

From Web to Semantic Web

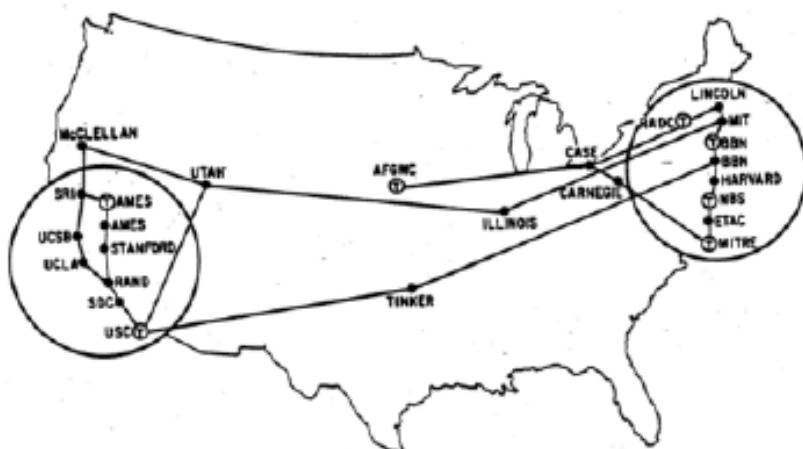
From Internet to Web 2.0



December 1969



June 1970



March 1972



July 1977

First Generation: The Internet

HPI

Hasso
Plattner
Institut

Computer Centered Processing

- How did the user get the information?

A terminal window titled "Terminal — bash — 74x20" is displayed. The window shows the command line: "Last login: Tue Nov 4 15:33:53 on ttys000" followed by "harald-sacks-macbook:~ harald\$". The rest of the terminal window is blacked out.

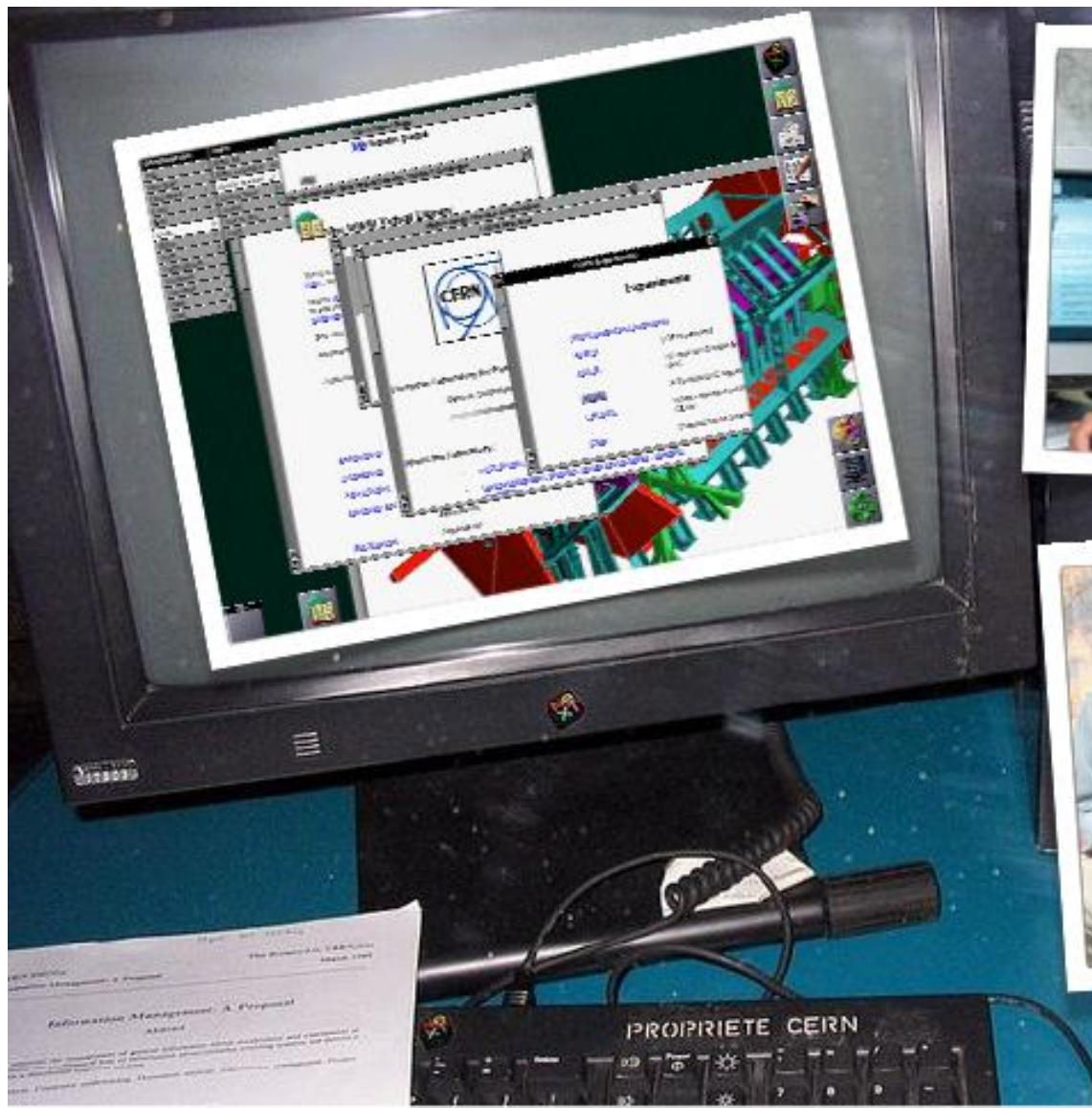
1. open terminal
2. connect to remote computer
3. retrieve file system data from remote computer
4. download file from remote to local computer
5. read file on local computer

Problem:

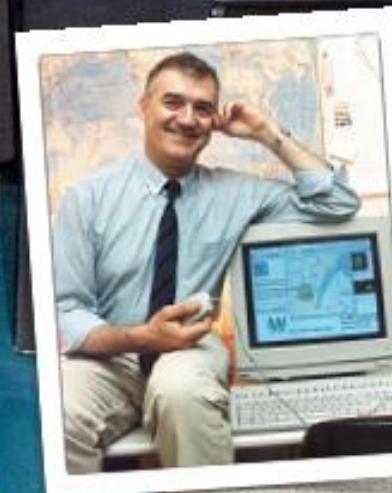
- Information access requires expert knowledge
- Information access is expensive...
- Information retrieval is very expensive...

MAP OF THE INTERNET
SATELLITE CIRCUIT
ISDN
T1
PLURIBUS UNUM
NOTE: THIS MAP DOES NOT SHOW ARPA'S EXPERIMENTAL SATELLITE CONNECTIONS

Juli 1977



Tim Berners-Lee



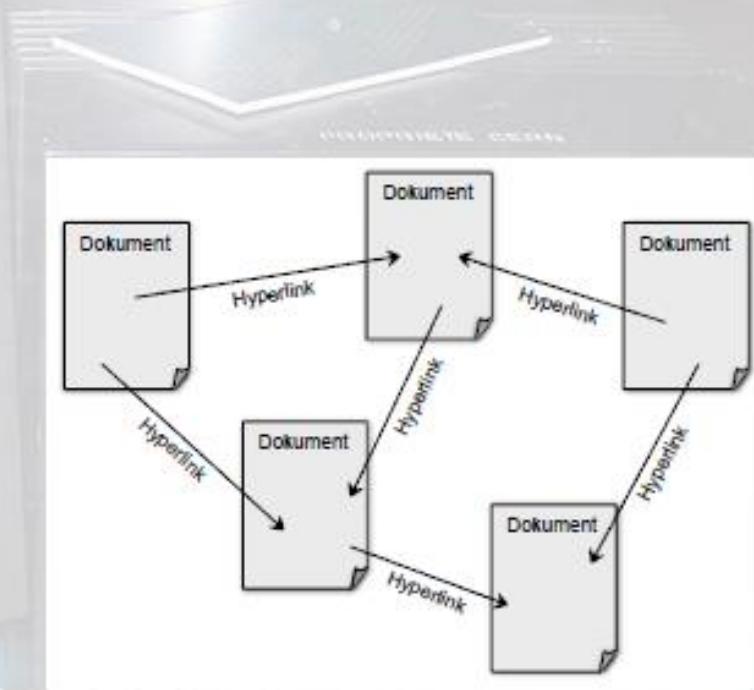
Robert Cailliau

The World Wide Web was born at the European Nuclear Research Center CERN in 1990...

Second Generation: The Web

Document Centered Processing

- How did the user get the information?



1. open browser
2. load document
3. click on next hyperlink
4. ...

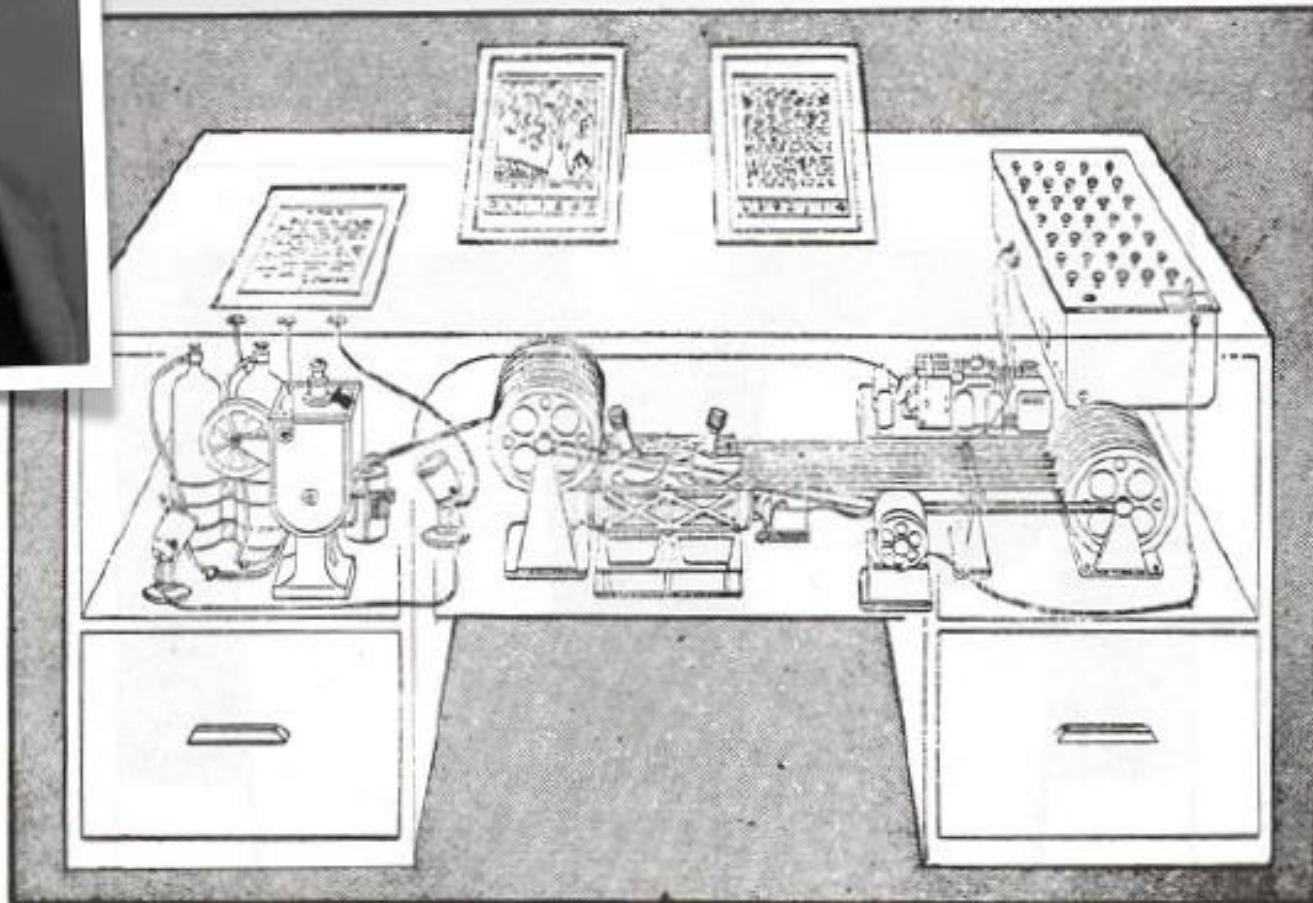
Advantages:

- No expert knowledge required
- Simple information access
- Information retrieval via search engines



Vannevar Bush
(1890-1974)

Vannevar Bush proposed the first Hypertext-System „MEMEX“ in 1945



Memex in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference (LIFE 19(11), p. 123).

Internet Domain Survey Host Count



<http://www.isc.org/>
(Stand: 10/2012)

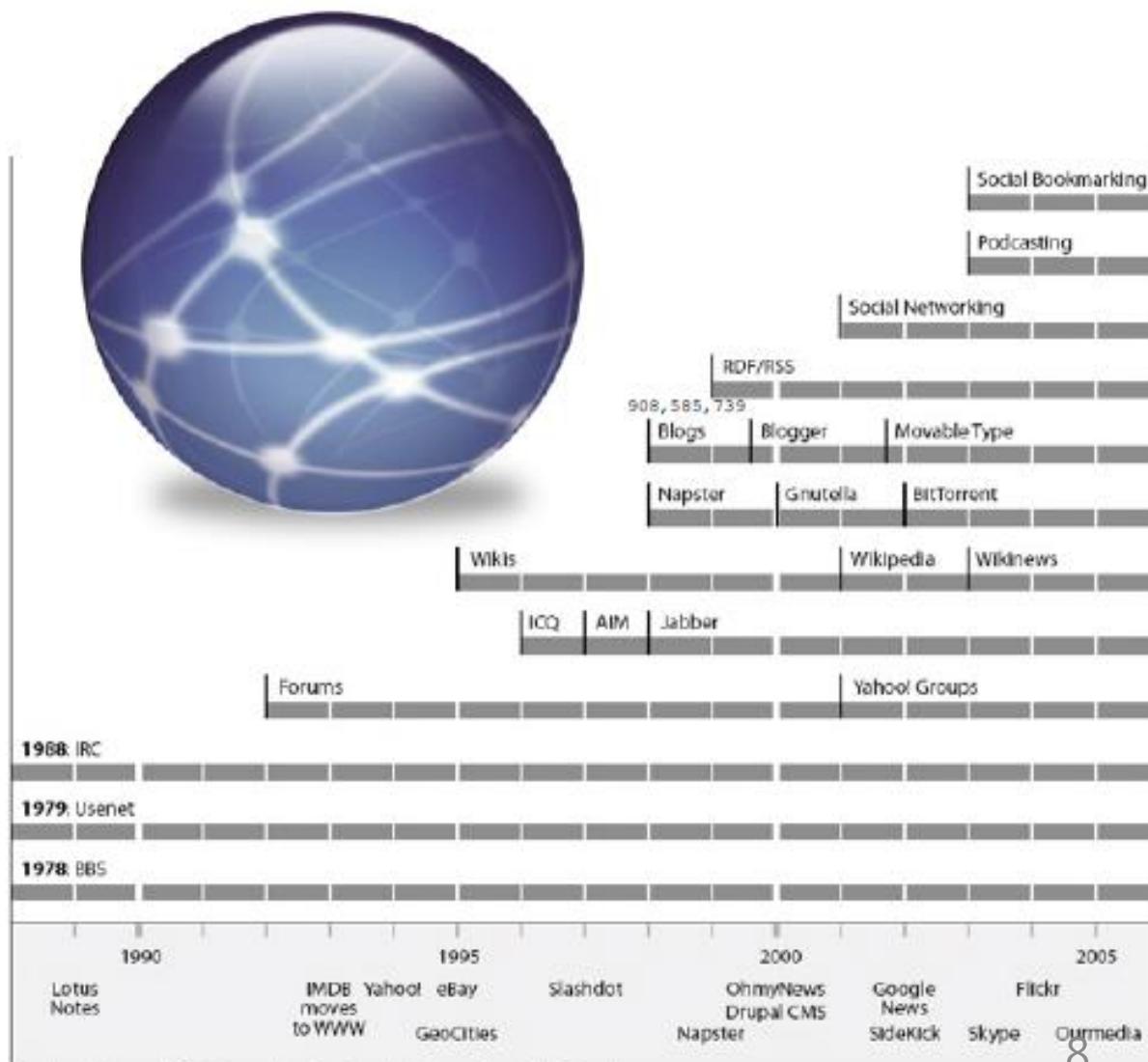
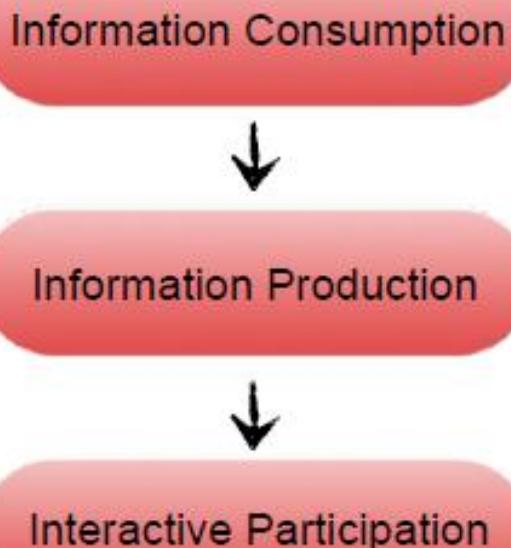
Source: Internet Systems Consortium (www.isc.org)

숙제: Update this page

From Web 1.0 to Web 2.0

Web Content and Applications are Changing

L7



• 웹 2.0 의 개념 (~~ 2004)



Tim O'Reilly

오라일리 미디어 설립자
IT 기술 전문가

Dale Dougherty

1. 웹은 플랫폼이다.
The Web As Platform
2. 집단지성을 활용한다.
Harnessing Collective Intelligence
3. 데이터가 차별화의 열쇠다.
Data is the Next Intel Inside
4. 소프트웨어 배포 주기란 없다.
End of the Software Release Cycle
5. 가볍고 단순하게 프로그래밍한다.
Lightweight Programming Models
6. 소프트웨어는 PC에 얹매이지 않는다.
Software Above the Level of a Single Device
7. 사용자들에게 더 많은 편리함을 제공한다.
Rich User Experiences

Web Content and Roles are Changing...



flickr.com: 사용자가 사진을 올리고, 올린 사진에 대해 태깅을 달고 다른 사용자는 댓글을 달아 검색을 용이하게 하기 위한 서비스이다.

Welcome to Flickr - Photo Sharing - Microsoft Internet Explorer

파일(F) 편집(E) 보기(V) 즐겨찾기(A) 도구(I) 도움말(H)

뒤로 ← → × 검색 ☆ 즐겨찾기

주소(D) http://www.flickr.com/ 미동 연결

Already a member? [Sign in](#).

flickr GAMMA

The best way to **store, search, sort** and **share** your photos.

[Sign up!](#)
Or, [learn more](#).

A [puppy](#) is but a [dog](#), plus [high spirits](#), and [minus common sense](#). - Agnes Repplier

Find a photo of... [GO](#)

Or, browse [popular tags](#)?

Explore Flickr
Browse interesting photos shared over [the last 7 days](#).

[Sign Up](#) | [Learn More](#) | [About Flickr](#) | [Terms of Use](#) | [Privacy Policy](#)

Copyright © 2006 Yahoo! Inc. All rights reserved.

a **YAHOO!** company

인터넷

del.icio.us: 온라인 상의 북마크 사이트이다. 사용자는 온라인 상에서 북마크를 할 수 있으며 또한 다른 사용자들은 그 사이트가 얼마나 많이 저장되었는지, 어떤 사용자들이 저장을 했는지 등에 대한 정보를 얻을 수 있다.

The screenshot shows the del.icio.us homepage as it would appear in Microsoft Internet Explorer. The title bar reads "del.icio.us - Microsoft Internet Explorer". The menu bar includes "파일(F)", "편집(E)", "보기(V)", "즐겨찾기(A)", "도구(I)", and "도움말(H)". The toolbar contains icons for back, forward, stop, search, and other common functions. The address bar shows "주소(D) http://del.icio.us/".

The main content area features the del.icio.us logo and navigation links: "keep", "share", and "discover". A "discover favorites:" section includes a search bar and a "search" button. A "What's a tag?" section explains what tags are and provides a link to learn more. A "sign up now" form asks for "username", "password", "password again", and "email", with a "register" button.

Below the main content, a "the del.icio.us hotlist" section displays recent bookmarks:

hot right now	1. U.S.D.C. Eastern District of Virginia	save this	44 people
	first posted by TheGoogly ... on july 31	9/11 trial moussaoui 911 exhibits tags	
	2. MacBook Pro with noise under the keyboard	save this	44 people
	3. What's New in Edge Rails: Simply RESTful Support - And How to Use It	save this	46 people

Another section shows bookmarks from "1 hour ago":

1 hour ago	1. Reviews and free downloads at Download.com	save this	77 people
	first posted by DanielGenser ... on nov 22, 2005	itunes software ipod mac apple tags	
	2. How The Piratebay Raid Changed Sweden · TorrentFreak, torrents and more	save this	45 people
	3. Wired 14.08: Pinch My Ride	save this	45 people

The bottom right corner of the window shows the Windows taskbar with icons for Internet Explorer and the Start button.

Wikipedia: 온라인 백과 사전으로 누구나 글을 쓰거나 수정/보완 할 수 있음.
다른 사람의 위키에 첨가, 삭제, 변형이 가능. 하루 평균 870만회 조회수. 전
세계 네티즌에 의해 200여개 언어로 작성

Wikipedia – Microsoft Internet Explorer

파일(F) 편집(E) 보기(V) 즐겨찾기(A) 도구(I) 도움말(H)

뒤로 뒤로 이동 검색 즐겨찾기

주소(D) http://www.wikipedia.org/ 이동 연결

WIKIPEDIA

English
The Free Encyclopedia
1 290 000+ articles

Français
L'encyclopédie libre
337 000+ articles

日本語
フリー百科事典
240 000+ 記事

Italiano
L'enciclopedia libera
181 000+ voci

Português
A encyclopédia livre
164 000+ artigos

Deutsch
Die freie Enzyklopädie
439 000+ Artikel

Polski
Wolna Encyklopedia
265 000+ haset

Nederlands
De vrije encyclopedie
216 000+ artikelen

Svenska
Den fria encyklopedin
176 000+ artiklar

Español
La encyclopédia libre
139 000+ artículos

search • suche • rechercher • szukaj • 檢索 • zoeken • ricerca • sök • busca • buscar

English >

10 000+

한국어 • עברית • ไทย • 中文 • Ελληνικά • Български • Русский • Српски • Українська • Bosanski • Català • Česky • Dansk • Førti • Simple English • Esperanto • Furlan • Gallego • Hrvatski • Ido • Bahasa Indonesia • Íslenska • Lädtöövõrasõnastik •

인터넷

Newsmap: 뉴스를 읽기 위해 회원 가입이나 광고를 볼 필요가 없을 뿐 아니라, 뉴스를 선택하여 보기 위해 사이트를 이동할 필요도 없는, 즉 모든 뉴스를 모아둔 서비스이다.

- Google Adsense: 자신의 웹 페이지나 Blog에 광고를 올릴 수 있고 클릭 횟수에 따른 수익금을 받을 수 있다. 웹 컨텐츠와 관련된 광고가 동적으로 디스플레이되며, 광고 장소 제공자는 클릭 횟수에 따른 수익금을 얻는다.

likejazz.COM - Microsoft Internet Explorer

파일(E) 편집(E) 보기(V) 즐겨찾기(A) 도구(I) 도움말(H)

뒤로 뒤로 이동 검색 즐겨찾기 주소(D) http://www.likejazz.com/?p=2

Home > Page 2

레알

월드컵도 끝난 마당에 레알(Real)을 보게된 이유는 [Fivespotting의 추천](#) 때문이었다.

5년간의 제작기간, 5개국 로케이션, 5가지 옴니버스 이야기가 명실공히 세계 최고의 축구클럽 레알 마드리드(Real Madrid)의 드라마와 함께 담겨있다. 옴니버스 드라마의 대표격인 러브 액츄얼리(Love Actually)를 생각하면 틀리지 않다. 단순히 레알 마드리드의 홍보 드라마쯤으로 치부하기엔 5가지 에피소드가 차지하는 비중또한 결코 작지않다.

하지만 이 영화의 주인공은 레알 마드리드이고 "레알 마드리드라는 축구팀이 어떻게 많은 팬들의 삶과 그들의 열정에 영향을 끼쳤는가 [Fivespotting]"에 대한 이야기다. 특히 후반부 바르셀로나(FC Barcelona)와 경기장면은 축구가 얼마나 멋진 스포츠인지 느끼기에 충분하다.

또다른 멋진 축구영화로 Seo님이 추천한 "[반드시 봄야 할 축구영화](#)"도 빼놓을 수 없다.

어쨌든 굳이 이 포스팅을 남긴 이유는 다른데 있다. 바로 영화중간, 축구에 전혀 관심없던 역사선생 마틴이 레알 마드리드에 관심을 갖기 시작하는 모습에 등장하는 화면 때문.



[About Me](#)

[Recent Comments](#)

[Lens](#)

[1,630 readers](#)

[Goooooogle 광고](#)

[프로토콜 스트리밍](#)

서버 CPU 사용률 33%
감소 보다 빠른 네트워크 성능 구현
[www.Intel.com](#)

[미 사이트에 광고하기](#)



이 저작물은 [크리에이티브 커먼즈 라이센스](#)에 따라 이용하실 수 있습니다.

완료 인터넷

creativecommons.org: License 계약을 체결하기 위한 일종의 표준 약관과 같다. 저작재산권자는 저작물의 이용을 허락할 수 있고, 허락을 받은 자는 "허락 받은 이용방법 및 조건의 범위 안에서" 저작물을 이용할 수 있다.

The screenshot shows the Creative Commons homepage as it would appear in Microsoft Internet Explorer. The top navigation bar includes links for 'visit ccMixter', 'visit science commons', 'visit iCommons', and 'Support the Commons'. A 'Worldwide' section allows users to 'Select a jurisdiction'. The main content area features several sections: 'Find' (Music, photos, and more) with a 'cc' logo icon, 'Publish' (Your stuff, safely and legally) with a user profile icon, and a 'Learn more...' link. On the left, there are categories for 'Audio', 'Images', 'Video', 'Text', and 'Education'. Below these is a 'Weblog' section with a message about the new search UI. A 'Featured Content' section highlights the animated series 'Odd Job Jack'.

Welcome | Creative Commons - Microsoft Internet Explorer

파일(F) 편집(E) 보기(V) 즐겨찾기(A) 도구(I) 도움말(H)

뒤로 ← → × 검색 ☆ 즐겨찾기

주소(D) http://creativecommons.org/ 미동 연결

visit ccMixter visit science commons visit iCommons Support the Commons

Worldwide Select a jurisdiction

creative commons

Creative Commons is a nonprofit organization that offers flexible copyright licenses for creative works.

Audio music, sounds, speeches...

Images photos, illustrations, designs...

Video movies, animations, footage...

Text books, blogs, essays...

Education lesson plans, course packets, textbooks...

Find Music, photos, and more

Publish Your stuff, safely and legally

Learn more...

Creative Commons licenses provide a flexible range of protections and freedoms for authors, artists, and educators. We have built upon the "all rights reserved" concept of traditional copyright to offer a voluntary "some rights reserved" approach. We're a nonprofit organization. All of our tools are free.

Weblog

New CC search UI Today we turned on a new search interface offering tabbed access to Yahoo! and Google CC-enabled web searches and Flickr CC image search. We hope that this new search user interface highlights the ...

weblog archives | RSS

Featured Content

ODD JOB Jack

You can catch the new season of the hilarious animated series *Odd Job Jack* – about a temp worker's myriad employment misadventures – on the Comedy Network in Canada. Better yet, make your own version of the hit cartoon! *Odd Job Jack*'s creators recently launched a site called *Free Jack*, in which the master Flash files and bitmaps of every piece of art used in this season of the show are being released under the Creative Commons

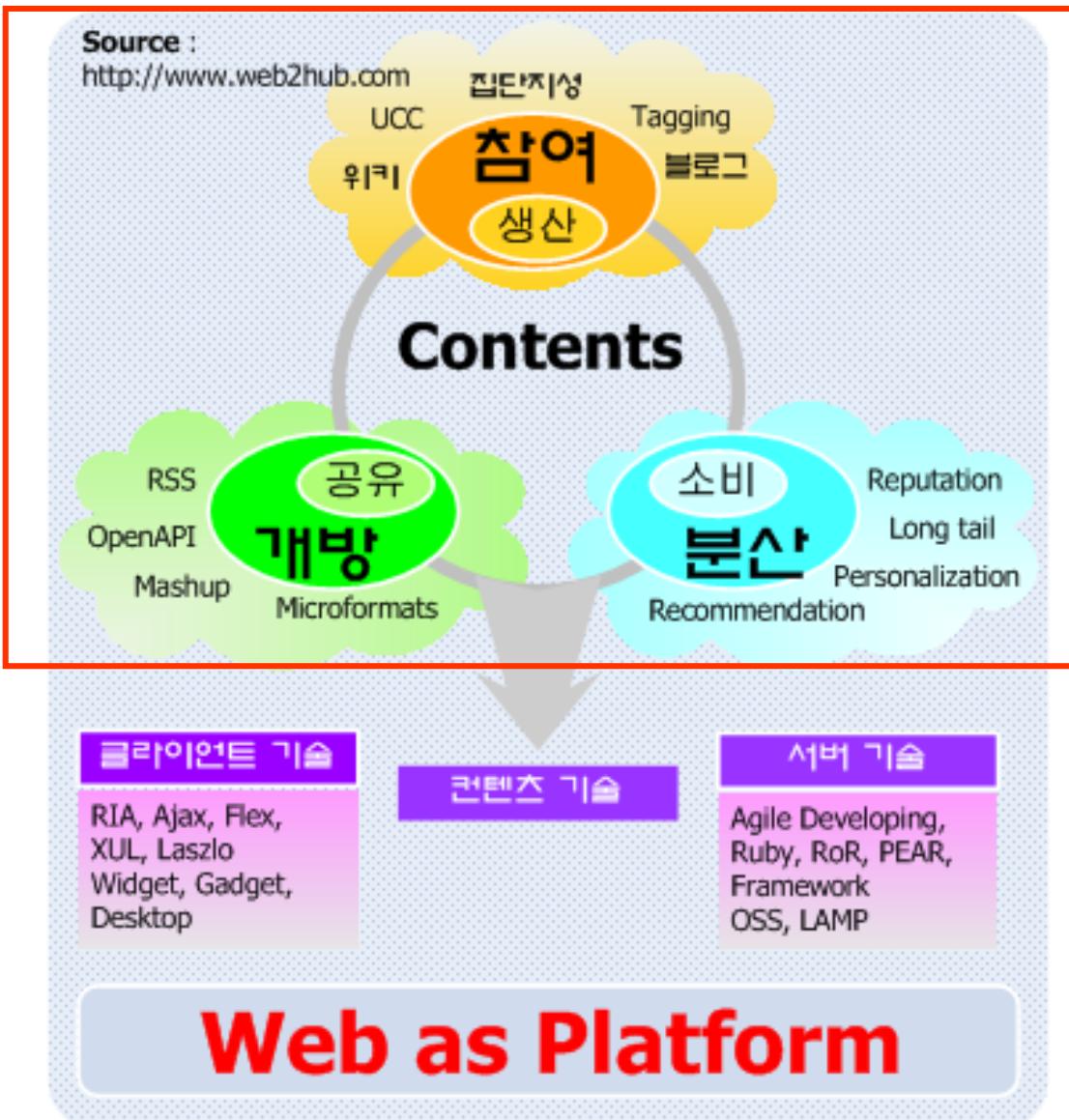
Web2.0 특징: 참여를 통한 컨텐츠 생산



UCC(User-Created Contents)



Web 2.0: 컨텐츠의 생산과 유통



Web2.0 vs Semantic Web [1/2]

- Semantic Web
 - 기존의 Web Contents에 의미를 저장 & 제공하여 사람과 컴퓨터가 좀더 협력을 잘 할수 있는 웹 (by Tim Burners Lee)
 - 컴퓨터끼리 정보해석력을 높여서 자동화 처리가 향상되는 웹
- Web 2.0
 - Web 2.0은 Semantic Web을 경제적 관점에서 보는 표현
 - Semantic Web에 의해 개발된 기술을 이용, 좀 더 효율적이고 지능적인 웹생활의 구현이 Web 2.0의 목표
- Business Sector → Web 2.0
- Academic Sector → Semantic Web
- Web 2.0 ~~ Semantic Web ~~ 차세대 웹 ~~ 미래의 웹

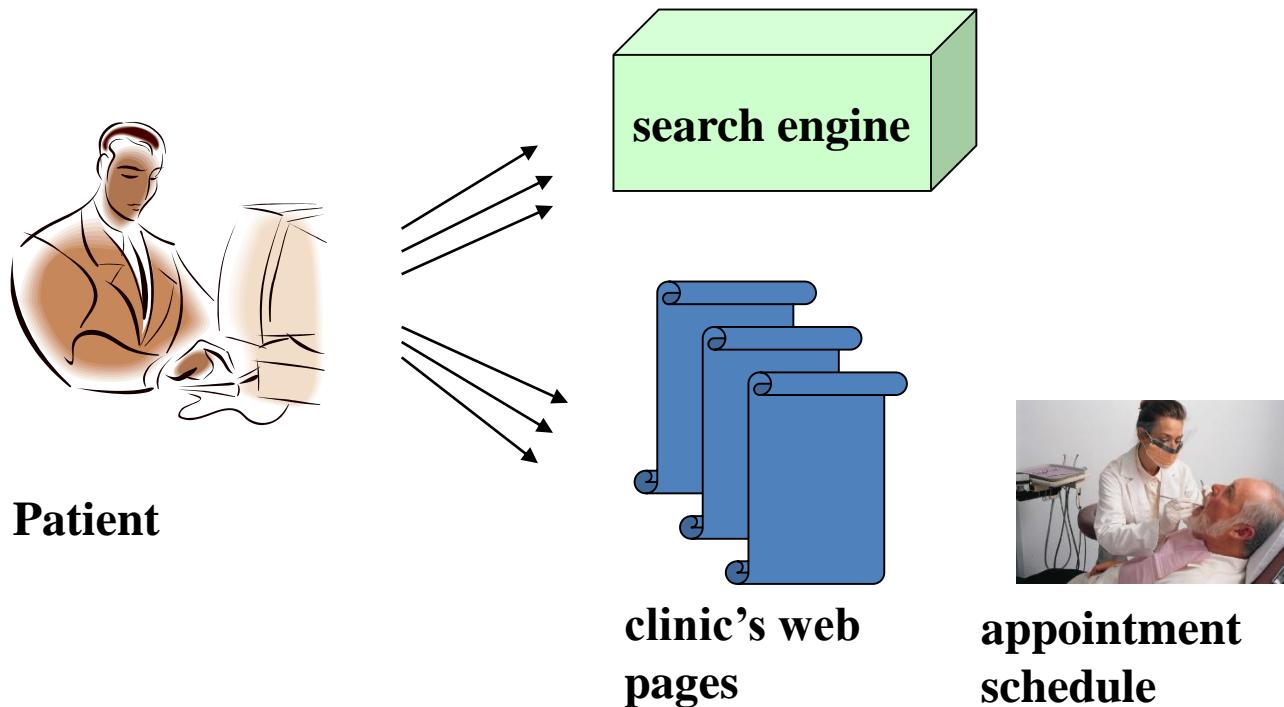
Web2.0 vs Semantic Web [1/2]

- Web2.0: 풍부한 Web-contents를 위해서 사용자의 참여를 적극적으로 유도하는 개방적 체계 (User Created Contents)
 - 오래된 기술: Blog, P2P
 - 최근의 기술: RSS, Tagging
- Semantic Web: 기존의 Web에서 표현되는 정보의 의미 (semantics)를 새로운 기법으로 강력하게 표현하고 저장 (Web Intelligence)
 - Ontology 표현기술: RDF, OWL
 - 지능형 검색과 지능형 웹서비스를 가능
 - 컴퓨터끼리 대화하고 자동화된 지능형 웹을 추구

Web Intelligence의 예 (1/2)

- 기존의 web:

- 1) 환자가 검색 엔진에서 치과를 검색
- 2) 자신의 장소와 가까운 치과의 홈페이지를 찾음
- 3) 치과의 진료 스케줄을 확인하고 자신과 시간이 맞을 경우 예약
- 예약을 하기까지 다수의 반복 작업 필요

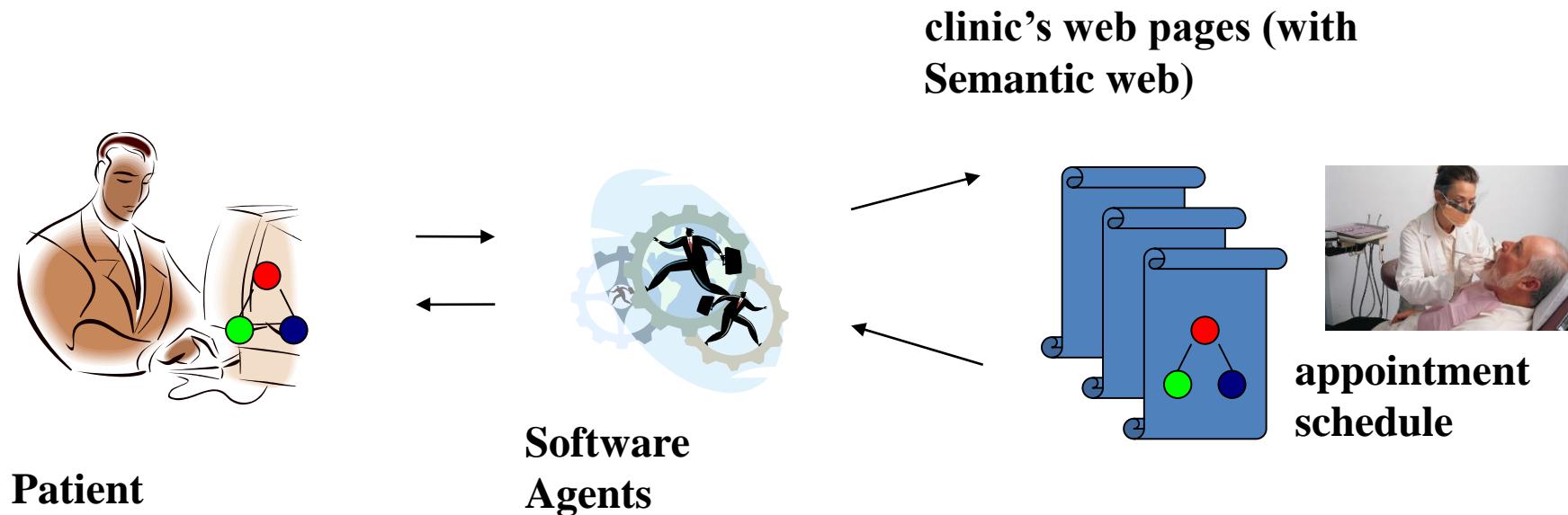


Web Intelligence의 예 (2/2)

- Semantic web으로 다음의 정보가 구축된 상태라면
 - 환자의 개인 스케줄, 각 치과의 위치, 진료 과목, 진료

(1) 환자는 software agent에게 예약 요청

(2) 각 병원의 홈페이지의 내용이나 구조가 다르더라도 software agent가 환자와 치과의 시멘틱웹 데이터를 분석, 환자의 시간과 위치에서 진료 가능한 치과를 예약해 줌



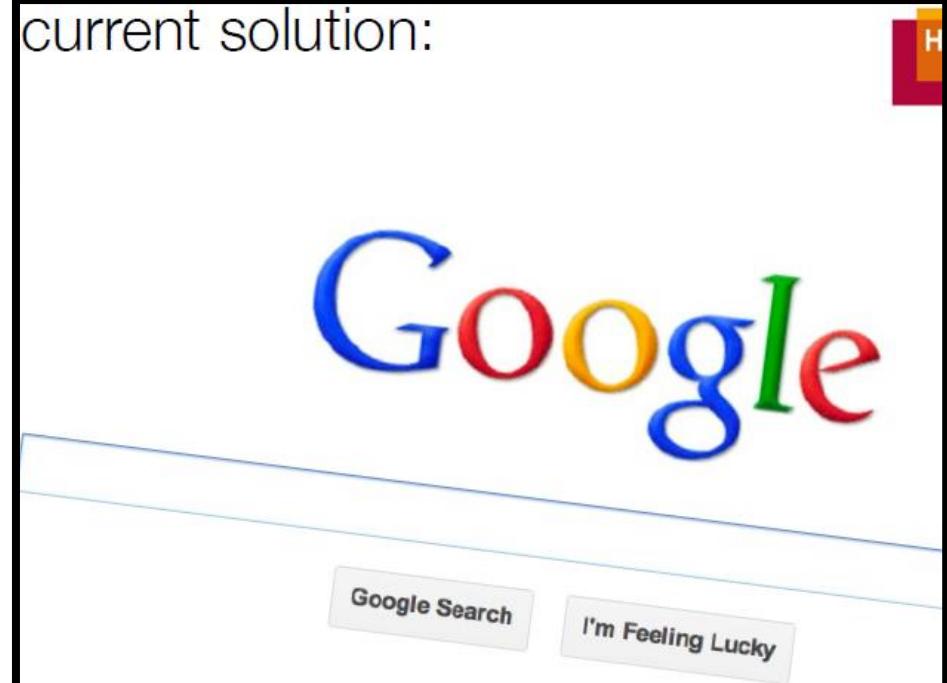
Super Challenge!

Intelligent Information Retrieval (지능형 정보 검색)

How can we find something
in the Web?



current solution:



The Web is really big...

- ca. 25×10^9 indexed documents in search engines

(TNL Blog: Google has 24 billion items index,
considers MSN search nearest competitor, September 2005)

- Web Crawler: $> 10^{12}$ (1 trillion) documents

(The Official Google Blog: We knew the Web was Big....., Juli 25, 2008)

- Google Search Index Caffeine comprises
ca. 100 Million Gigabytes i.e. 10^{17} Byte

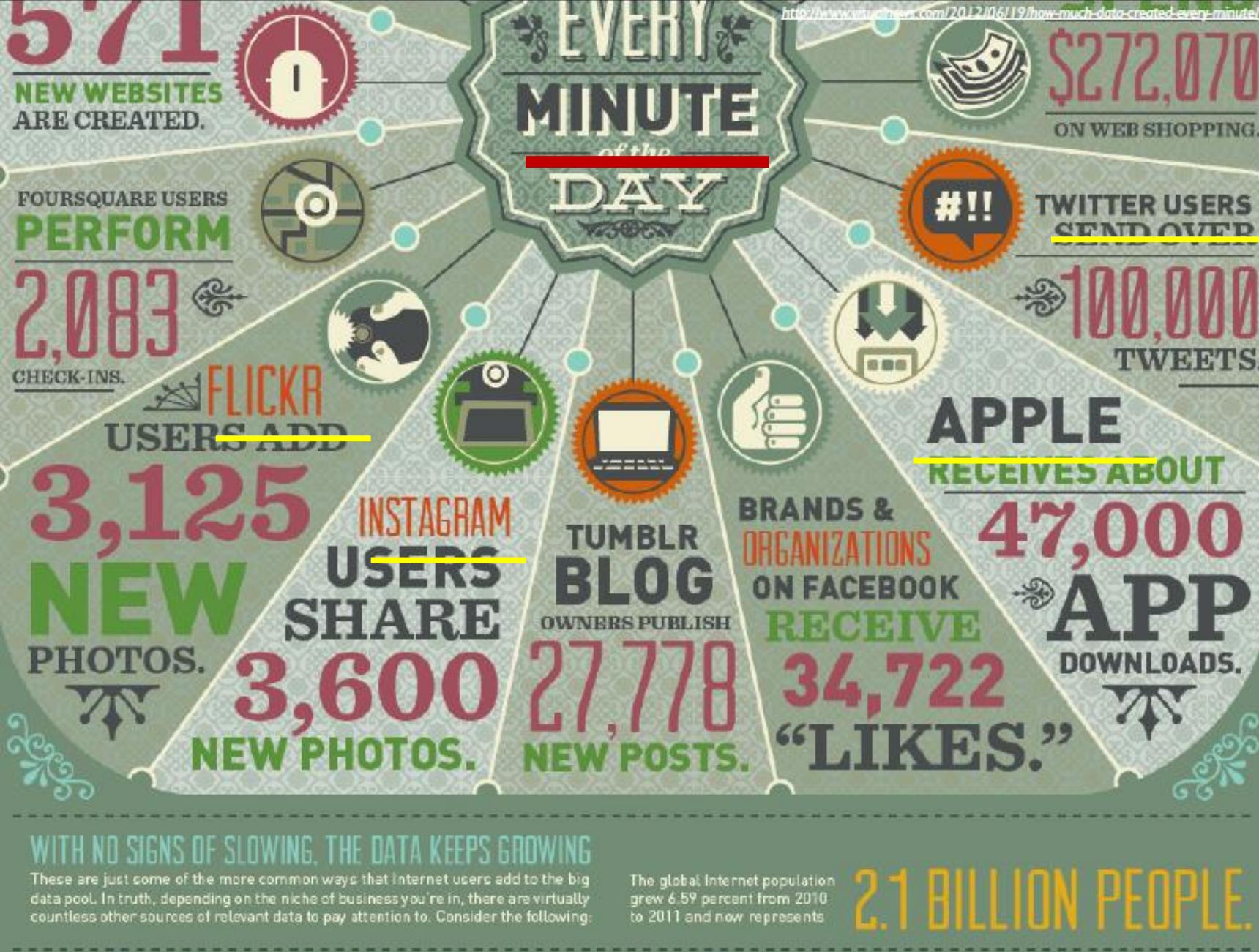
(SMX Video: Google's Matt Cutts On Caffeine Launch, June 9, 2010,
<http://searchengineland.com/smx-video-googles-matt-cutts-on-caffeine-launch-43933>)

- DeepWeb (Darkweb) estimated to be about 550 times
bigger than Surface Web

(Bergman, 2001)

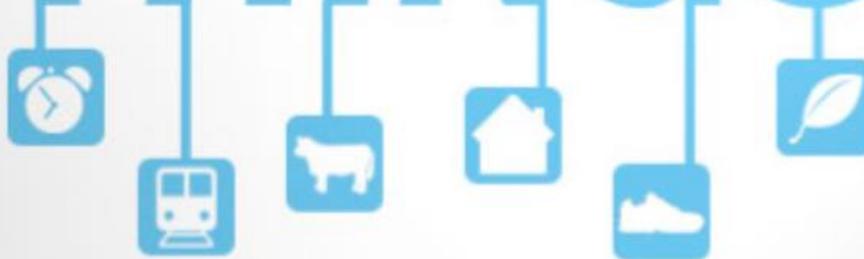
숙제: Update this page





숙제: Update this page

The INTERNET *of* THINGS



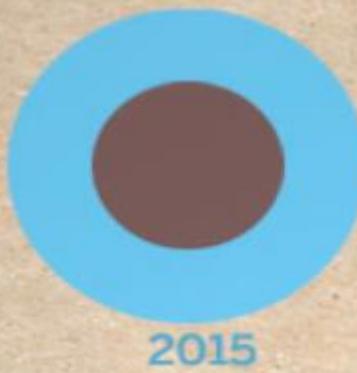
During 2008, the number of **things** connected to the Internet exceeded the number of **people** on earth.



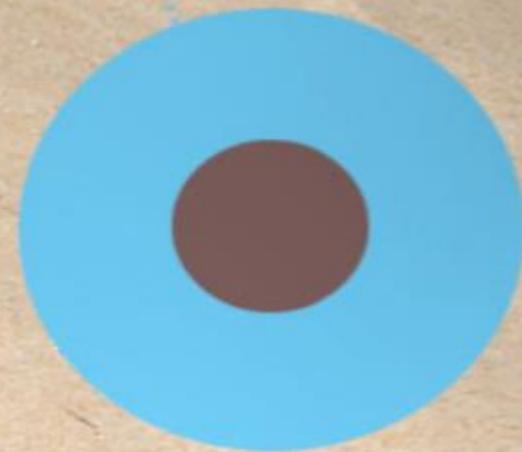
2003



2010



2015



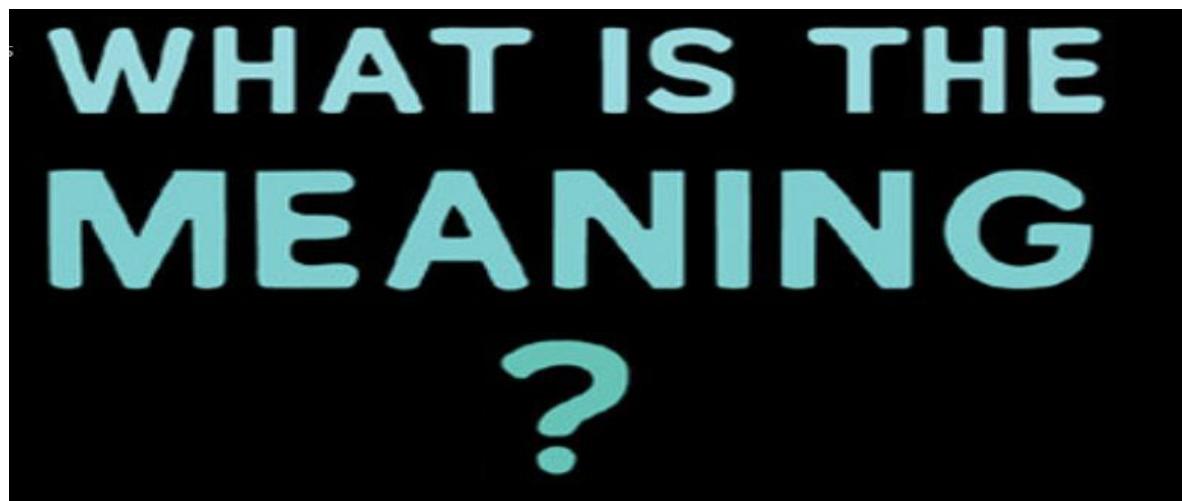
By 2020 there will be 50 billion.

Information in the WWW

- what is **important** and how do you know?
- what is information, what is **advertisement**?
- what does the information **mean**?
- how **credible/trustworthy** is the information?
- what belongs **together**?
- what is **redundant**?

- Humans have **contextual knowledge**, **world knowledge** and **experience** to solve the problem

- The Web is supposed to be used by humans
- 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



Semantics

Semantics (greek σημαντικός = *pertains to the character, the study of meaning*) is part of the linguistics focussed on

- **Sense** and
- **Meaning**

of language or symbols of language. It 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**.

The semantics of a message depends of its **context** and **pragmatics**.

Syntax

Syntax (greek σύνταξις = *Arrangement, Ordering*) as in grammatics 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.

Context

Context (lat. *contextus* = *interweaved*) denotes the surrounding of a symbol (concept) in an expression resp. its relationship with surrounding expressions (concepts) and further related elements,

Contexts denotes all elements of any sort of communication that define the interpretation of the communicated content, as e.g.,

- **general contexts:**
place, time, interrelation of action in a message
- **personal or social contexts:**
relation between sender and receiver of a message

Pragmatics

Pragmatics (greek. πράγμα = *action*) 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.

Experience

Experience considers all information that you have learned and put in context with the world you are living in.

For **successful communication**,

- information has to be correctly transmitted (**Syntax**)
- the meaning (**Semantics**) of the transmitted information must be interpreted correctly (= **understanding**)
- **understanding** depends on
 - the **context** of both sender and receiver and
 - the **pragmatics** of the sender
- the context of sender and receiver depends on
 - the **experience** (knowledge of the world) of both sender and receiver

Meaning



sender

receiver

Experience

Concept

symbolizes

refers to

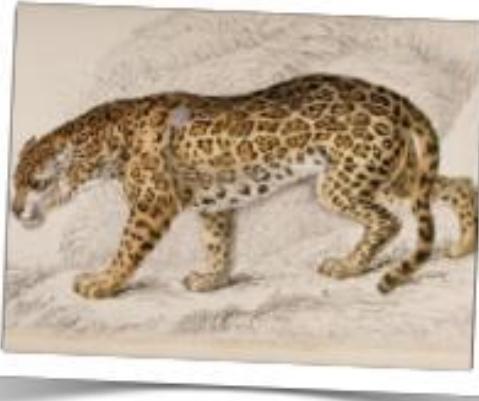
Experience

Symbol

stands for

Object

"Jaguar"



Pragmatics

How do we get the meaning in the Web...?

Google jaguar

Search Images Maps Videos News Shopping More Show more

10,300,000 results 254,000,000 after filters

Jaguar International - Market selector page
www.jaguar.com

Official worldwide web site of Jaguar Cars. Please users to pages tailored to region and model-specific information.

Jaguar USA: Our Lineup
Used Pre-Owned Jags
www.Jaguar.com/rightsidesite.com/2009

Our mission at Jaguar has been to create and introduce some of the most beautiful cars. The XK, XJ and XJL bring the...

Jaguar™ Australia
www.jaguar.com.au/en

2 Aug. 2006 Explore Jaguar's range of Luxury Cars. Choose from the XJ, XF & XK. Explore Jaguar™

More videos for jaguar >

Jaguar – Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Jaguar

The jaguar is a big cat, a feline in the Panthera genus, and is the only Panthera species found in the Americas. The jaguar is the largest feline in the Americas, and is the third largest in the world.

Jaguar Cars – Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Jaguar_Cars

Jaguar Cars Ltd., known simply as Jaguar is a British manufacturer, headquartered in Whitley, Coventry, England. It is part of the Jaguar Land Rover business, a subsidiary of the Indian company Tata Motors.

Jaguars, Jaguar Pictures, Jaguar Facts - National Geographic

Jaguar

Jaguar Cars Ltd., known simply as Jaguar, is a British luxury and sports car manufacturer, headquartered in Whitley, Coventry, England. It is part of the Jaguar Land Rover business, a subsidiary of the Indian company Tata Motors.

Facebook See results about:

Jaguar The jaguar is a big cat, a feline in the Panthera genus, and is the only Panthera species found in the Americas.

- traditional keyword-based search leads to **many not relevant results**
- different meanings
- Polysemy
- different contexts

Problem 1: Information Retrieval

Polysemy 다의어 ↔ Monosemy 35

About 424,000 results (0.38 seconds)

Ad related to "Panthera onca" (1)

Jaguar[Pics - Panthers Works to Save Wild Cats](#)

Big Cats

[Jaguar - Wikipedia, the free encyclopedia](#)
en.wikipedia.org/w/index.php?title=Jaguar&oldid=5191110The jaguar, *Panthera onca*, is the only extra-gensis. DNA evidence shows the lion, tiger, jaguar. Cars - Jacksonville Jaguars - Jagua**Panthera Onca**[Panthera onca \(jaguar\)](#)www.iucnredlist.org/details/31911/0

by A Caso - 2011 - Cited by 10 - Related articles

[Panthera onca: Status_no_of Status_dd_of Status_lc_of Status_mt_en Status_su_of Status_en_of Status_o_of Status_ew_of Status_ex_of: Summary ..](#)[ADW: Panthera.onca: INFORMATION](#)animaldiversity.ummz.umich.edu/accounts/Panthera_onca/[Panthera.onca:jaguar: By Jonathan Noguera: Geographic Range; Habitat; Description; Reproduction; Lifespan / Longevity; Behavior; Communi](#)[Panthera | Jaguares](#)www.panthera.org/species/jaguarThe jaguar (*Panthera onca*) is the third largest cat in the world, after the lion, but it is the largest feline in the Western Hemisphere. During the[Jaguar \(Panthera.onca\) - theBigZoo.com](#)www.thebigzoo.com/Animals/Jaguar.asp

The Jaguar is the largest cat in the Western Hemisphere and the third largest in the world (after the Lion and the Tiger.) It is also one of the four roving cats.

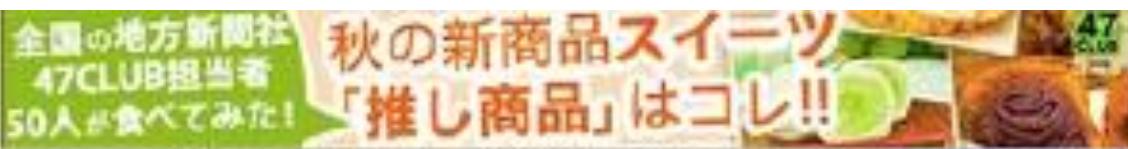
[Jaguar - Panthera onca - NatureWorks](#)**jaguar**

The jaguar is a big cat, a feline in the *Panthera* genus, and is the only *Panthera* species found in the Americas. The jaguar is the third-largest feline after the tiger and the lion, and the largest in the Western Hemisphere. Wikipedia

Scientific name: *Panthera onca*

- traditional keyword-based search
- **does not find all results**
- synonyms and metaphors
- missing context definition

Problem 1: Information Retrieval



天地人 | 社説 | 訃報 | お悔やみ | 青宮
プロ野球 | MLB | Jリーグ | 大相撲 | 高校野球

Go ウェブ

エンタメ | コラム | 新刊本 | 新作映画 | 音楽CD | 格闘技

東奥日報ご愛読者限定サービス!

ハートナビ

Information Extraction

- can only be solved 'correctly' by a human agent
- heterogeneous distribution and order of information
- Software agent does not have sufficient
 - knowledge of contexts
 - world knowledge and
 - experience
- to solve the problem

青森のニュース 2012年10月14日 22:37 RSS ブログバースト

第10回 弘前・白神アップルマラソン写真特集

「青めぐり」DVD発売 青森県初実放 青日本大賞

Problem 2: Information Extraction

・中古車10万円
・中古車ネットガーデン
・ネット住宅展示場
・メール申し込み
・Miniガイド

ロレックス

正規品取扱店

マエバー

人材派遣なら
東奥日報人材センター アウト

青森01-フライネット

サンワアリーナ

風家

高木ちゃん・福井

クレバーホーム

Cubic Four

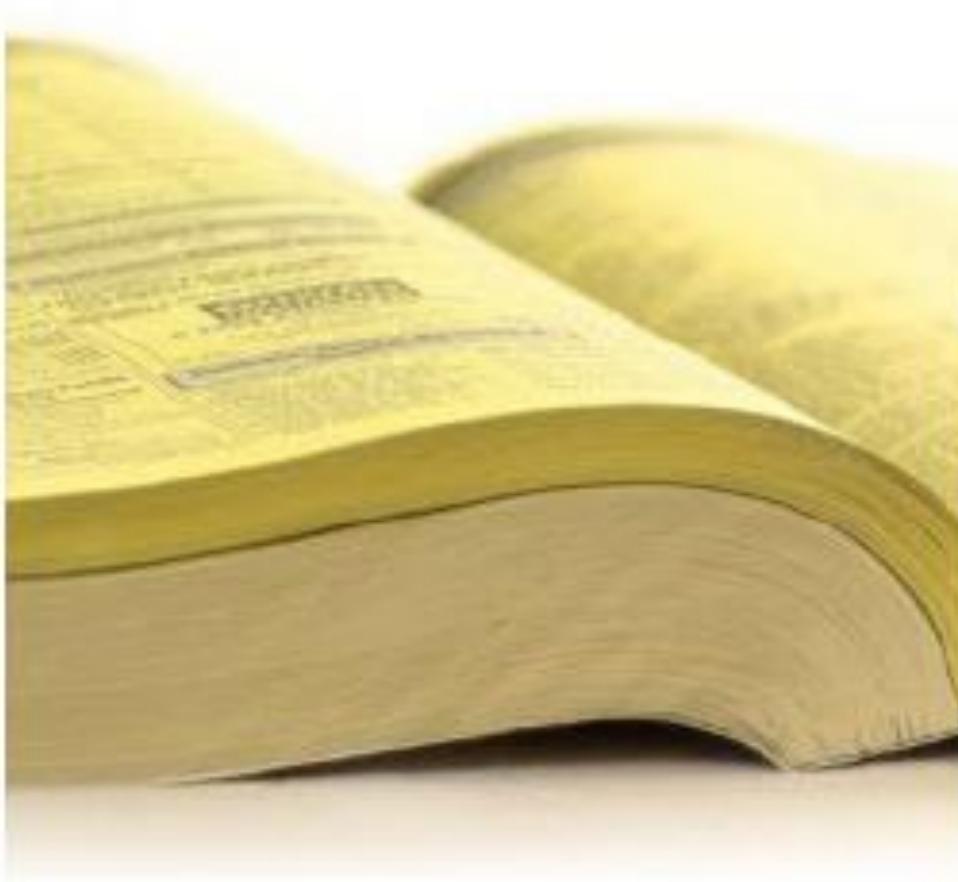
弘前初音便

お問い合わせはご当地の

NEW

- **implicit knowledge**, i.e. information does not have specified explicitly, but must be derived via logical deductions from available information.





Problem 3: Maintenance

- the more complex and voluminous a website, the more complicated is the maintenance of the only weakly structured data.
- Problems:
 - syntactic and semantic (link) consistency
 - correctness
 - timeliness

- Adaption of the presented information content to personal requirements
- Problems:
 - from where do we get the required (personal) information?
 - personalization vs. data security



From the World Wide Web to the Web of Data

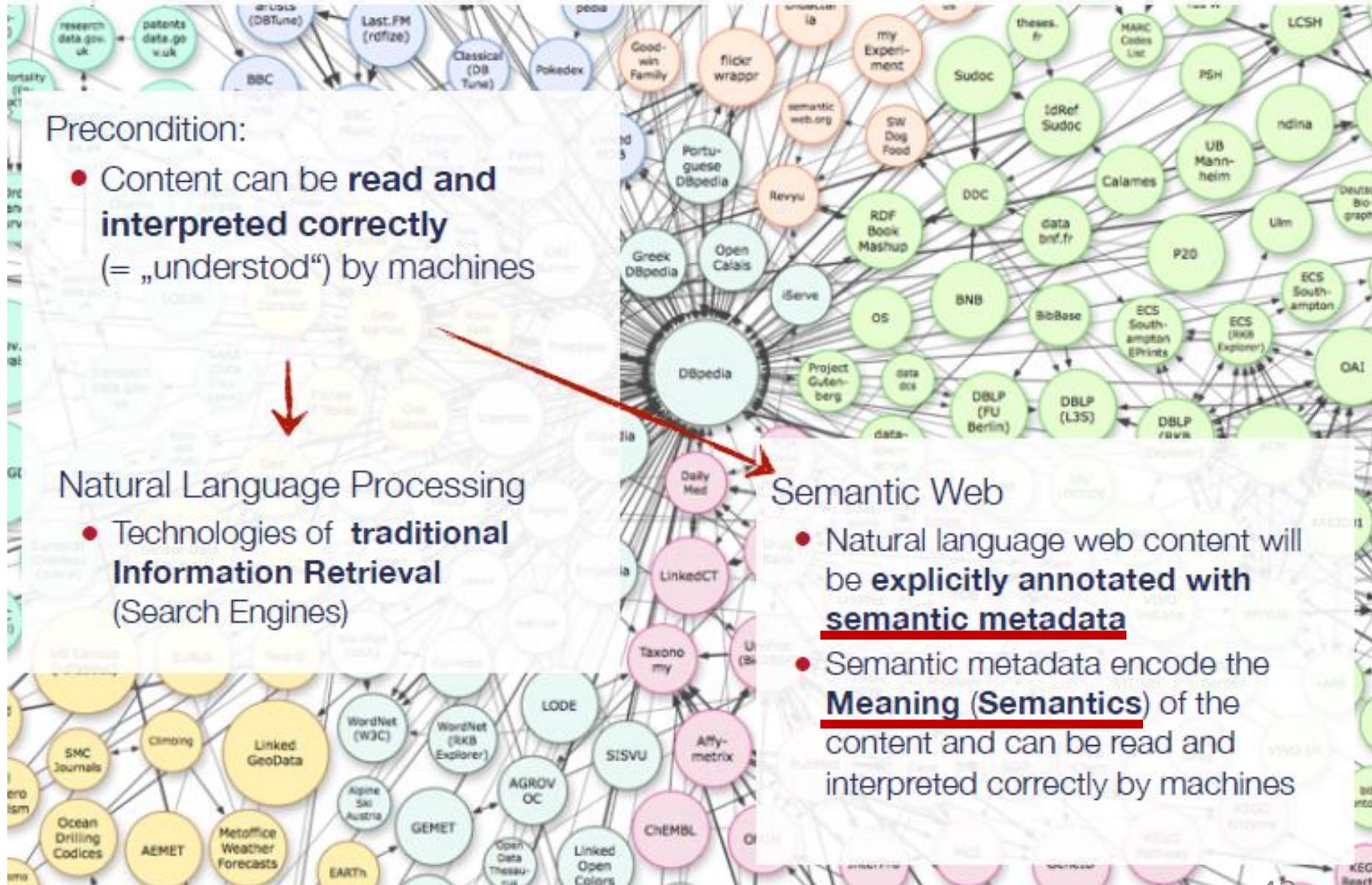


Vision of Semantic Web

„The Web was designed as an information space, with the goal that it should be useful not only for human-human communication, but also that machines would be able to participate and help...“

Tim Berners-Lee, Semantic Web Roadmap, Sept 1998

From the World Wide Web to the Web of Data



„Understanding“ Content on the Web

Snub 모욕하다

9



- **Disambiguation**

solution of linguistic ambiguities

Text: „Why snub me, Fab?“

Entity Mapping
Disambiguation

Fab ?

a detergent

Fab ?

a Canadian gay magazine

FAB ?

the Film Advisory Board

FAB ?

Federación Atlética de Bolivia

Fab Melo ?

Brazilian Basketball Player

Fabrizio Moretti ?

Drummer of The Strokes

Fabio Capello ?

Manager of the UK national Football team

„Understanding“ Content on the Web



Text: „Why snub me, Fab?“

Entity Mapping

Fabio Capello

Entity

is a

Soccer Manager

Class

is subclass of

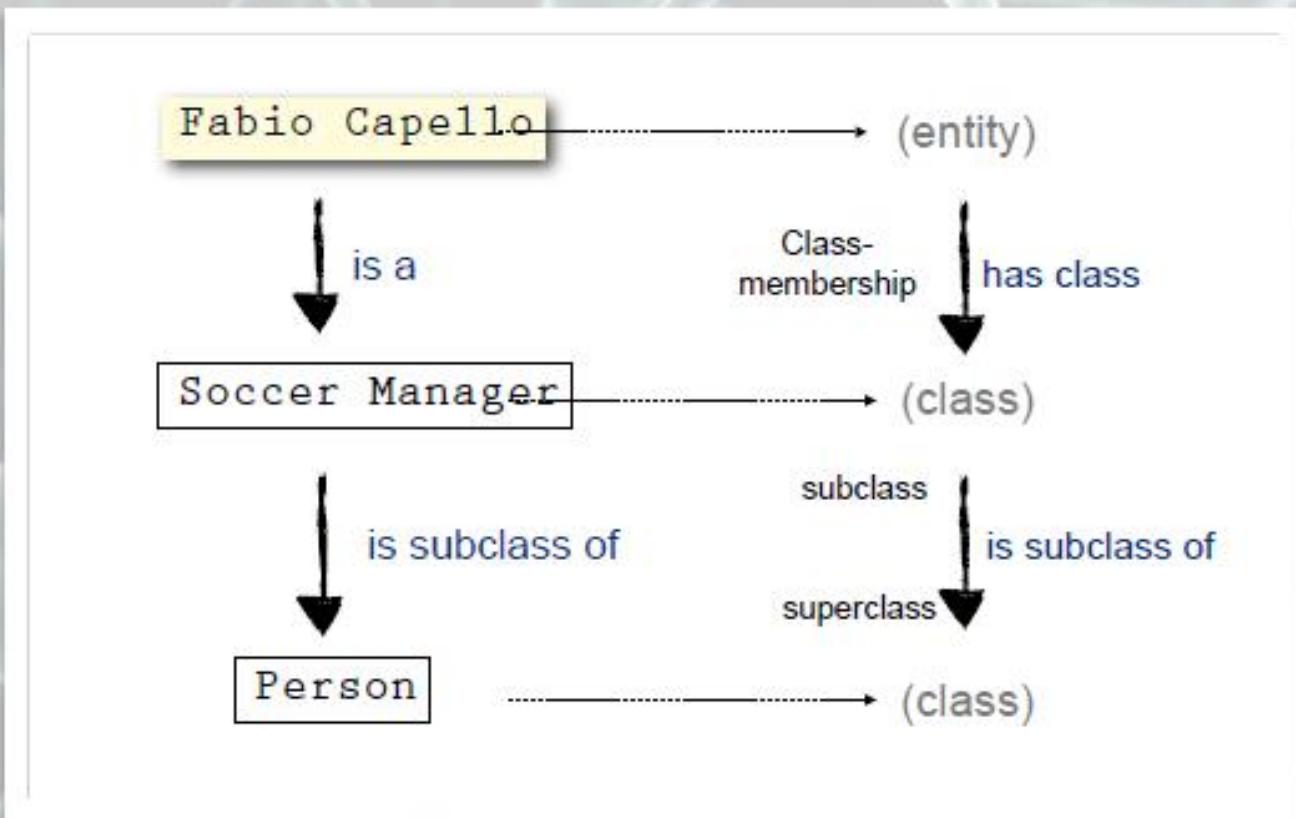
Person

Class

- The **Meaning (Semantics)** of entities and classes must be defined explicitly.

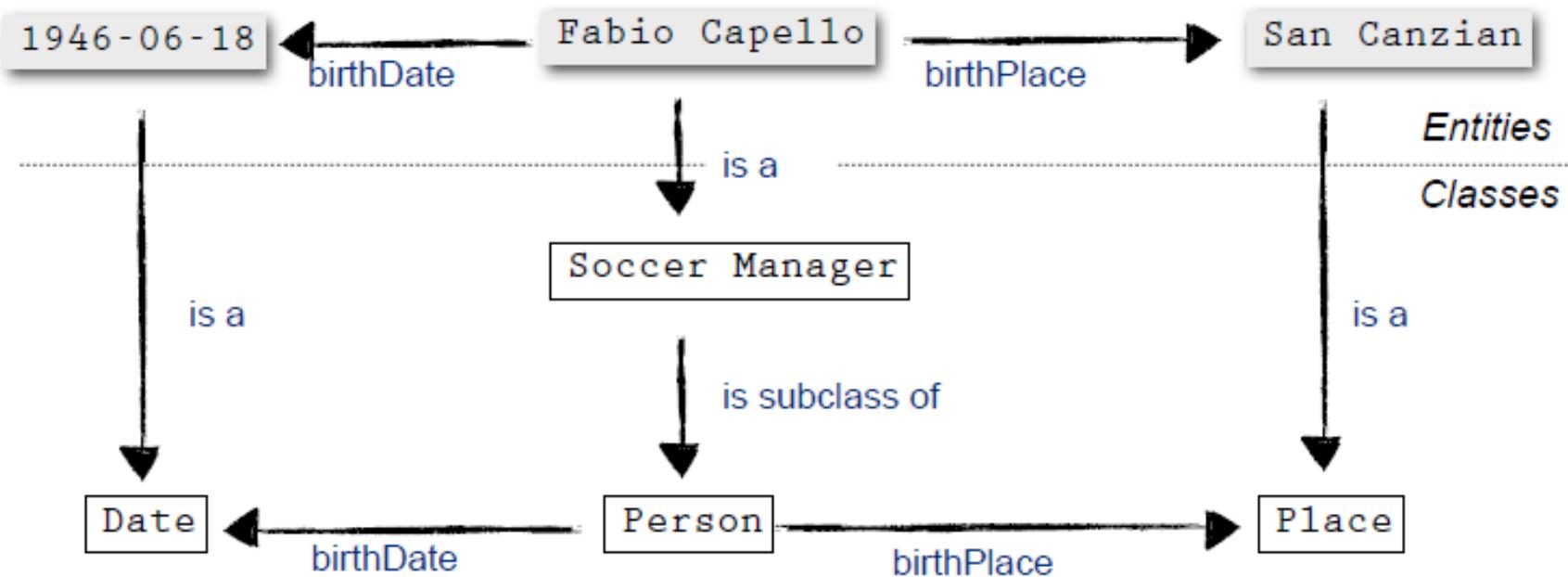
„Understanding“ Content on the Web

- The Meaning (Semantics) is expressed with the help of appropriate knowledge representations (Ontologies)

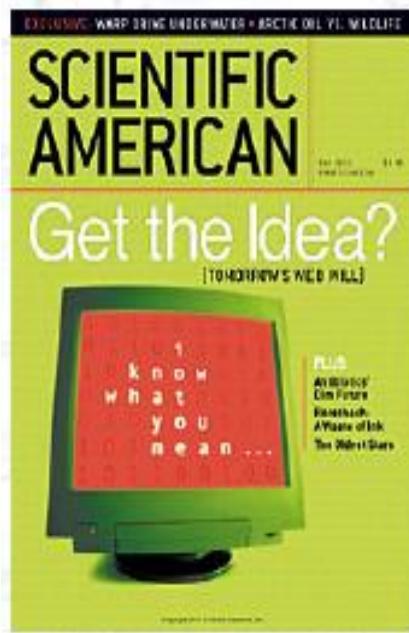
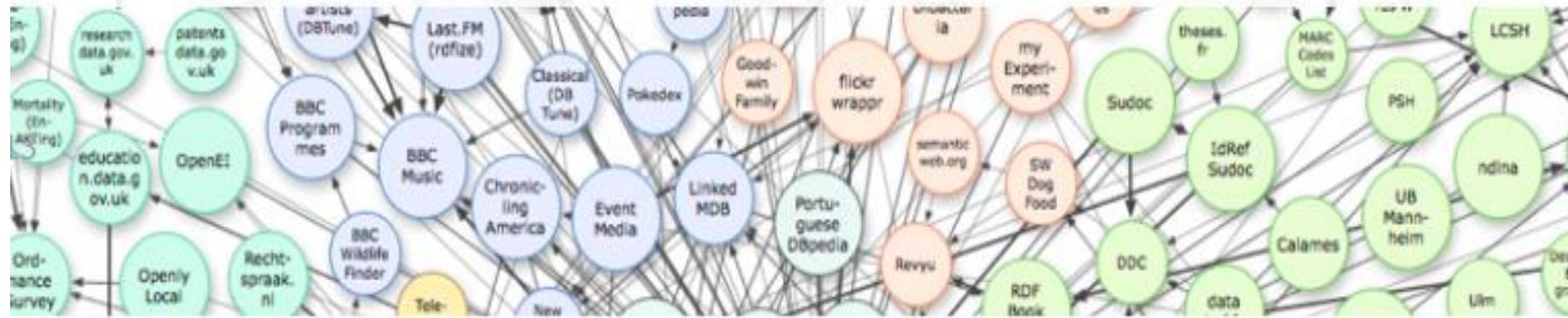


„Understanding“ Content on the Web (III)

- The Meaning (Semantics) is expressed with the help of well suited knowledge representations (Ontologies)



What is the Semantic Web?



„The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation“

Tim Berners-Lee, James Hendler, Ora Lassila: *The Semantic Web*,
Scientific American, 284(5), pp. 34-43(2001)

The Semantic Web - a Web of Data

- The Meaning of Information (Semantics) is made explicit by **formal (structured) and standardized knowledge representations (Ontologies)**.
- Thus it will be possible,
 - to process the meaning of information automatically
 - to relate and integrate heterogeneous data
 - to deduce implicit (not evident) information from existing (evident) information in an automated way

The Semantic Web is kind of a **global database** that contains a **universal network of semantic propositions**.

Semantic Web

...towards an „intelligent“ Web

The Semantic Web Technology Stack

(not a piece of cake...)

Most apps use only a subset of the stack

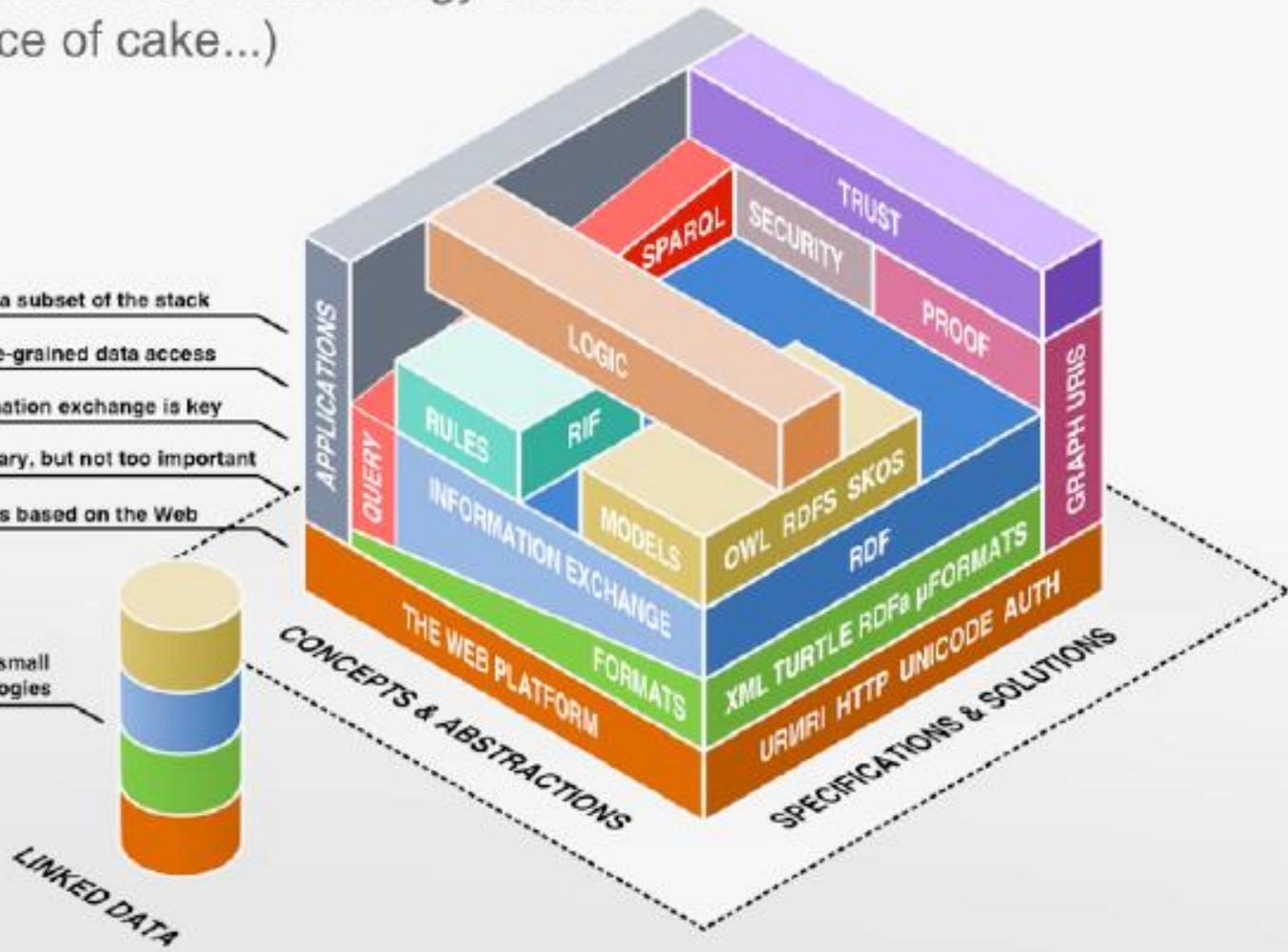
Querying allows fine-grained data access

Standardized information exchange is key

Formats are necessary, but not too important

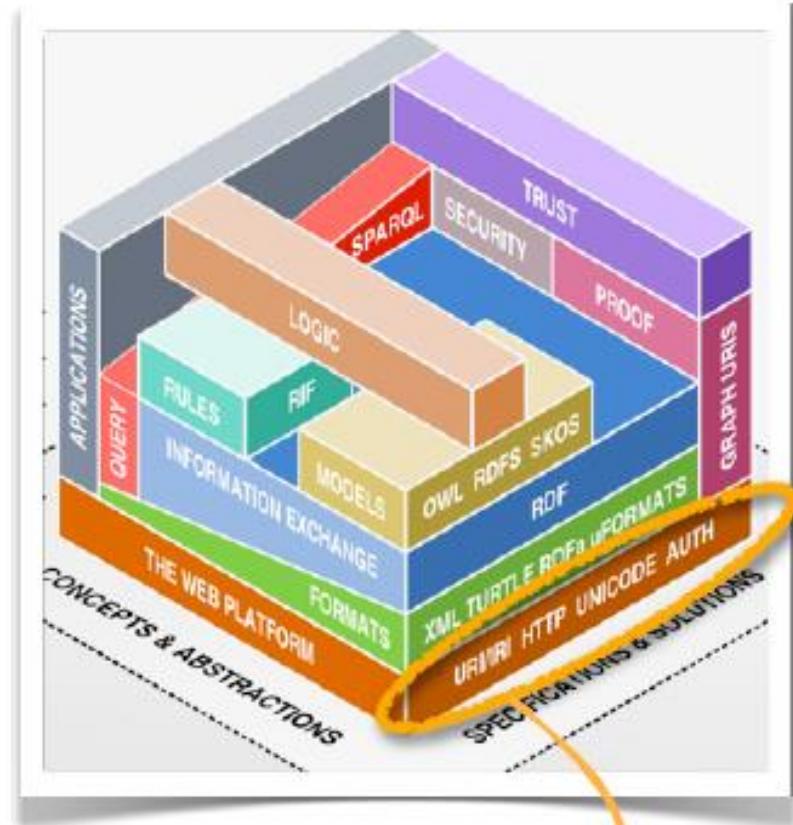
The Semantic Web is based on the Web

Linked Data uses a small selection of technologies



Semantic Web

...towards an „intelligent“ Web



URI - Uniform Resource Identifier

Fabio Capello

→ http://dbpedia.org/resource/Fabio_Capello

Semantic Web

...towards an „intelligent“ Web

→ http://de.wikipedia.org/wiki/Fabio_Capello

The screenshot shows the Wikipedia page for Fabio Capello. The sidebar on the left contains links like Main page, Contents, Pastored versions, Current events, Random article, Donate to Wikipedia, Help, About Wikipedia, Community portal, Recent changes, Contact Wikipedia, and Thesaurus. The main content area features a large portrait of Fabio Capello. Below it is a table of his career statistics, which is highlighted with a red border. The table includes columns for Years, Team, Apps, and Goals.

Years	Team	Apps	Goals
1984–1985	OFV Al	40	103
1985–1990	Hellas (managed)	82	171



WIKIPEDIA
The Free Encyclopedia

Player Information			
Full name	Fabio Capello		
Date of birth	18 June 1948 (age 61)		
Place of birth	San Cesario s'Appia, Italy		
Height	1.76 m (5 ft 9 in)		
Playing position	Middle		
Club Information			
Current club	Hellas (managed)		
National career			
1982–1984	OFV Al		
Senior career			
Years	Team	Apps	Goals
1984–1985	OFV Al	40	103
1985–1990	Hellas (managed)	82	171

RDF Version of Wikipedia

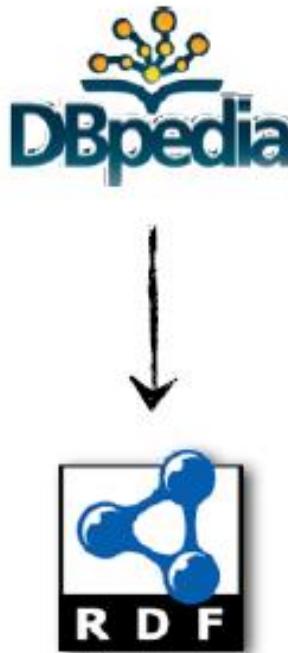
→ http://dbpedia.org/resource/Fabio_Capello

DBpedia
51

Semantic Web

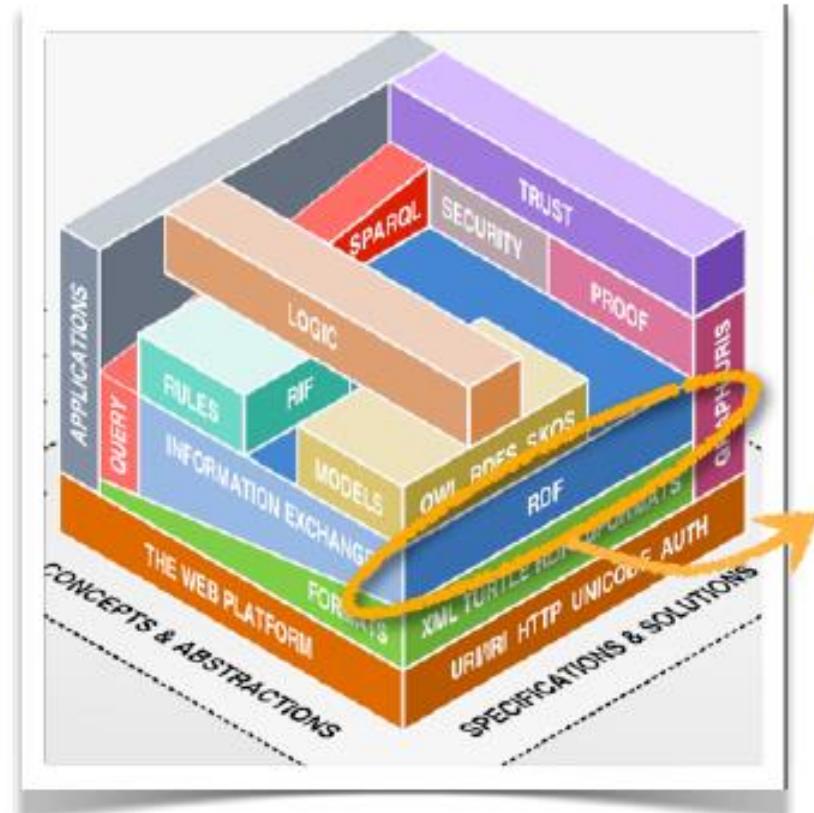
...towards an „intelligent“ Web

http://dbpedia.org/resource/Fabio_Capello



Semantic Web

...towards an „intelligent“ Web



http://dbpedia.org/resource/Fabio_Capello

```
:Fabio_Capello dbpp:birthPlace :San_Canzian .  
:Fabio_Capello dbpp:birthDate "1946-06-18" .  
. :Fabio_Capello rdfs:type dbpo:SoccerManager .  
. :Fabio_Capello rdfs:type dbpo:Person .  
...
```



RDF Resource Description Framework



RDF Tripel

:Fabio_Capello



RDF Subject

rdf:type



RDF Property

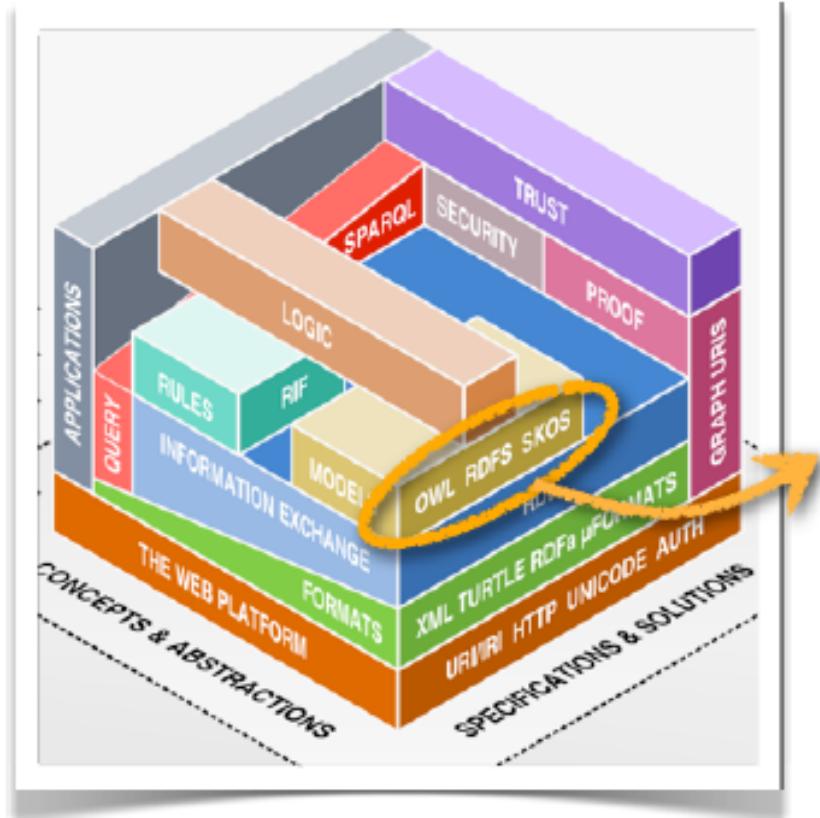
dbpo:SoccerManager .



RDF Object

Semantic Web

...towards an „intelligent“ Web

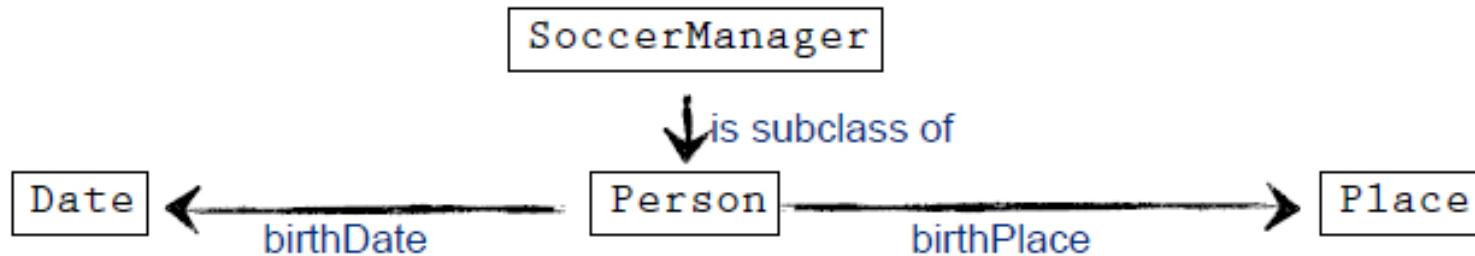


<http://dbpedia.org/ontology/SoccerManager>

```
dbpo:SoccerManager rdf:type owl:Class .  
dbpo:SoccerManager rdfs:subClassOf dbpo:Person .  
dbpo:SoccerManager rdfs:label "SoccerManager" .  
dbpp:birthPlace rdf:type rdf:Property .  
dbpp:birthPlace rdfs:domain dbpo:Person .  
dbpp:birthPlace rdfs:range dbpo:Place .  
dbpp:birthDate rdf:type rdf:Property .  
dbpp:birthDate rdfs:domain :Person .  
dbpp:birthDate rdfs:range xsd:date .  
...
```



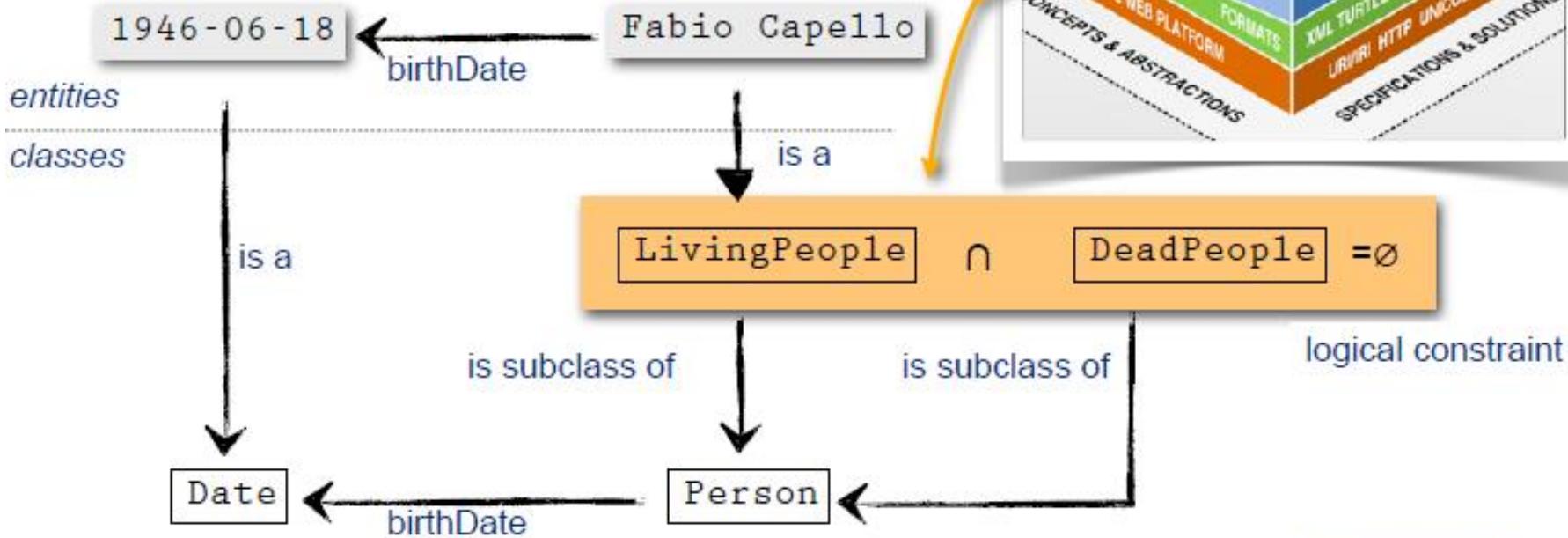
RDF Schema



Semantic Web

...towards an „intelligent“ Web

0



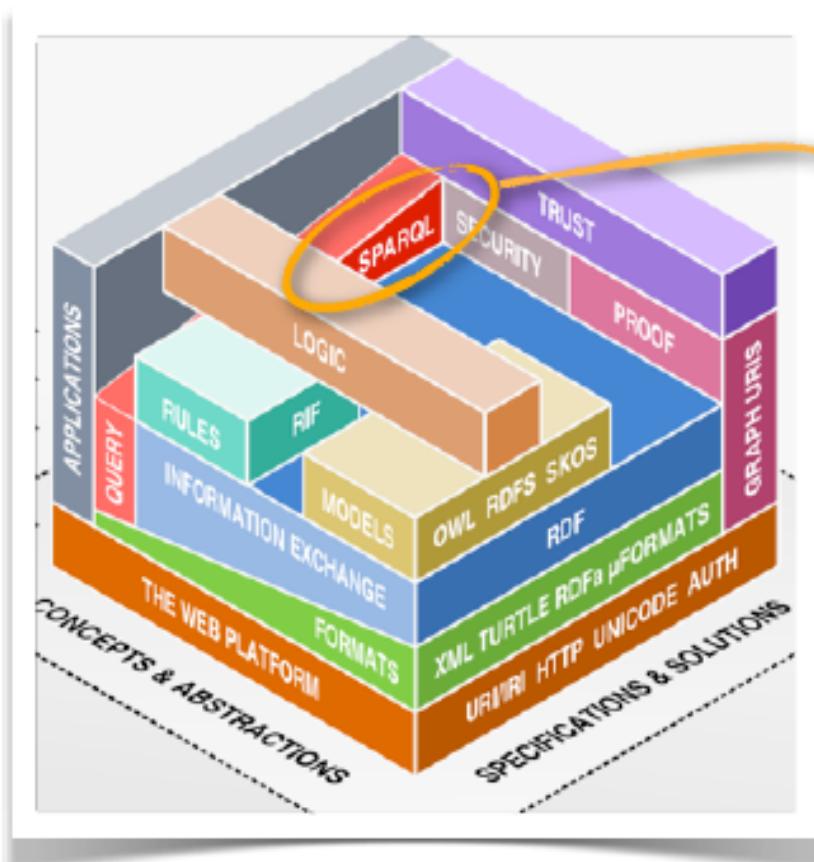
$\forall x. \exists y. \text{hasDeathDate}(x, y) \wedge \text{Person}(x) \wedge \text{Date}(y)$
 $\rightarrow \text{DeadPeople}(x)$



(Description Logics)
+ Rules

Semantic Web

...towards an „intelligent“ Web



SPARQL

Look for all Soccer Players
that have scored more than
10 goals as member of a
national team.

PREFIX dbpo: <<http://dbpedia.org/ontology>>
PREFIX dbpp: <<http://dbpedia.org/property>>

```
SELECT DISTINCT ?I ?I2 ?g
FROM <http://dbpedia.org>
WHERE {
    ?s rdf:type dbpo:SoccerPlayer .
    ?s dbpp:nationalteam ?o .
    ?s rdfs:label ?I
        FILTER langMatches( lang(?I), "EN" ) .
    ?s dbpp:nationalgoals ?g
        FILTER(?g>10).
    ?s dbpp:nationalteam ?nat .
    ?nat rdfs:label ?I2
        FILTER langMatches( lang(?I2), "EN" ) .
}
ORDER BY DESC(?g)
```

Look for all Soccer Players that have scored more than 10 goals as member of a national team.

W3C SPARQL		
Lars Christiansen (handballer)"@en	"Iceland national handball team"@en	1544
Jérôme Fernandez"@en	"Denmark national handball team"@en	1503
Stefan I Ävgren"@en	"France national handball team"@en	1284
"Wissem Hmam"@en	"Sweden national handball team"@en	1170
"Iker Romero"@en	"Tunisia national handball team"@en	745
"Alberto Entrerríos"@en	"Spain national handball team"@en	723
"Kim Andreasen"@en	"Spain national handball team"@en	667
"Lars Krogh Jeppesen"@en	"Switzerland national handball team"@en	649
"Vahid Shamsaei"@en	"Italy national handball team"@en	613
"Alessandro Rosa Vieira"@en	"Spain national futsal team"@en	602
"Raúl Entrerríos"@en	"Brazil national futsal team"@en	553
"Jorginho (beach soccer)"@en	"Brazil national beach soccer team"@en	417
"Benjamin (beach soccer)"@en	"Denmark national handball team"@en	356
"Dhanraj Pillay"@en	"Iran national futsal team"@en	349
"Victor Tomás"@en	"Brazil national futsal team"@en	336
"Buru (beach soccer)"@en	"Spain national handball team"@en	295
"Bruno Mallas"@en	"Brazil national beach soccer team"@en	288
"Christine Sinclair"@en	"Brazil national beach soccer team"@en	288
"Birgit Prinz"@en	"India"@en	280
"Mokhtar Dahan"@en	"Spain national handball team"@en	234
"Konstantin Yeryomenko"@en	"Brazil national beach soccer team"@en	201
	"Brazil national beach soccer team"@en	192
	"Canada women's national soccer team"@en	134
	"Germany women's national football team"@en	128
	"Malaysia national football team"@en	125
	"Russia national futsal team"@en	122

Problem: Data Quality....

Semantic Web Technologies

Applications on the Web of Data

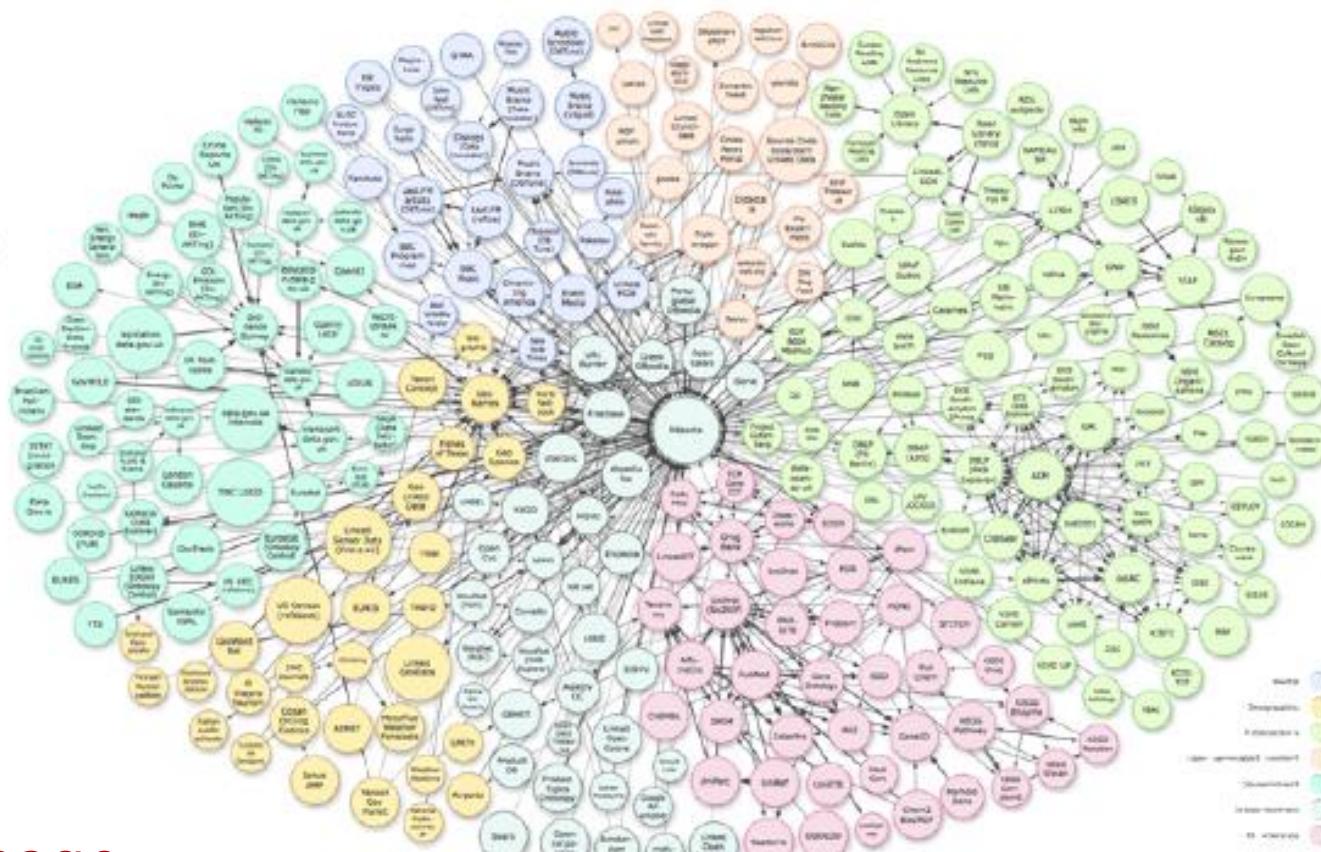
Linked Data

- Linked Open Data (LOD) denote publicly available (RDF) Data in the Web, identified via URI and accessible via HTTP. Linked data link to other data via URI.



The 'Web of Data'

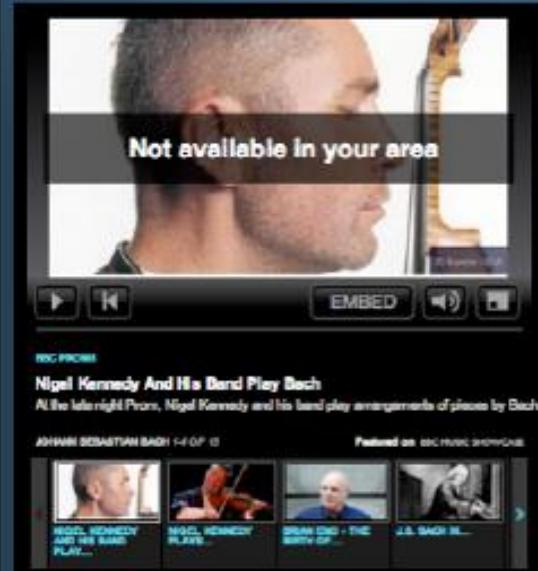
- >31 billion facts
 - >500 million links
(Oct. 2011)



숙제: Update this page

Johann Sebastian Bach

Born 21 March 1986 Died 28 May 1720



Biography

Born in the German town of Eisenach, J. S. Bach was a chorister then violinist before taking his first organist post at Arnstadt while still a teenager. It was in Weimar, as court organist from 1708, that Bach began to produce monthly cantatas, and wrote many of his great organ works, as well as organ transcriptions of concertos by Vivaldi.

In 1717 Prince Leopold offered him the position of Kapellmeister at Cöthen, where he wrote the Brandenburg Concertos, the four Orchestral Suites and the violin concertos, and married his second wife Anna Magdalena, who bore 13 children. Bach's famous duties in his final job, as Kantor of the Thomasschule in Leipzig from 1723 until his death, involved teaching Latin and Music, choir-training, and writing and directing music for the services.

Nevertheless he managed also to write the Mass in B minor, the six-choral Motet, The Art of Fugue, The Musical Offering and Goldberg Variations during this time. His inventive counterpoint became unfashionable soon after his death until the early 19th century, since when his reputation has remained unquestioned.

Profile © Edward Bhawuk

Links & Information

145

Elements of interaction and the modern interface design checklist

William F. Buckley Jr., *The Conservative Mind*

www.sagepub.com/journals

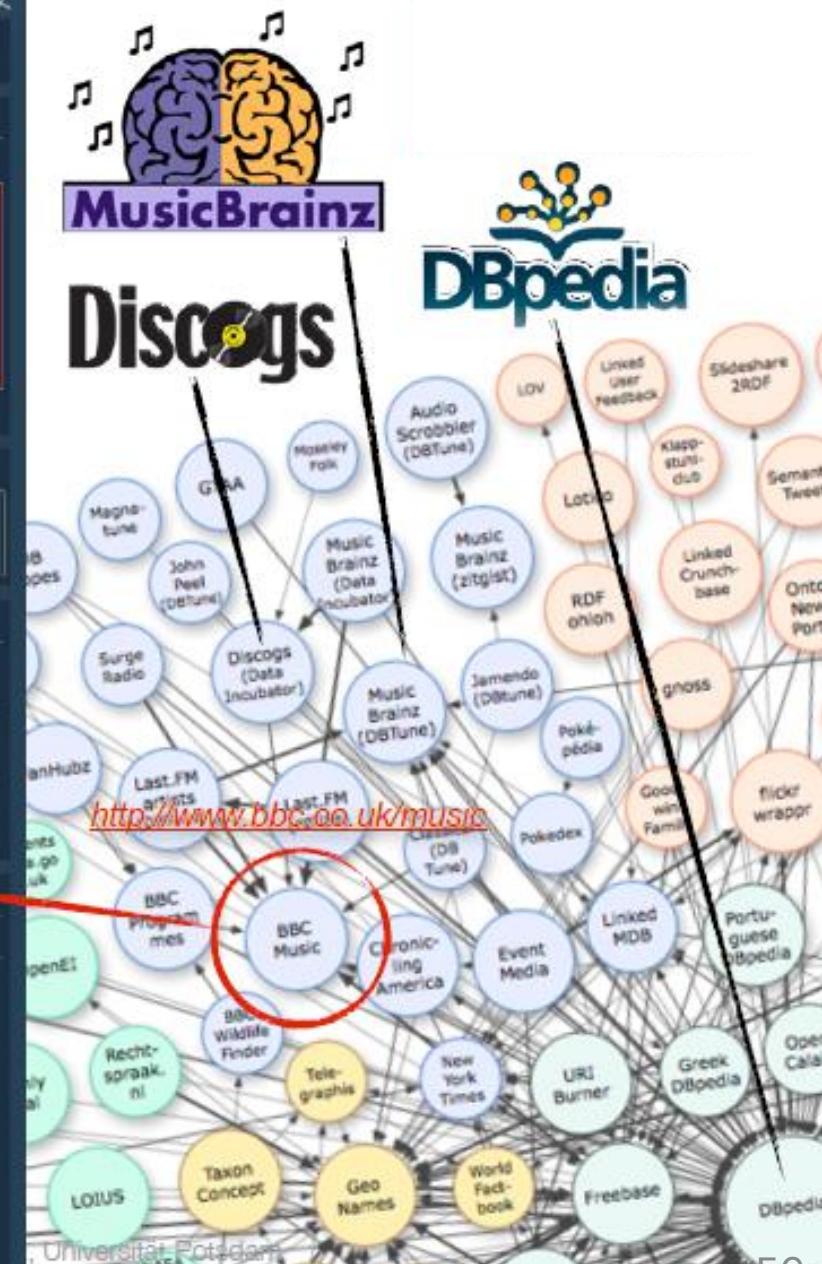
www.oriental.com

Learn more on Johann Sebastian Bach.

Discover at decogs.com/artist/07-ann-schmidt

Percent of Johann Christoph Friedrich Bach, Johann Christian Bach, Wilhelm Friedemann Bach and Carl Philipp Emanuel Bach

A screenshot of the BBC Radio 3 website's 'Composers A-Z' section. The top navigation bar includes 'Share This Page', 'Facebook', and 'Twitter' buttons. Below this, the main heading 'Now On The BBC' is displayed above a 'Composers A-Z' link and a descriptive text about exploring more composers and their works at Radio 3. The central content area features a large, bold title 'Composers A-Z' with a red '3' logo in the top right corner. Below it is a sub-headline 'Discover the music of the great composers with Radio 3'.



What is so special about the BBC Music Website?

- Information is dynamically aggregated from external, publicly available data (Wikipedia, MusicBrainz,...)
- no Screen Scraping
- no specialized API
- data available als Linked Open Data
- data access via simple HTTP Request
- data is always up-to-date without manual interaction

Search Engines - Document Retrieval

- Search Engine Query String:

“Jaguar of 2000“



- Refine „jaguar“

CAR or ANIMAL or OPERATINGSYSTEM etc. ?

Refine: **CAR**

Search Engines - Document Retrieval

- Search Engine Query String: “Jaguar of 2000“



- Name: **Jaguar E Type**
- Category: Car
 - ...
 - ConstructionYear <Integer>
 - technicalParameter
 - CylinderCapacity <Integer>
 - Power <Integer>
 - ...
 - ...

Knowledge Representation
(Ontology, Linked Data)

- Refine '2000'
ConstructionYear or technicalParameter ?
Refine: **ConstructionYear**

Search Engines - Document Retrieval

- General Problems:
 - correct interpretation of query string
 - correct identification of entities
 - automatic disambiguation
 - usability
 - personalization



Search Engines - Fact Retrieval

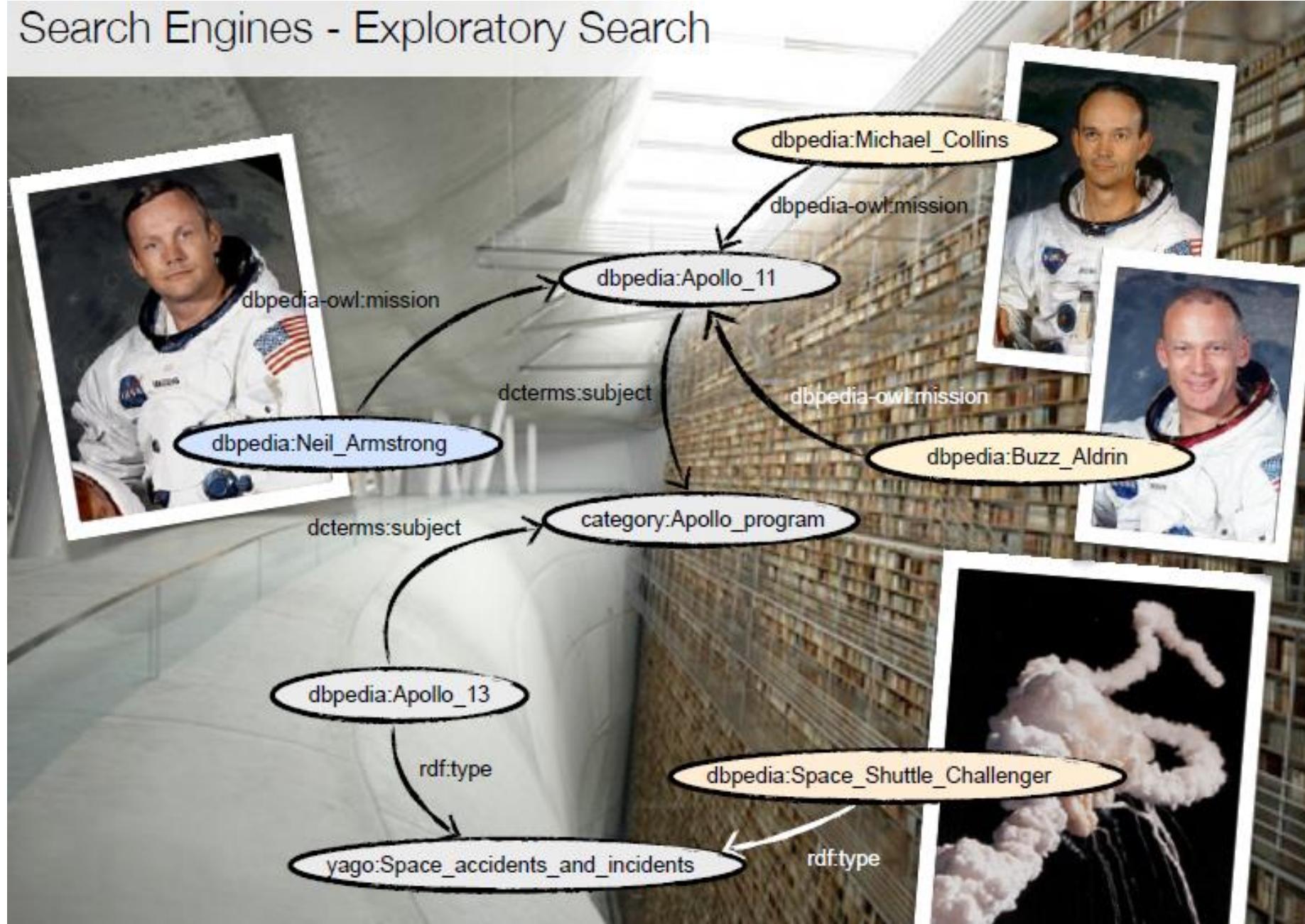
- Query String:

“Where can I fill up the tank at a considerable discount?”



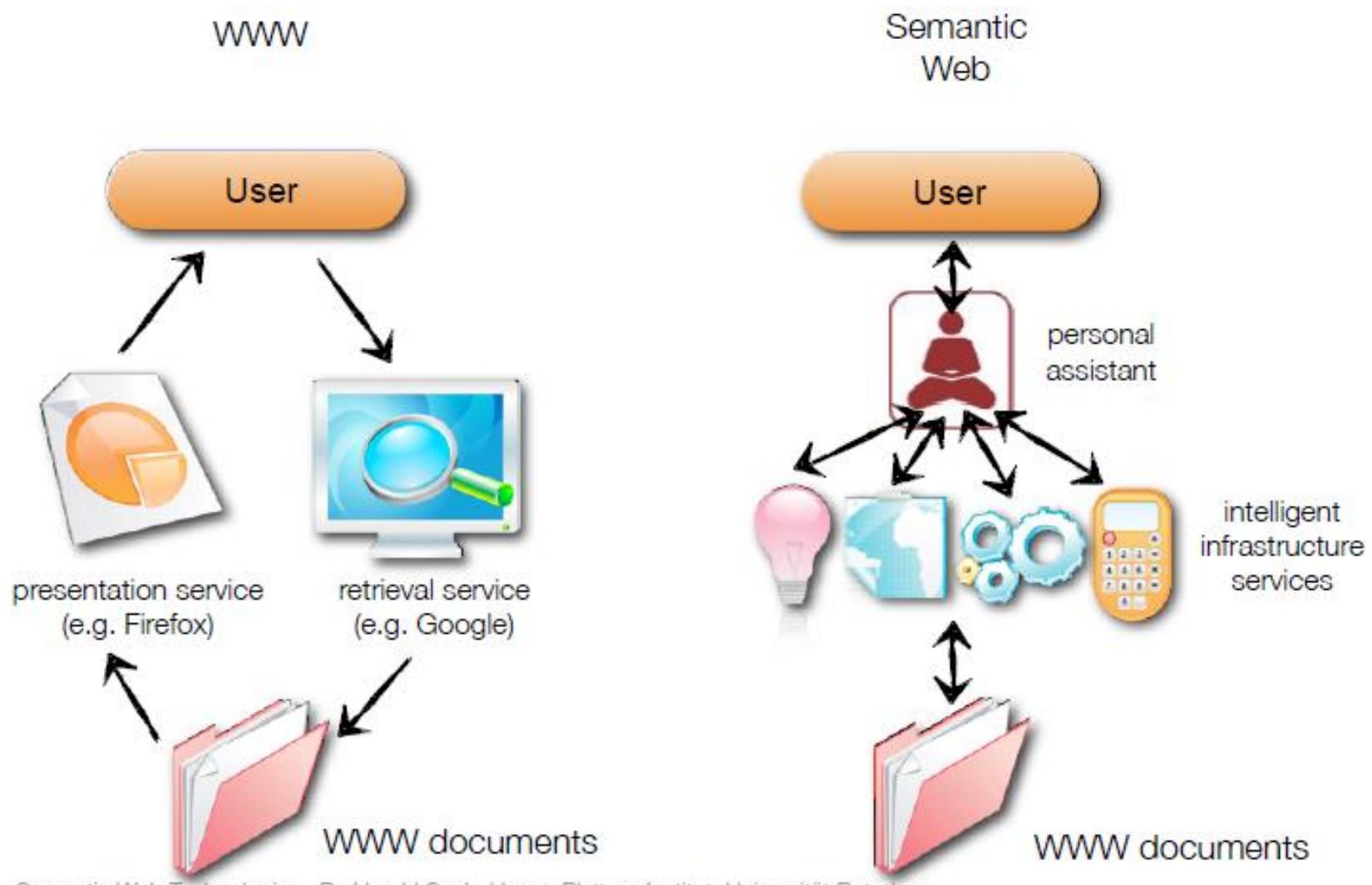
Answer :
- Hohenfelden, xy-Str. 32 -> Super leaded, 1,99 €
- fuel-efficient route will be passed to navigation
- drive only at half throttle for saving fuel...

Search Engines - Exploratory Search

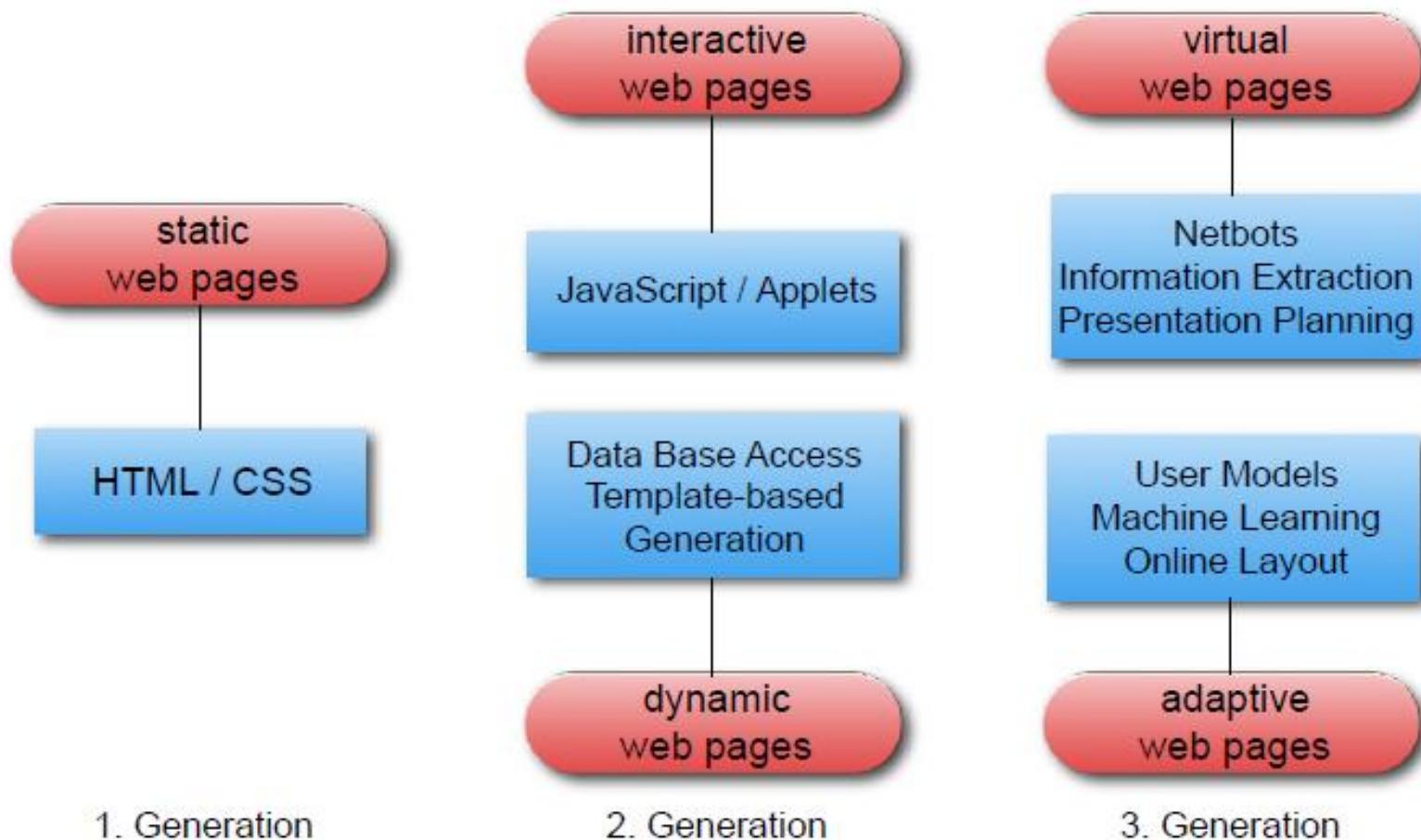


Intelligent Agents in the Semantic Web

2



3 Generations of Web Documents



1. Generation

2. Generation

3. Generation