CAS Applied Data Science - Module 1 - www | data bases

Data Acquisition and Management

PD Dr. Sigve Haug

Bern, 2019-08-23

Questions from yesterday

•••

- …

Module 1
Third day

13:30 Discussion session

14:00 Retrieving data from www
- Introduction, notebook tutorial

15:00 Break

15:30 Retrieving data from www
- Notebook tutorial
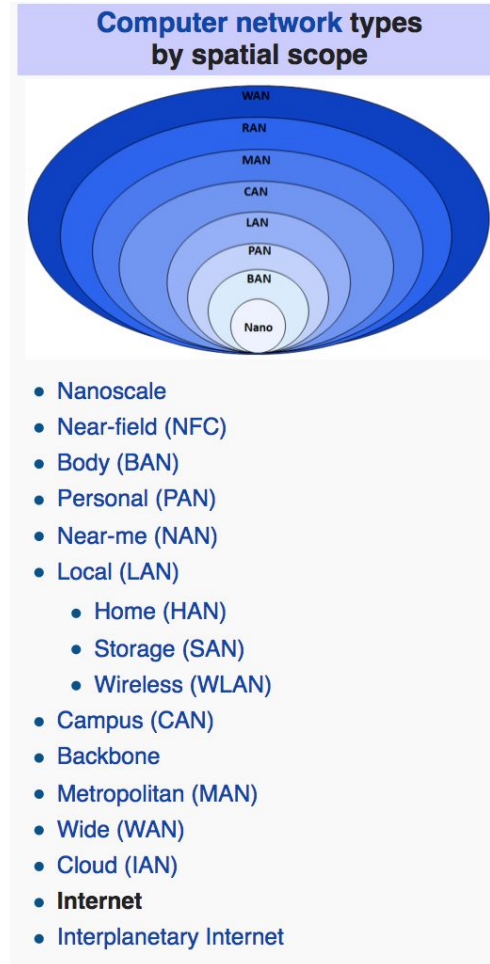- Project clarificationsl

17:00 Apero

___

Retrieving data from www

The **World Wide Web** (**WWW**), also called **the Web**, is an information space where documents and other web resources are identified by Uniform Resource Locators (URLs), interlinked by hypertext links, and accessible via the Internet.
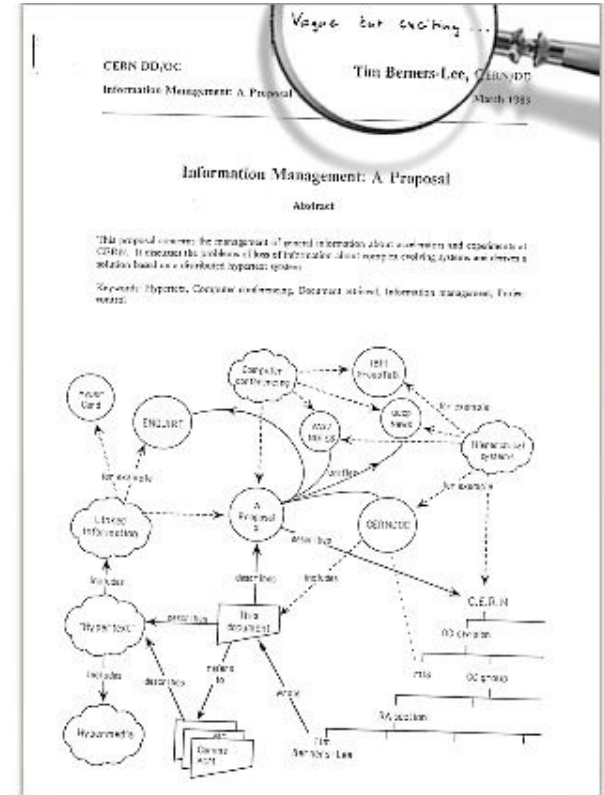
Internet

History

- Early developments in 1960s

- ARPANET 1969 (Advanced Research Projects Agency)

- Packets switching

- TCP/IP standardized in 1982

- Email, phone, file sharing … www

Computer network types by spatial scope

- Nanoscale
- Near-field (NFC)
- Body (BAN)
- Personal (PAN)
- Near-me (NAN)
- Local (LAN)
  - Home (HAN)
  - Storage (SAN)
  - Wireless (WLAN)
- Campus (CAN)
- Backbone
- Metropolitan (MAN)
- Wide (WAN)
- Cloud (IAN)
- **Internet**
- Interplanetary Internet

5

World Wide Web

History

- 1989 Tim Berners Lee proposal at CERN

  - HTML, URI/L, HTTP

- 1990 First www server and browser

- Rest is history

- https://webfoundation.org/about/vision/history-of-the-web/

World Wide Web

Usage

-
- https://www.slideshare.net/Micro-Focus/growth-of-internet-data-2017

early milestones	Key Layers of the Internet	milestones
email@-1971 Ray Tomlinson	CONTENT	1987-HyperCard Bill Atkinson
Archie-1990 Emtage & Deutsch	SEARCH ENGINE	1998-Google Brin & Page
DOS Houdini-1986 Neil Larson	BROWSERS	1993-Mosaic Marc Andreessen
(Vannevar Bush, Ted Nelson, Douglas Engelbart)	WORLD WIDE WEB	1990-http:// Tim Berners-Lee
ARPANET-1969 J.C.R. Licklider	INTERNET	1975-TCP/IP Cerf & Kahn
SAGE-1956 George Valley	NETWORKS	1973-Ethernet Robert Metcalfe
Z3-1941 Konrad Zuse	COMPUTERS	1976-Apple Jobs & Wozniak

How to access www data sources?

GUI or from shell, scripts and programs

- With application programming interface (API)

- Gives you the set of functions/methods (in python modules) to access (remote) services

- Can automate the access process

...

- Often provided by service provider

- Also "unofficial" API by "private" programmers

Let us jump to the examples (notebook of today)

CAS Applied Data Science - Module 1

Data Acquisition and Management

End of www

Bern, 2019-08-23