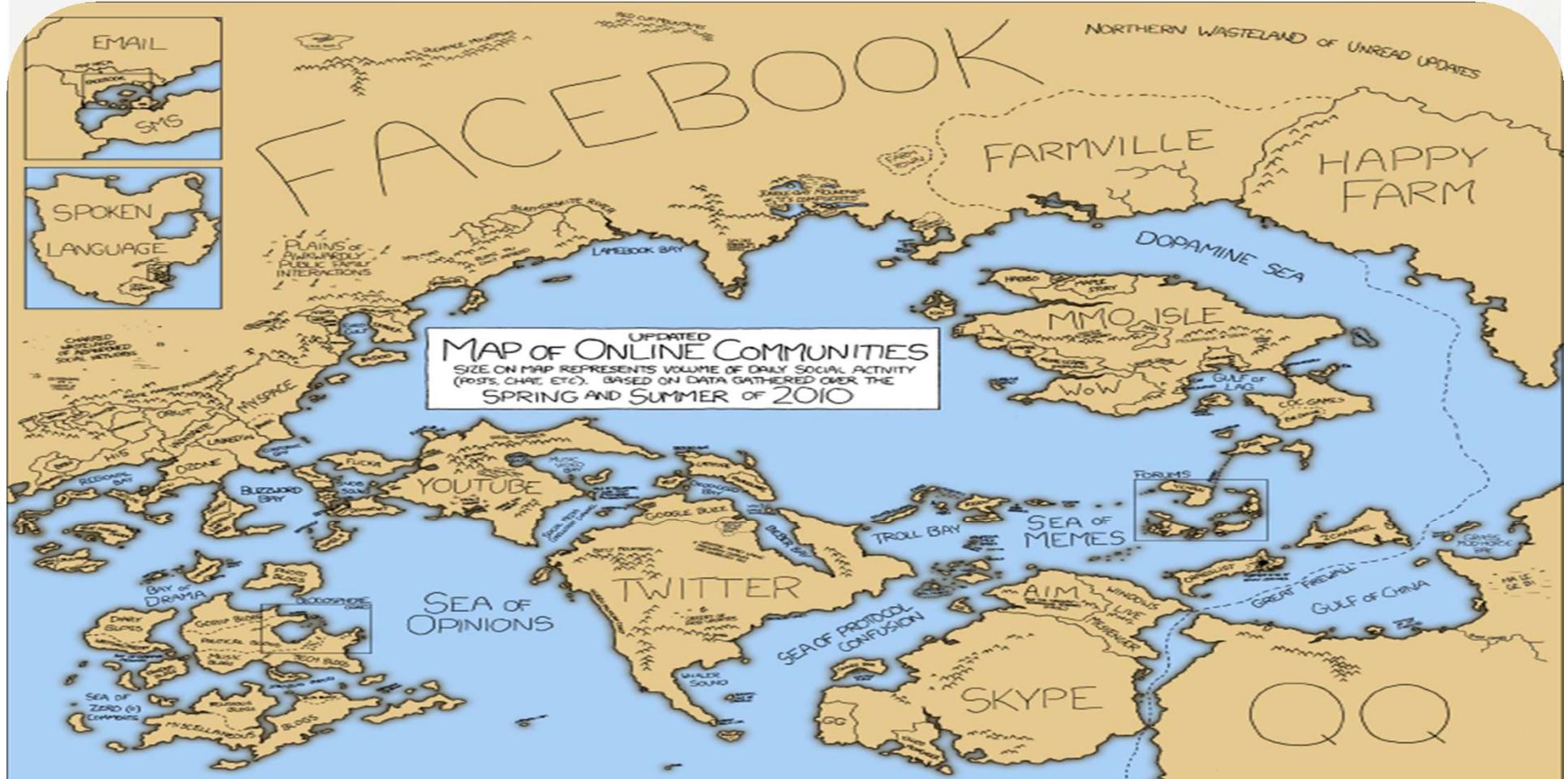


- Arama motorlarını desteklemek (Google, Bing, Yahoo, v.b.)
- Özelleştirilmiş arama motorları, örnekler: news, shopping, papers, recipes, v.b.
- İş mantığı: Potansiyel rakiplerin ya da iş ortaklarını takip edebilmek
- İlgilenilen Web sayfalarını izleyebilmek
- Kötü amaçlı kullanıcıları: e-posta adreslerinin “spamming”, “phishing” gibi kötü amaçlar için bulunması

Arama motoru içindeki Crawler



Online Communities



ABOUT THIS MAP

COMMUNITIES RISE AND FALL, AND TOTAL MEMBERSHIP NUMBERS ARE NO LONGER A GOOD MEASURE OF A COMMUNITY'S CURRENT SIZE AND HEALTH. THIS UPDATED MAP USES SIZE TO REPRESENT TOTAL SOCIAL ACTIVITY IN A COMMUNITY — THAT IS, HOW MUCH TALKING, PLAYING, SHARING, OR OTHER SOCIALIZING HAPPENS THERE. THIS MEANT SOME COMPARING OF APPLES AND ORANGES, BUT I DID MY BEST AND TRIED TO BE CONSISTENT.

ESTIMATES ARE BASED ON THE BEST NUMBERS I COULD FIND, BUT INVOLVED A GREAT DEAL OF GUESSWORK, STATISTICAL INFERENCE, RANDOM SAMPLING, NONRANDOM SAMPLING, A 20,000-CELL SPREADSHEET, EMAILING, CARDLING, TEA-LEAF READING, GOAT SACRIFICES, AND GUT INSTINCT (I.E. MAKING THINGS UP).

SOURCES OF DATA INCLUDE GOOGLE AND BING, WIKIPEDIA, ALICE, GIGABOX.COM, SPUNJELLY.COM, WORDPRESS, ANGSTAT, EVERY WEBSITE STATISTICS PAGE I COULD FIND, PRESS RELEASES, NEWS ARTICLES, AND INDIVIDUAL SITE EMPLOYEES. THANKS IN PARTICULAR TO FOLKS AT LAST.FM, LIVE.JOURNAL, REDDIT, AND THE NEW YORK TIMES, AS WELL AS SYSADMINIS AT A NUMBER OF SITES WHO SHARED STATISTICS ON CONDITION OF ANONYMITY.



❑ Web Kullanım Madenciliği nedir?

- ❑ Web Kullanım Madenciliğinde kullanılan veriler, web üzerindeki çeşitli sunucularda tutulan kullanıcı erişim hareketlerinin yer aldığı çeşitli log dosyalarından elde edilir.
 - ❑ İstemcilerden gelen her istek, bir kayıt olarak, metin tabanlı log dosyalarına eklenir.
 - ❑ Log dosyasındaki kayıt formatı verilen servis çeşidine ve kullanılan işletim sistemine göre faklılıklar gösterir.
 - ❑ Bu log dosyalarından bazıları, access log (erişim), mail log, error log, ftp log şeklindedir.

❑ Sonuç olarak;

- ❑ Web madenciliği günümüzde internetin yoğun bir şekilde kullanımının artması nedeniyle üzerinde önemli ölçüde araştırma yapılan bir alan haline gelmiştir.
- ❑ Web madenciliği, kullanıcıların web sitesindeki davranışlarını çeşitli kayıtlar üzerinden inceleyerek web sitelerinin yeniden tasarım yada geliştirilmesi konusunda ipuçları sunar.



Why Should I Know about All This?

Gartner

- “80% of business is conducted on **unstructured** information”
- “85% of all data stored is held in an **unstructured** format”
- “7 million **Web pages** are being added every day”

Butler Group
a Datamonitor Company

- “**Unstructured** data doubles every three months”

TEŞEKKÜRLER