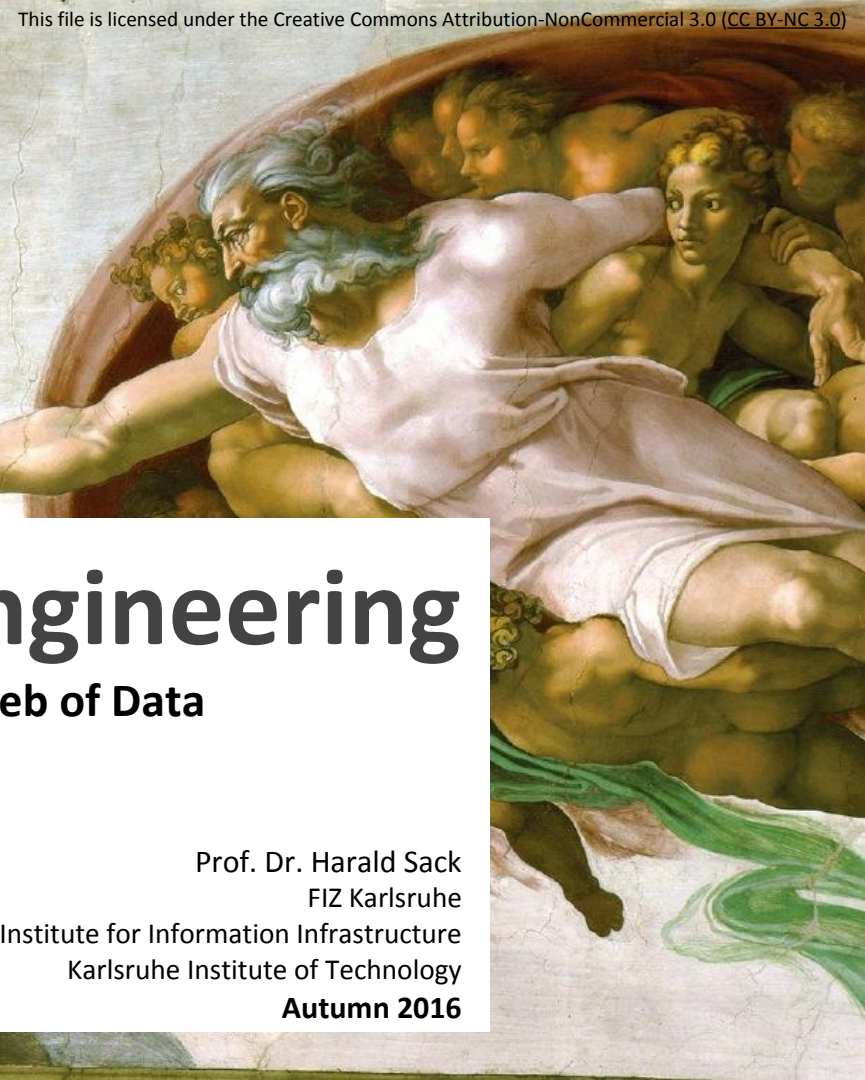




1.1 The Story of the Web so far



Leibniz Institute for Information Infrastructure
Karlsruhe Institute of Technology
Autumn 2016



J.C.R. Licklider
ARPA IPTO Director
(1962-1964)



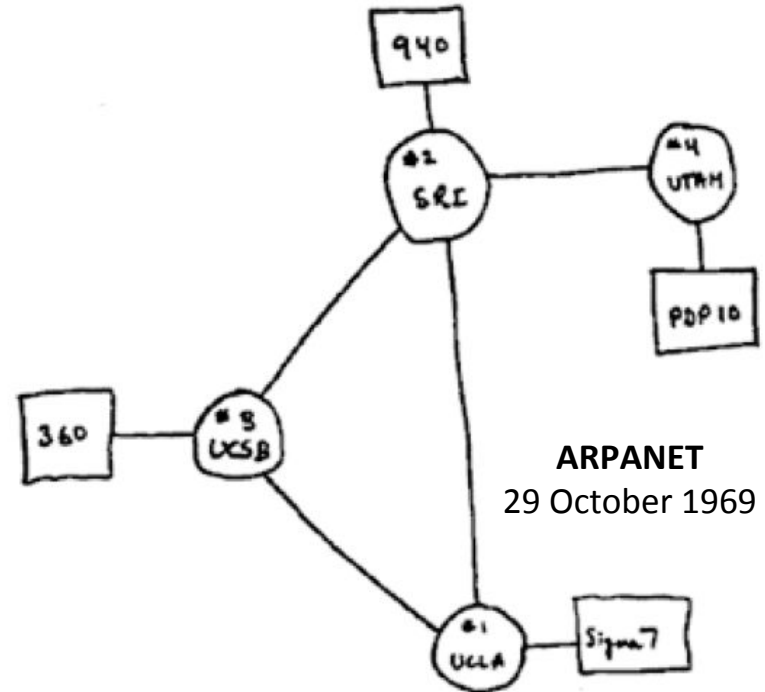
Robert Taylor
ARPA IPTO Director
(1966-1969)



Larry Roberts
ARPA IPTO Chief Scientist
(1966-1973)

"[...] we are entering a technological age in which we will be able to interact with the richness of living information - not merely in the passive way that we have been accustomed to using books and libraries, but as active participants in an ongoing process, bringing something to it through our interaction with it, and not simply receiving something from it by our connection to it."

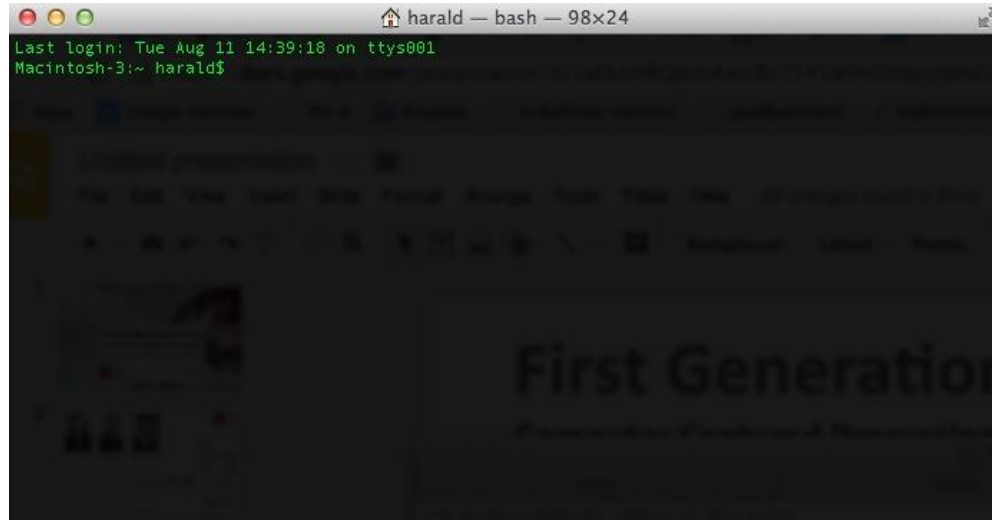
J.C.R. Licklider, Robert Taylor: [*"The Computer as a Communication Device"*](#).
Science and Technology 76, pp. 21-31 April 1968.



First Generation: The Internet

Computer Centered Processing

- How does the user get the information?



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

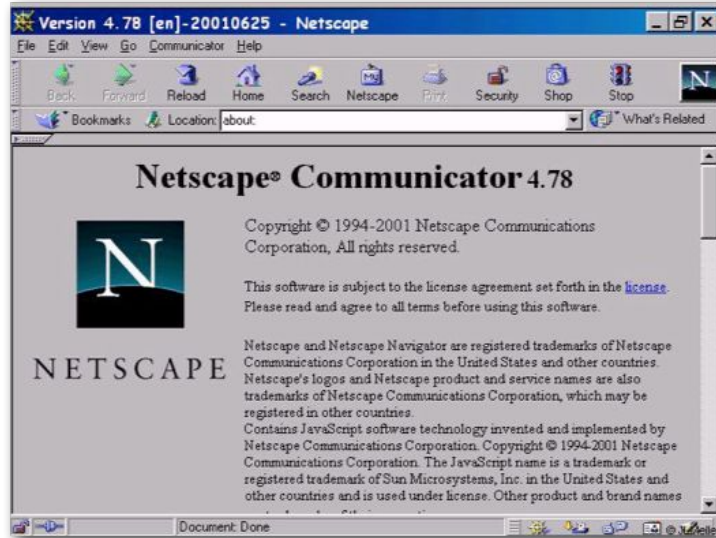
Problems

- information access requires expert knowledge and is expensive
- information retrieval is even more expensive

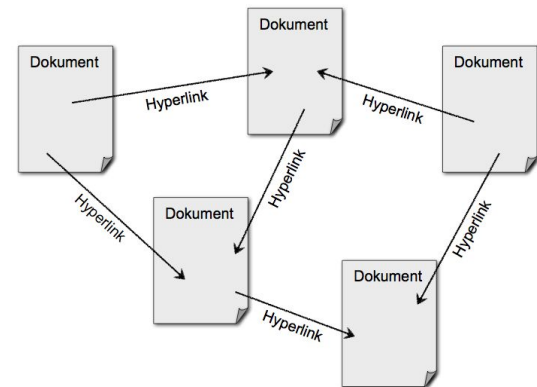
Second Generation: The Web

Document Centered Processing

- How does the user get the information?



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



Advantages

- information access does not require expert knowledge
- information retrieval via search engines

Third Generation: The Web of Data

Data Centered Processing

- The **Web of Data** is an upgrade of the Web of Documents
- It's the Web as a huge decentralised database (knowledge base) of **machine-accessible data**

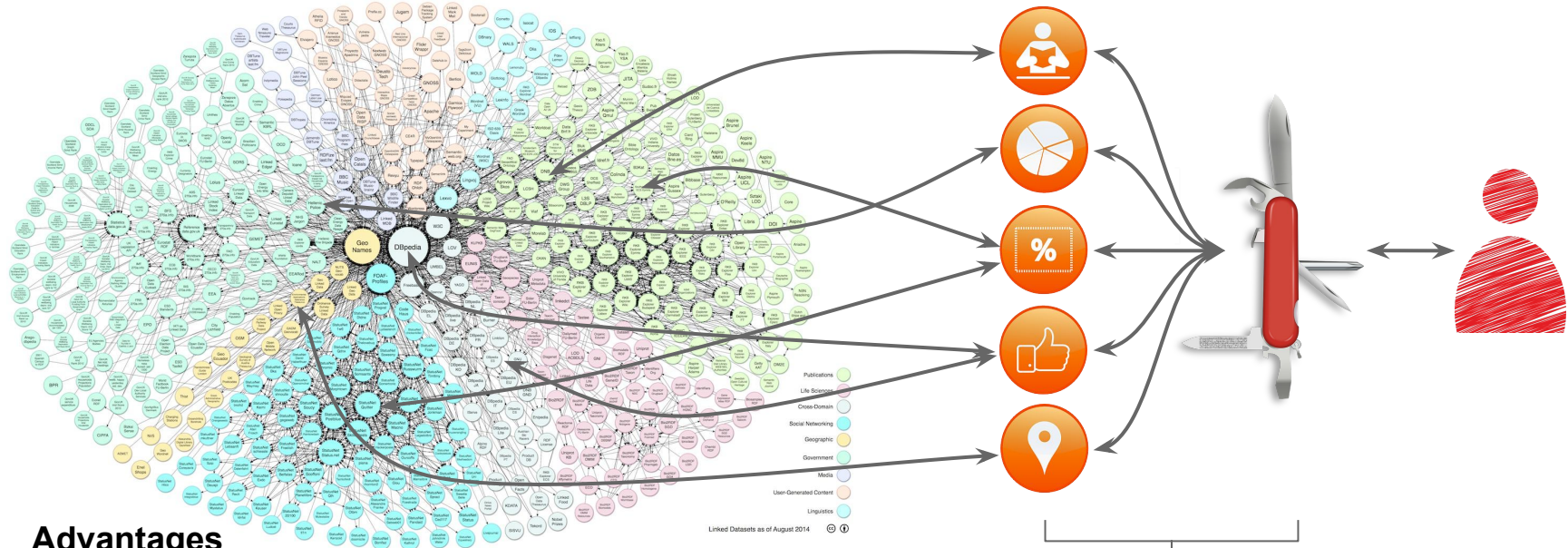
„The web of **human-readable document** is being merged with a web of **machine understandable data**. The potential of the mixture of humans and machines working together and communication through the web could be immense.“

Tim Berners-Lee, [The World Wide Web: A very short personal history](#), May 1998

Third Generation: The Web of Data

Data Centered Processing

- How does the user get the information?

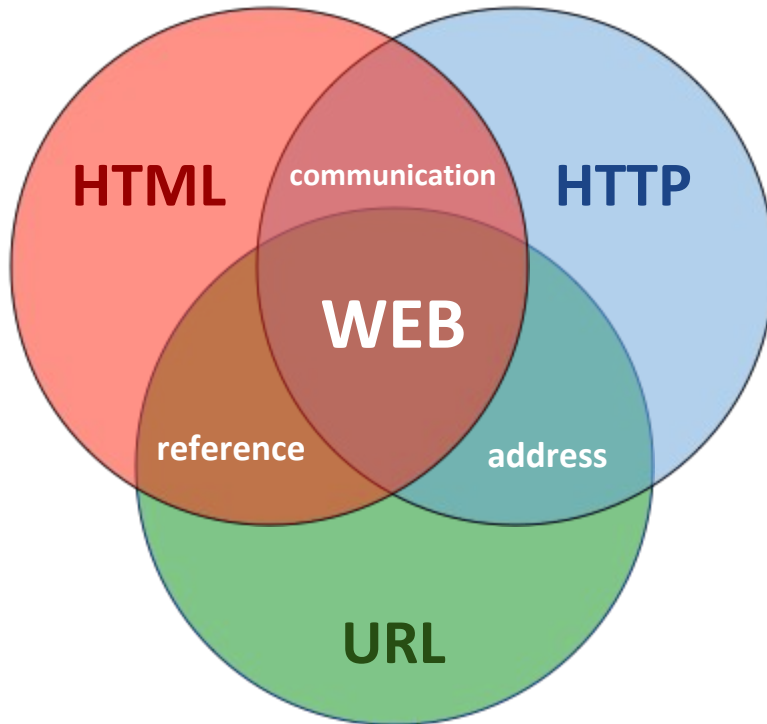


Advantages

- Information can be automatically selected, aggregated, remixed and published according to personal preferences

Information aggregation & filtering

The Architecture of the Web



1. Identification (URI) & address (**URL**)

e.g. `http://hpi.de`

2. Communication / protocol (**HTTP**)

`GET /index HTTP/2`

`Host: hpi.de`

3. Representation language (**HTML**)

`Tabea works at`

`<a href="http://hpi.de">HPI</a>.`

The (Document) Web is for 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

```
überschneiden sich thematisch. Hilf mit, die Artikel besser voneinander  
abzugrenzen oder zu vereinigen. Beteilige dich dazu an der <a  
hre  
k.2  
200  
Bau  
hre  
tit  
hre  
Q">  
tit  
(CE  
</t  
</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">
```

The Information Retrieval Dilemma



- ambiguity of natural language (polysemy)
- different words/expressions for the same concept (synonyms, metaphors)

The logo consists of the letters 'J' and 'E' in a bold, white, sans-serif font, followed by three slanted red parallel lines.

HEISENBERG CAN WAIT.

SUNDAY NIGHT FOOTBALL #SNF

Implicit Knowledge, i.e. information does not have to be specified explicitly, but must be derived via logical deductions from available information.

A close-up, slightly blurred image of a web browser's address bar. The text 'http://www' is visible in a black, sans-serif font on a light blue background. The image is tilted diagonally, and the focus is on the 'http' part, with the rest of the URL being out of focus.

http://www

In the (traditional) Web there is no explicit semantics



Next: 02 - Understanding Data on the Web

OpenHPI - Course Linked Data Engineering - Lecture 1: Linked Data and the Web of Data