



Outline



- What is required to understand the content on the Web?
- How web evolution has taken place?
- What are expectation from the Web?

- Johannes Gutenberg developed a movable type printing in 1447.
 - primary goal was to develop a mechanism to speed the printing
 - Result
 - books, journals, newspapers, scientific papers, etc....

- Tim Berners-Lee (with colleagues) created the World Wide Web in 1989.
 - independent contractor at CERN in Geneva (European Organization for Nuclear Research).
 - primary goal was to provide rapid, electronic access to the online technical reports and other documents created by the world's high-energy physics laboratories.
 - Result
 - fundamental change of human communication and social interaction.

- Collection of **interlinked hypertext documents** accessible using internet.
- The Web
 - **HTTP** (how to transfer data) GET/index.html
 - **URI** (how to address data) <http://www.uclab.khu.ac.kr/>
 - **HTML** (to mark up data for human reader) ..
<html><head><title>.....

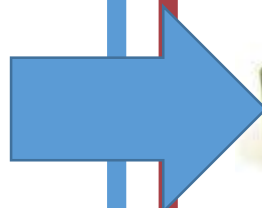


- Immensely successful.
- Huge amounts of data.
- Syntax standards for transfer of structured data.
- Human-readable documents.

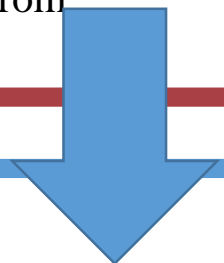
- Billions of diverse resources/documents online



Retrieving documents



Extracting relevant data from
retrieved documents



Combining information from different sources
to achieve a particular goal

Web Challenges



Too much information with **too little structure** and made for human consumption



Web resources are heterogeneous in terms of contents, in terms of structure, and in terms of character encoding



Humans can derive new (implicit) information from given pieces of information; machines are required to do the same

Understanding the Web

The screenshot shows the NHK Japan website interface. At the top, there's a navigation bar with the NHK logo and the text '北米どこでも24時間日本語チャンネル'. Below this, there are buttons for '視聴方法' (Viewing Method), '番組表' (Program Schedule), and 'カスタマーサービス' (Customer Service). A search bar is also present. The main content area is divided into sections for 'おすすめ番組' (Recommended Programs), '映画' (Movies), 'ドキュメンタリー' (Documentaries), 'バラエティ' (Variety), and '音楽' (Music). Each section lists specific programs and their details.

- What information is important and how do you know?
- What is news and what is commercial?
- What information is related by content?
- What does the information mean?

The (Document) Web for Humans

```
arrows.svg/44px-Merge-arrows.svg.png" width="44" height="18" /></a></td>
<td>Die Artikel <strong class="selflink">Klassifikator
(Informatik)</strong> und <a href="/wiki/Klassifikationsverfahren"
title="Klassifikationsverfahren">Klassifikationsverfahren</a>
überschneiden sich thematisch. Hilf mit, die Artikel besser voneinander
abzugrenzen oder zu vereinigen. Beteilige dich dazu an der <a
href="/wiki/Wikipedia:Redundanz/M%C3%A4rz_2009#Klassifikator_.28Informati
k.
20
Ba
hr
ti
hr
e"
title="Wikipedia:Redundanz/M%C3%A4rz_2009#Klassifikator_.28Informati
(CET)</small></td>
</tr>
</table>
<p>Ein <b>Klassifikator (Informatik)</b> sortiert Dokumente in Klassen
ein. Diese Methodik kommt vor allem im <a href="/wiki/Webmining"
title="Webmining" class="mw-redirect">Webmining</a> zum Einsatz.</p>
<p><br /></p>
<table id="toc" class="toc">
<tr>
```

- The web is based on the **markup language HTML**
- HTML describes
 - How information is **presented**
 - How information is linked
 - But **not**, what the information means

The Information Retrieval Dilemma

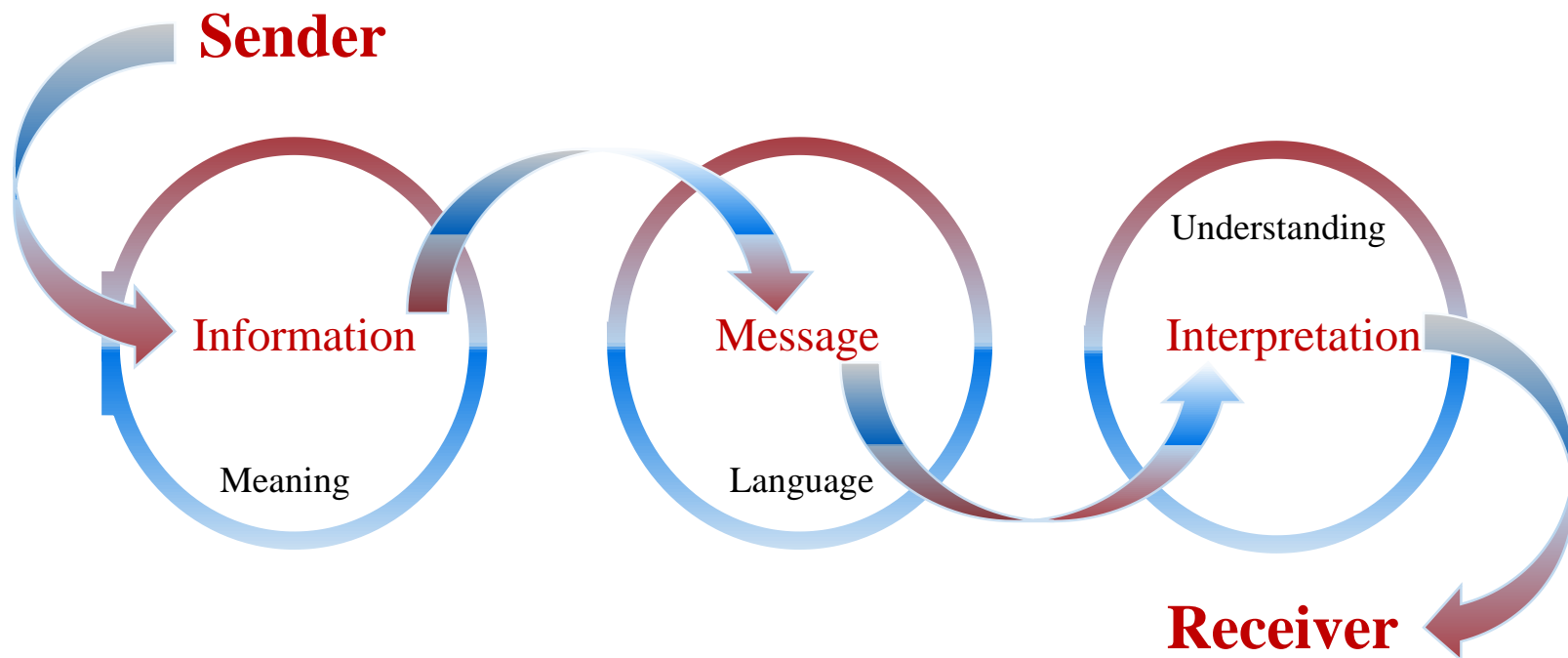
- Ambiguity of natural language (**polysemy**)
- Different words/ expressions for the same concepts (**synonyms, metaphors**)



- **Implicit Knowledge** – Information does not have to be specified explicitly, but must be derived via logical deductions from available information
- **What is the meaning?**

Meaning and Comprehension

So what does correct interpretations depends on?





Syntax



- =[greek] Arrangement, Ordering
- In grammar syntax denotes the study of the principles and processes by which sentences are constructed in particular languages.
- In formal languages, syntax is just a set of rules, by which well formed expressions can be created from a fundamental set of symbols (alphabet).
- In computer science, syntax defines the normative structure of data.

Cannot specify the meaning



Semantics



- =[greek] pertains to the character, the study of meaning
- is part of the linguistics focussed on **Sense and Meaning of language or symbols of language.**
- is the study of **interpretation of signs or symbols** as used by agents or communities within particular **circumstances and contexts.**
- Semantics asks, **how sense and meaning of complex concepts can be derived from simple concepts based on the rules of syntax.**

Cannot monitor environment/ context



Context



- denotes the surrounding of a **symbol (concept) and its relationship with the surrounding expression (concepts)**
- Context denotes **all elements of any sort of communication that define the interpretation of the communicated content**
- We distinguish
 - **General context**
 - (place, time, interrelation of action in a message)
 - **Personal or social contexts**
 - (relation between sender and receiver of the message)

Cannot clarify the intention



Pragmatics



- **reflects the intention** by which the language is used to communicate a message.
- In linguistics pragmatics denotes the study of **applying language in different situations**.
- It also denotes the **intended purpose of the speaker**.
- Pragmatics studies the **ways in which context contributes to meaning**.

Cannot replace the world knowledge



Experience



- **Experience** considers all information that you have learned and put in context with the world you are living in.
- Experience in this sense also often is referred to as **common sense knowledge or world knowledge**.

Defines the meaning and comprehension

- **Correct interpretations** depends on
 - Syntax
 - Semantics
 - Context
 - Pragmatics
 - Experience



Successful Communication



- For **successful communication**
 - Information has to be correctly transmitted (**syntax**)
 - The meaning (**Semantics**) of the transmitted information must be correctly interpreted (**Understanding**)
- **Understanding** depends on
 - The **context** of both sender and receiver
 - The **pragmatics** of its sender
- **Context** of sender and receiver depends on
 - The **experience** (knowledge of the world) of both sender and receiver



receiver

sender

Experience



Concept

symbolizes

refers to

Symbol

stands for

Object

"Jaguar"

Pragmatics

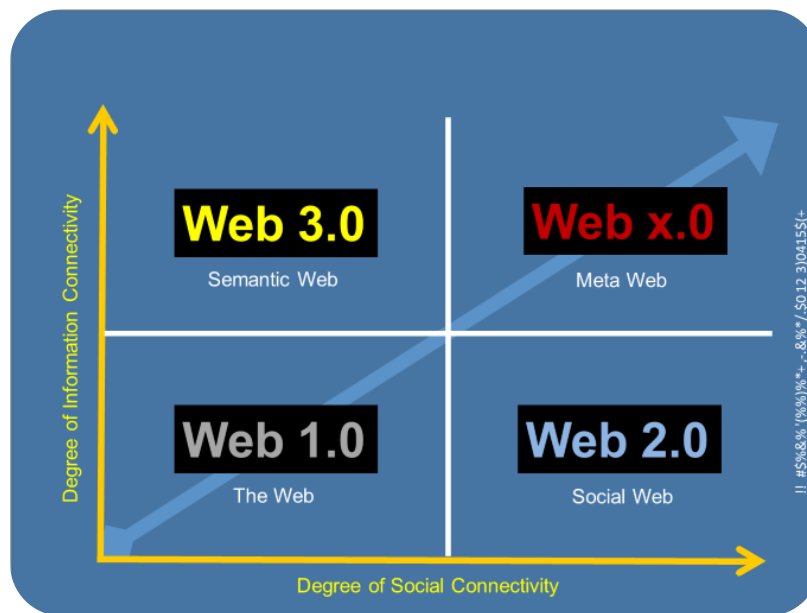
In the (traditional) Web there is no explicit semantics



Evolution of web can be categorized by the prospects of successful communication

Evolution of Web

- Web 1.0 – The static web
- Web 2.0 – The writing and participating web
- **Web 3.0 – The semantic executing web**
- Web 4.0 – “Mobile Web”
- Web 5.0 - Open, Linked and Intelligent Web = Emotional Web



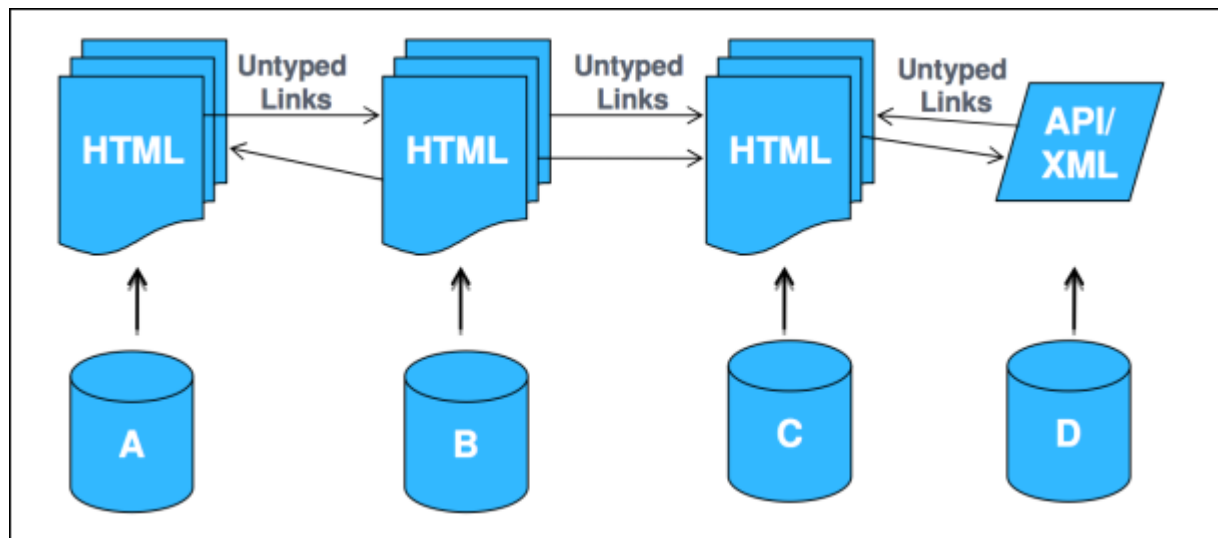


Web 1.0 – Info Centric Web



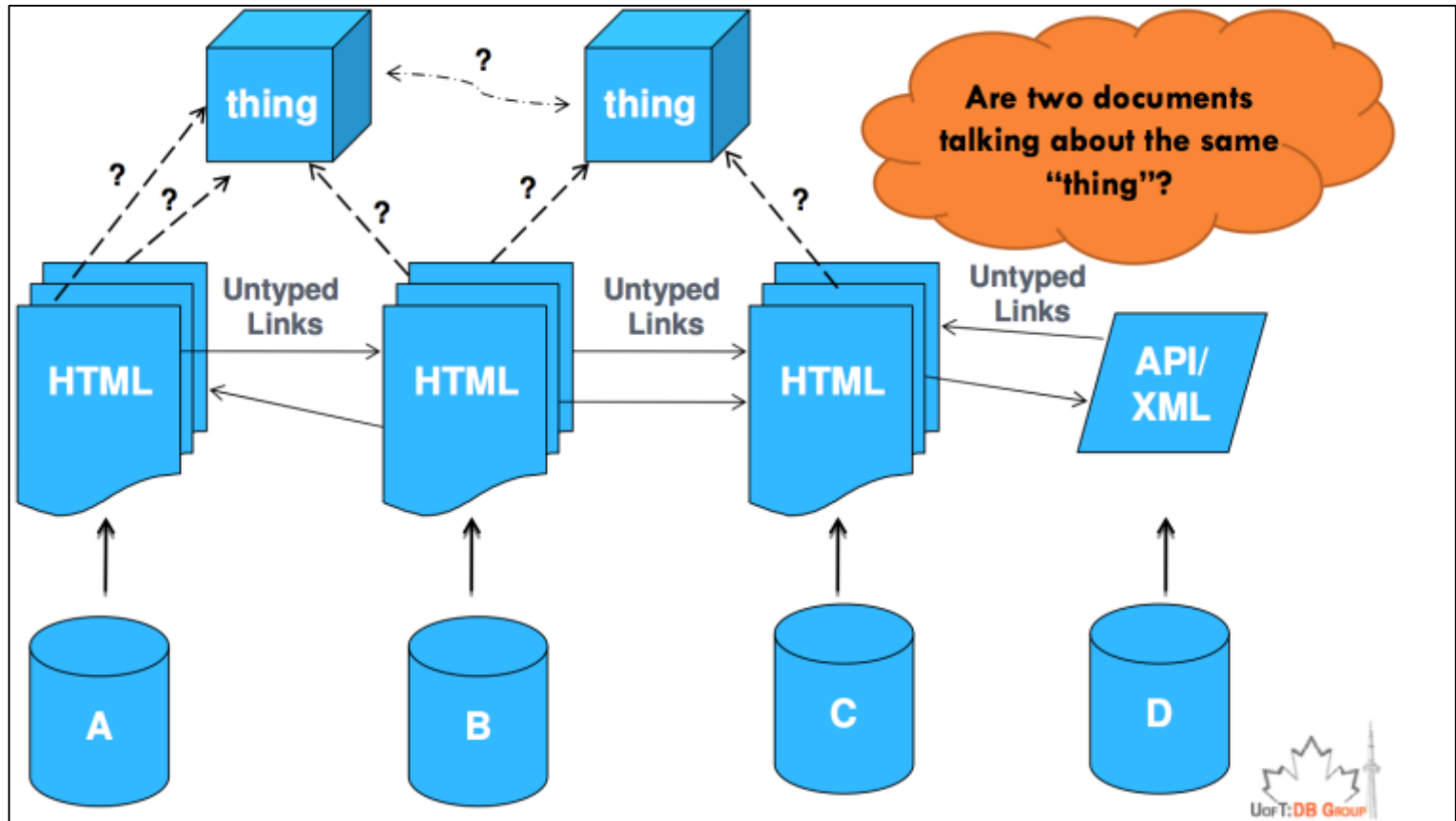
- The first generation of the World Wide Web (WWW), characterized by separate static websites.
- It is one-way broadcasting.
- It is invented 1989 by Tim Berners- Lee.
- It was widely used between 1998 and 2001, and it is still used beside Web 2.0 in almost all web sites.

Web of Documents



- Primary objects: **documents**
- Degree of structure in data: **low**
- Semantics of content: **Implicit**
- Designed for: **human consumption**
- Links between **documents**

Web of documents: The problem



The lack of active interaction of common users with the web



Web 2.0 – People Centric Web



Web 2.0 has no single definition but can be explained through a series of Internet trends, one being the empowerment of the user.

Deitel, Paul J; Deitel, Harvey M

- The year 1999 marked the beginning of a **Read-Write-Publish** era
- A non-technical user can actively interact & contribute to the web using different blog platforms.
- Ability to contribute content and interact with other web users.
- This era empowered the common user with a few new concepts like **Blogs, Social-Media & Video-Streaming**.
- Publishing your content is only a few clicks away! Few remarkable developments of Web 2.0 are **Twitter, YouTube, Flickr and Facebook**.



Web 3.0 – Machine Centric Web



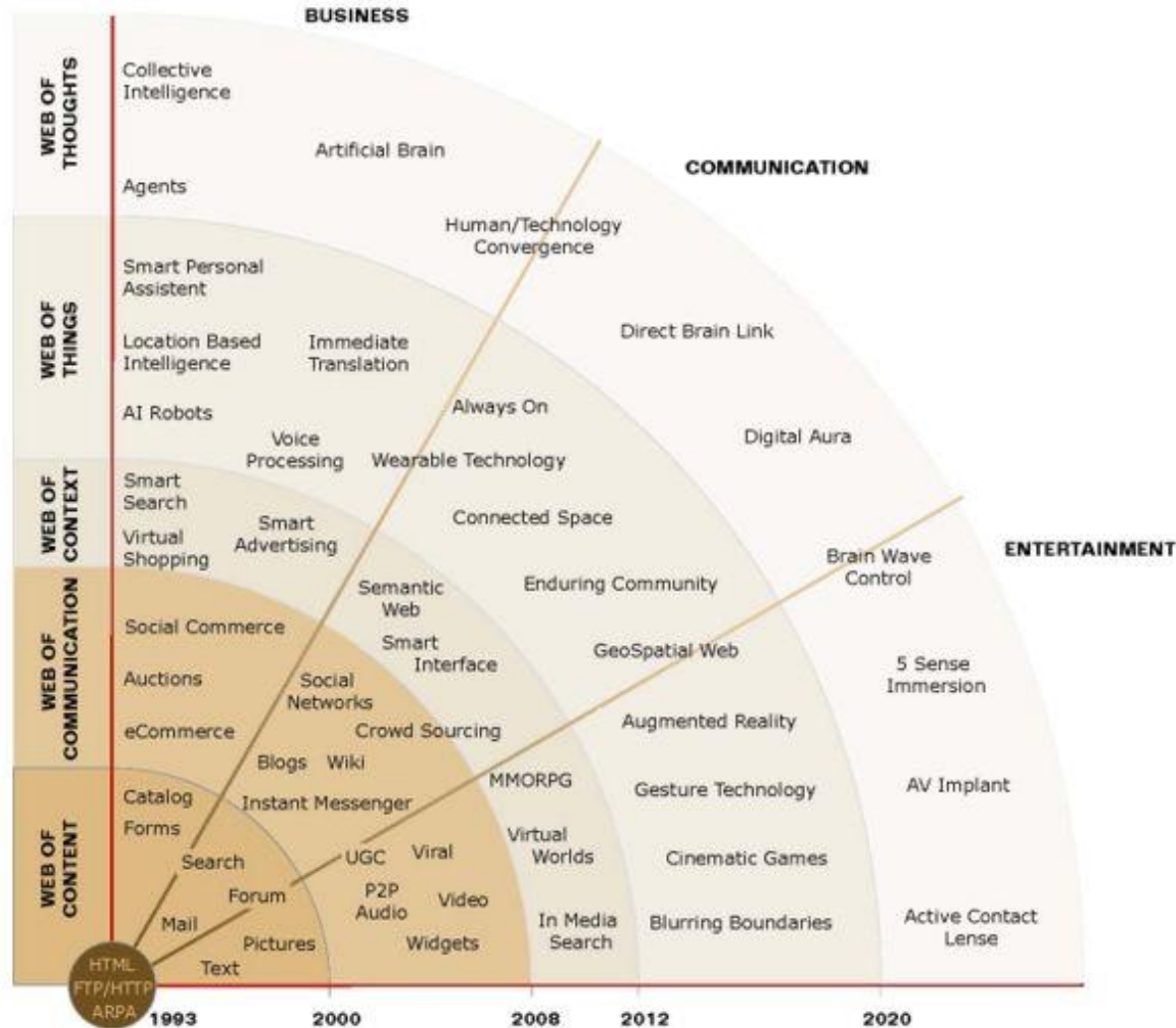
- Web 3.0 refers to a third generation of internet based services that collectively allow the emergence of Intelligent Semantic Web
- Semantic Web is a group of methods and technologies to **allow machines to understand the meaning** - or "semantics" - of information on the World Wide Web.
- The semantic web is a vision of information that is **understandable by computers, so computers can perform more of the tedious work involved in finding, combining, and acting upon information on the web.**



Comparison



Crawl	Walk	Run
Web 1.0	Web 2.0	Web 3.0





Understanding Semantic Web More

Web of data: The problem



- How about this query:
 - *How many **romantic comedy Hollywood movies** are directed by a **person who is born in a city that has average temperature above 15 degrees!**?*
- You need to:
 - Find reliable sources containing facts about **movies, directors, birthplaces of famous artists/directors, average temperature of cities across the world**, etc.
 - The result: several lists of thousands of facts
 - Integrate all the data, join the facts that come from heterogeneous sources

Understanding the Problem

- Let's organize a trip to Budapest using the Web!
- You may want to know something about Budapest; look for some photographs...

flickr
Home You Organize Contacts Groups

Budapest
Group Pool Discussion 1,418 Members Map Invite

Group Pool (19,017 items | Add photos or video)



NEW From apuc

NEW From André Fromont

NEW From carlogambino

NEW From Crashbandi

NEW From Crashbandi

NEW From Crashbandi

Discussion (33 posts | Post a new topic)

Google budapest Search Images
Moderate SafeSearch is on

Images Showing: All image sizes



E-mail this photo E-mail.
Budapest
550 x 412 - 40k - jpg
www.tripadvisor.com

Budapest, Hungary
450 x 338 - 43k - jpg
www.transitionabroad.com

Budapest looks its most beautiful at ...
1024 x 768 - 181k - jpg
web.kivf.bgf.hu

Danube Bridge Elisabeth in Budapest ...
1024 x 768 - 194k - jpg
budapest5.freeblog.hu

Budapest had 2421831 inhabitants in ...
422 x 422 - 25k - jpg
www.squidoo.com

budapest night
575 x 352 - 22k - jpg
www.wayfaring.info

Understanding the Problem

- Find a proper flight

Emirates website interface showing flight booking options. The page includes a search bar, a list of destinations, and a 'Make a Booking' section. The 'Make a Booking' section displays a table of flight options with columns for departure date, time, and price.

Departure	18 Jul	19 Jul	20 Jul	21 Jul	22 Jul	23 Jul	24 Jul
18 Jul							
19 Jul		77806 (FNR)	77806 (FNR)		77806 (FNR)	97056 (FNR)	94406 (FNR)
20 Jul		77806 (FNR)	77806 (FNR)		77806 (FNR)	97056 (FNR)	77806 (FNR)
21 Jul		77806 (FNR)	77806 (FNR)		77806 (FNR)	97056 (FNR)	77806 (FNR)
22 Jul							
23 Jul		77806 (FNR)	77806 (FNR)		77806 (FNR)	97056 (FNR)	77806 (FNR)
24 Jul							

MALEV Hungarian Airlines website interface showing flight booking options. The page includes a search bar, a list of destinations, and a 'SELECT DELIVERY METHOD' section. The 'SELECT DELIVERY METHOD' section displays a table of delivery methods with columns for delivery method, service fee, and address.

Delivery method	Service fee	Address
EMail	EUR 10	with e-invoice
Alport	EUR 30	FLM ticket office, departure hall 2
Office	EUR 30	FLM ticket office, departure hall 2
Counter		Counter or mail delivery is available only within the country of departure! You will be asked to provide a delivery address on the next page. Please note that in case of courier delivery we will be unable to deliver your ticket(s) to a PO box.
Mail	EUR 30	

Wizzair website interface showing flight booking options. The page includes a search bar, a list of destinations, and a 'new destination from London Luton' section. The 'new destination from London Luton' section displays a table of flight options with columns for departure date, time, and price.

Departure	18 Jul	19 Jul	20 Jul	21 Jul	22 Jul	23 Jul	24 Jul
18 Jul							
19 Jul		77806 (FNR)	77806 (FNR)		77806 (FNR)	97056 (FNR)	94406 (FNR)
20 Jul		77806 (FNR)	77806 (FNR)		77806 (FNR)	97056 (FNR)	77806 (FNR)
21 Jul		77806 (FNR)	77806 (FNR)		77806 (FNR)	97056 (FNR)	77806 (FNR)
22 Jul							
23 Jul		77806 (FNR)	77806 (FNR)		77806 (FNR)	97056 (FNR)	77806 (FNR)
24 Jul							



Understanding the Problem



- Find a proper hotel

The screenshot shows the HostelTraveler.com website. The header includes the site name and navigation links like 'Hostels', 'Reviews', 'Best deals', and 'Top Cities'. A central banner reads 'Find Hostels and Lodging at your destinations.' Below this, a search bar and filters are visible. A section titled '3 accommodations have been found matching your criteria.' lists results for 'Balaton, Hungary'. One result is 'Unity Hostel Balaton' with details like 'Rakoczi Ut 268', 'Hostel', and '8 Units'. A 'Make Reservations' button is present next to the listing.

The screenshot shows the Hilton website's search results page. The header features the Hilton logo and navigation links. A 'Sign in' section is visible. The main content area displays 'Search Results' for 'Hilton Hotels'. A sidebar on the left allows users to 'Change Your Search' with filters for 'Location' (City, State/Province, Country), 'Search Within' (radius and unit), and 'Brand' (Hilton Hotels, All Hilton Family Hotels). The main results section shows 'The following locations matched your request.' and lists 'Hilton Budapest WestEnd' as a result. A 'Compare' button is visible at the bottom right.



Resolve Data Heterogeneity



- There are multiple ways to encode same information
 - Muhammad, Mohammad, Mohammed, Muhammed (12 different spellings)
 - Barak Obama, Barak H. Obama, Barak Hussain Obama, Mr. Obama, President of America, President of United States of America, President of USA, American President.
- It is simply difficult to distinguish the **meaning between these two sentences:**
 - I am a computer engineer.
 - I am a computer engineer, you may think. Well, . .
- Only Web of Data is not the solution
- **We need to extend the current Web to a **standard** way for a “Web of Data”**



Semantic Web

The Technology & Experts Statements



- It involves publishing in languages specifically designed for data: **Resource Description Framework (RDF)**, **Web Ontology Language (OWL)**, and **Extensible Markup Language (XML)**:
 - HTML describes documents and the links between them.
 - RDF, OWL, and XML, by contrast, can describe arbitrary things such as people, meetings, or airplane parts.
- Tim Berners-Lee
 - “..., you’ll Have access to an unbelievable data resource”.
- Nova Spivak
 - “...It’s a set of standards that turns the Web into one big database,” .
 - “ ...I call it the World Wide Database”.



References



- Lecture Notes of Dr Haral Sack, Hasso-Plattner-Institute for IT Systems Engineering [Knowledge Engineering with Semantic Web Technologies]
- Introduction to the Semantic Web (tutorial), 2009 Semantic Technology Conference, San Jose, California, USA, June 15, 2009,
Ivan Herman, W3C
- Web 3.0, Maram Bani Younes, Marilu Cervantes Salgado, Professor A. Alsedik,
- Dr Asad Masood Khattak Lectures (KHU Faculty)



Thanks.

Next Lecture – Paradigm of Semantic Web