

Grundlagen der Informatik

VO und KV

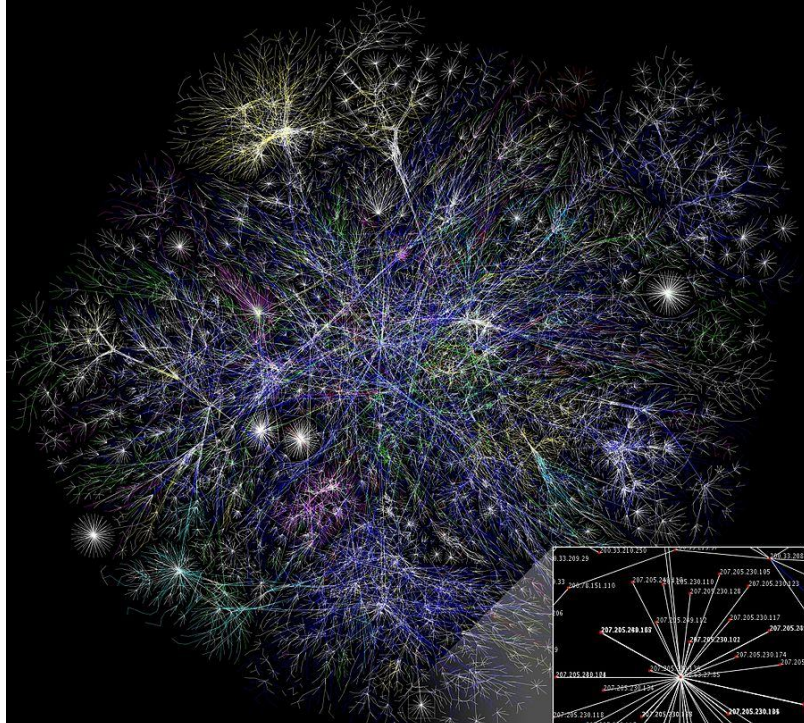
Internet und Web



Christopher Pollin | <https://chpollin.github.io>
 Zentrum für Informationsmodellierung, Universität Graz
 Digital Humanities Craft OG | <https://dhcraft.org>

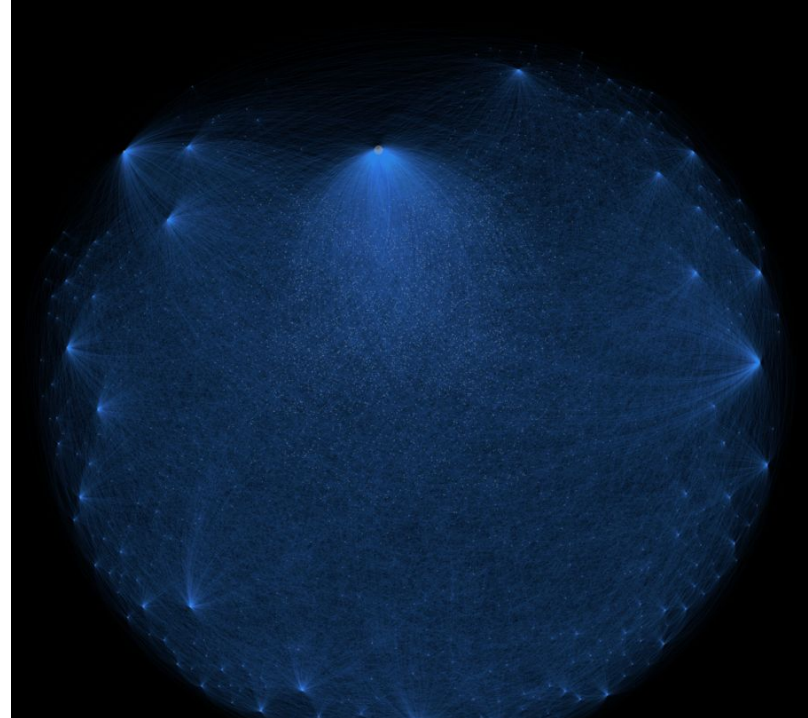
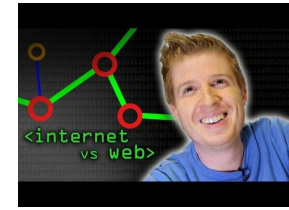


Internet



Visualisierung eines Teils des Internets

Web



Visualisierung des world wide web common crawls aus dem Jahr 2012
(44 mio. domains)

Clear Web

- Öffentliche Webseiten
- Erreichbar durch Suchmaschinen
- Für jeden sichtbar

Deep Web

- Private Datenbanken
- Medizinische Daten
- Unternehmensdaten
- Regierungsdaten
- Private Foto Cloud

Dark Web

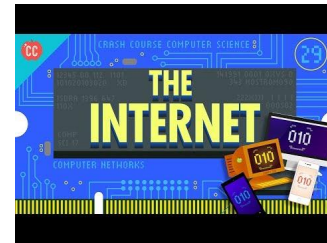
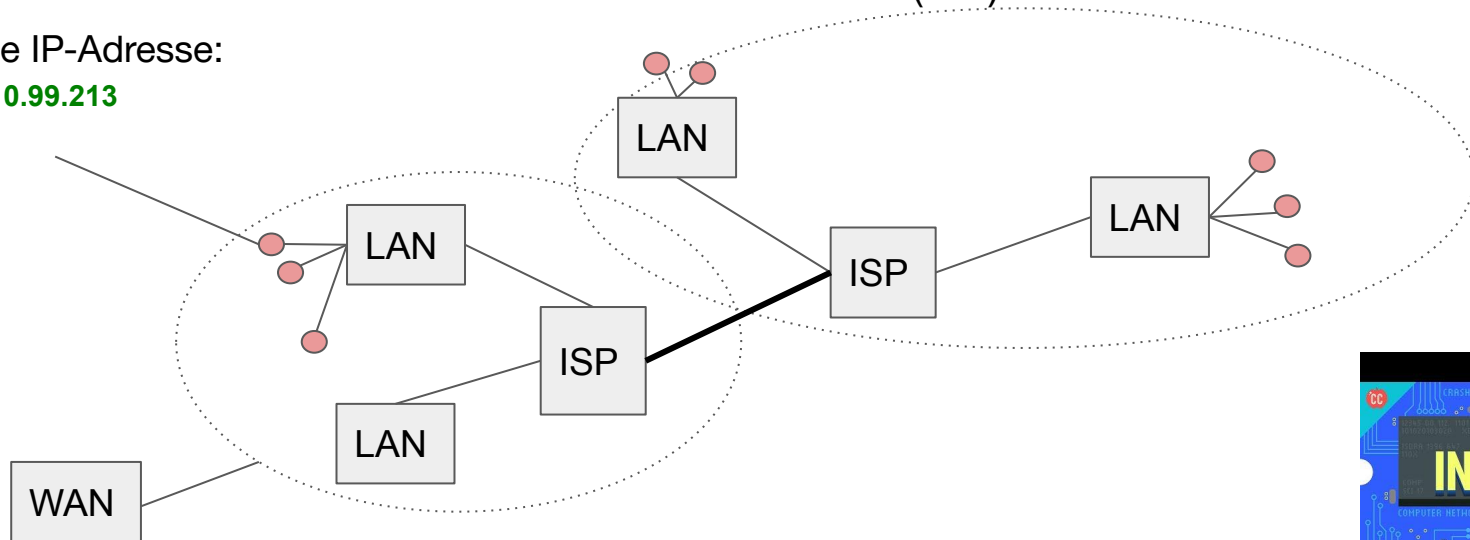
- Illegale Inhalte
- Nur über TOR erreichbar

https://de.wikipedia.org/wiki/Deep_Web

Internet: Die Infrastruktur und das Netz der Rechnernetze

- Host/Endgeräte: Rechner, Smartphone ●
- LAN (Local Area Network) ... jedes Gerät in einem Haushalt/Büro
- WAN (Wide Area Network) ... alle Netzwerke eines Internet Service Providers (ISP)

eindeutige IP-Adresse:
z.B: 80.110.99.213



<https://jvns.ca/cat.png>

Was passiert alles, wenn wir dieses Bild einer Katze im Web auf unserem Gerät anschauen wollen?

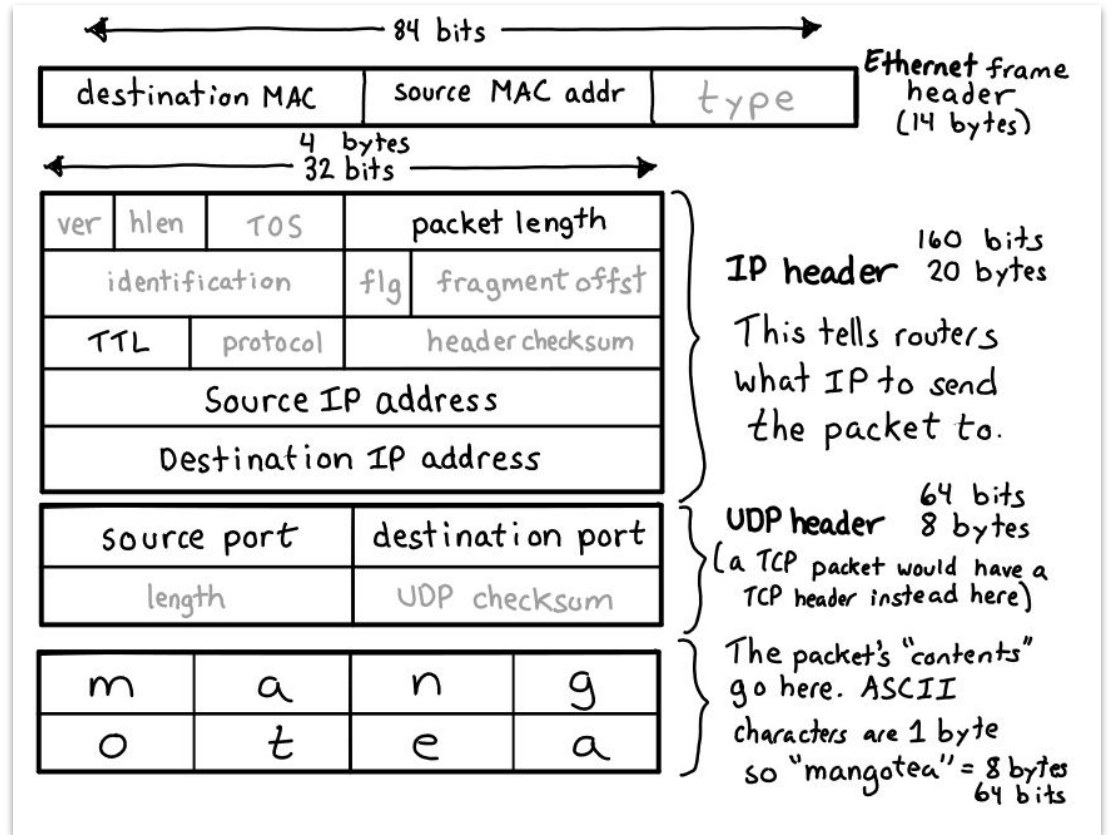
EVANS Julia, <https://wizardzines.com/#networking-ack>



Cat Ered Bread Infinite Energy -
Ace Energy

Pakete im Internet

- Daten werden über das Internet als Pakete übermittelt.
- Der Standard für diese Kommunikation ist das **Internet Protocol (IP)**
- **IP** bringt ein Paket zum richtigen Rechner im Internet
- Die Abbildung zeigt ein *UDP packet* mit dem Inhalt "mangotea"



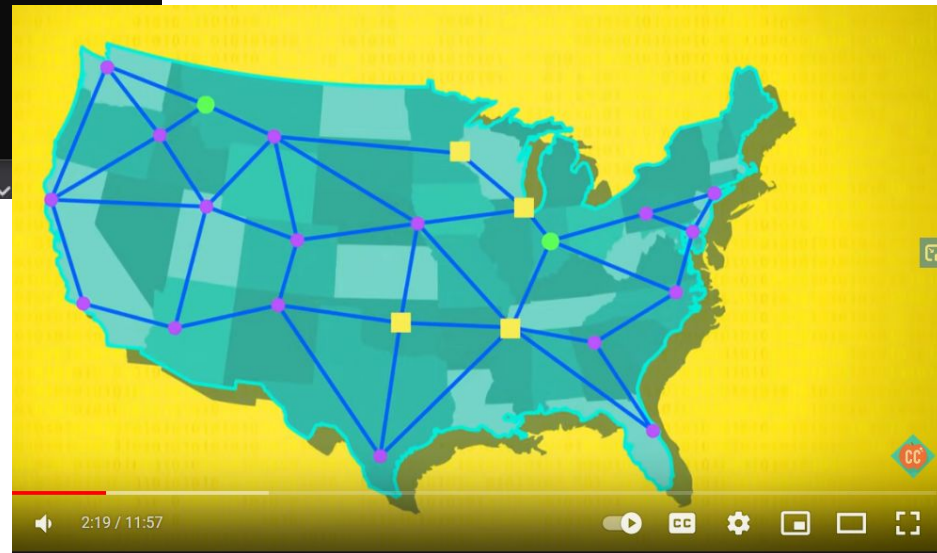
Routenverfolgung zu jvns.ca [172.67.179.90]

über maximal 30 Hops:

1	<1 ms	<1 ms	<1 ms	143.50.30.1
2	1 ms	1 ms	1 ms	143.50.223.97
3	*	*	*	Zeitüberschreitung der Anforderung.
4	1 ms	1 ms	1 ms	143.50.2.1
5	2 ms	1 ms	1 ms	143.50.11.3
6	2 ms	2 ms	2 ms	143.50.174.1
7	2 ms	2 ms	2 ms	graz2.aco.net [193.171.21.65]
8	*	*	*	Zeitüberschreitung der Anforderung.
9	5 ms	5 ms	9 ms	aconet-ias-geant-gw.vie.at.geant.net [83.97.88.1]
10	19 ms	20 ms	19 ms	ae8.mx1.mil2.it.geant.net [62.40.98.188]
11	17 ms	17 ms	17 ms	cloudflare.mix-it.net [217.29.66.167]
12	19 ms	17 ms	22 ms	188.114.100.9
13	17 ms	18 ms	17 ms	172.67.179.90

Ablaufverfolgung beendet.

Z:\>_



<https://www.ip-insider.de/was-ist-traceroute-tracert-a-681086/>

<https://www.youtube.com/watch?v=AEaKrq3SpW8&t=67s>

1. Domain Name System (DNS)

Namensauflösung von
IP/Domainnamen

2. Socket

Vom Betriebssystem bereitgestelltes
Objekt, das als
Kommunikationsendpunkt dient.

3. Transmission Control Protocol TCP

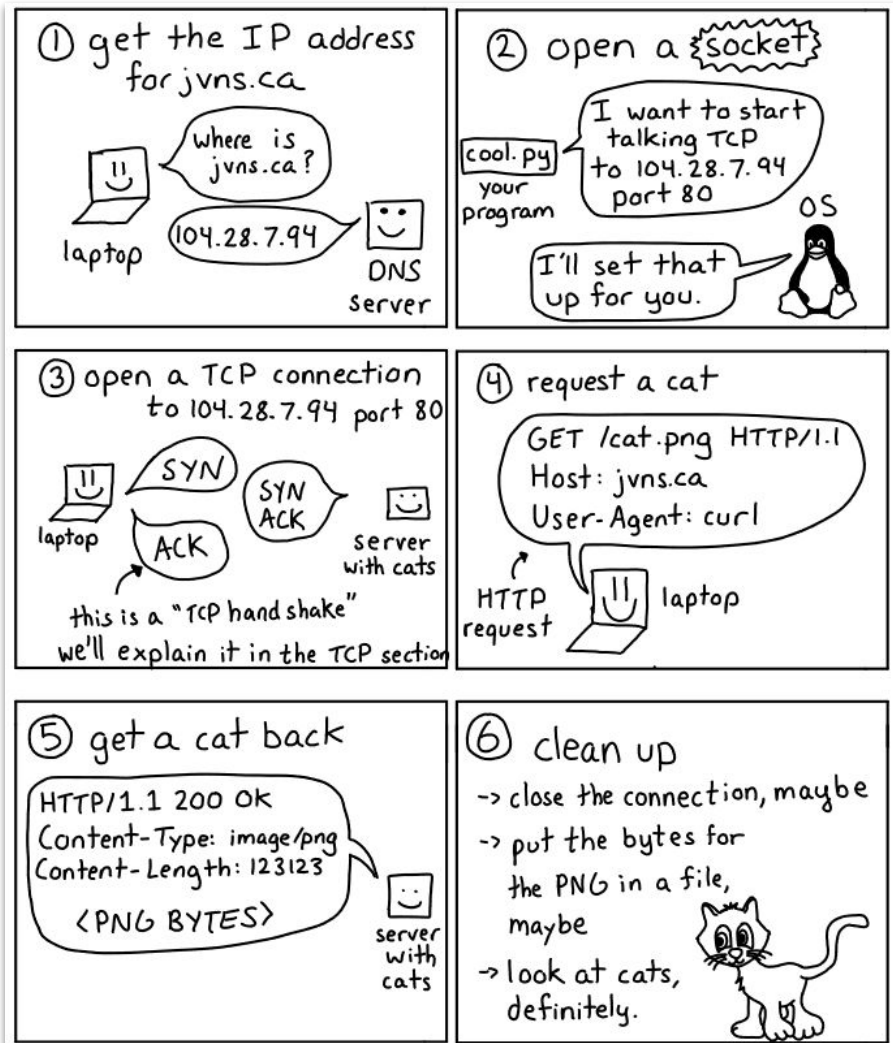
Transport Protokoll ("Regelwerk") im
Internet

4. Hypertext Transfer Protocol (HTTP) - Request

Protokoll um das Bild anzufordern

5. HTTP - Response

6. Clean up



I ♥ TCP

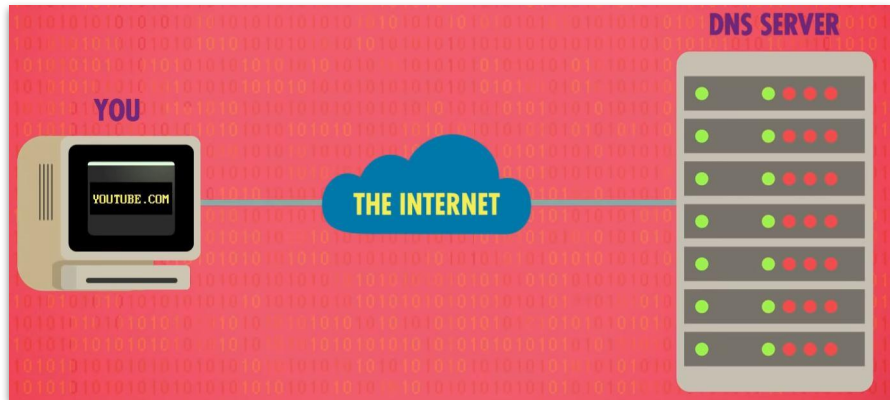


IP Adressen

Jeder Rechner ist eindeutig adressierbar über eine IP Adresse: <https://whatismyipaddress.com>

Diese IP kann man direkt in den Browser eingeben: 143.50.30.45 = <http://gams.uni-graz.at>

Das Internet hat einen speziellen Service, der diese Ziffernfolgen auf Text mappt: **Domain Name System DNS**



- Top Level Domains
.org .com, ...
- Second Level Domains
google.com, uni-graz.at
- Sub-Domain of Parent
images.google.com

Übung - Terminal: was passiert?

```
$ curl http://gams.uni-graz.at/context:szd
```

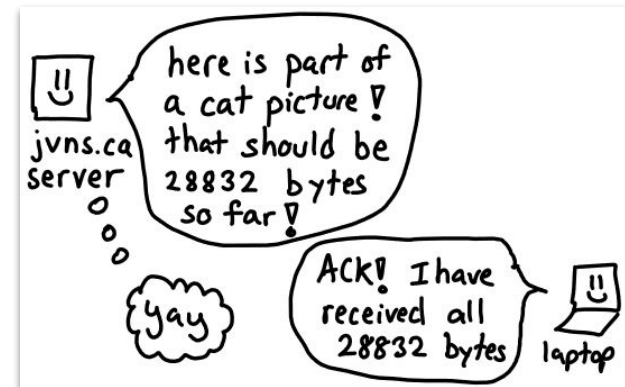
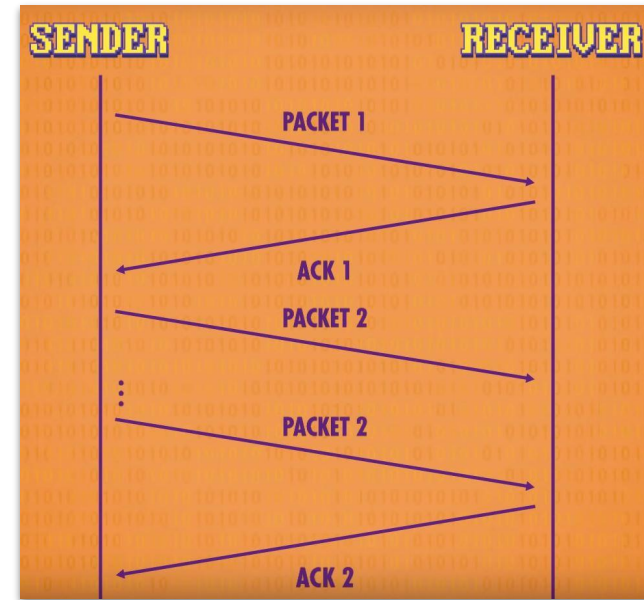
```
$ dig jvns.ca
```

<https://www.geeksforgeeks.org/curl-command-in-linux-with-examples/>

<https://mediatemple.net/community/products/dv/204644130/understanding-the-dig-command>

Transmission Control Protocol - TCP

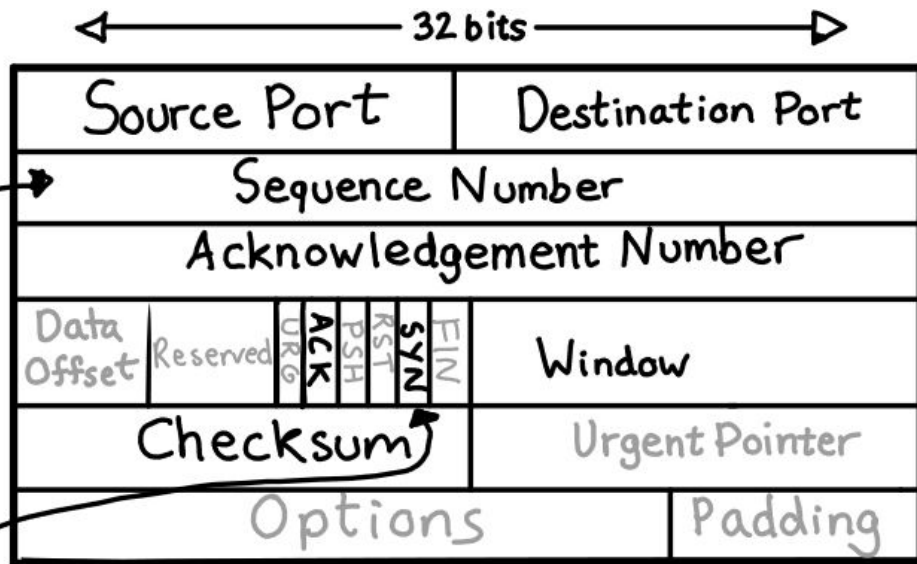
- Stellt sicher, dass ein Paket richtig ankommt.
- Paket 34, folgt auf Paket 35, folgt auf...
- Jedes Paket kann einen eigenen Weg gehen.
- Es werden ganz viele gleiche Pakete versendet.
- ACK ... Acknowledgement/Bestätigung
- Keine ACK: nach gewisser Zeit Wiederholung der Sendung des Pakets.
- Receiver hat Sequence Number, kann also alles wieder zusammenbauen.



This is what a TCP header looks like:

the "sequence number" lets you assemble packets in the right order 😊

this is the SYN bit



Every TCP connection starts with a "handshake". This makes sure both sides of the connection can communicate with each other.

Hyper Text Transfer Protocol

HTTP

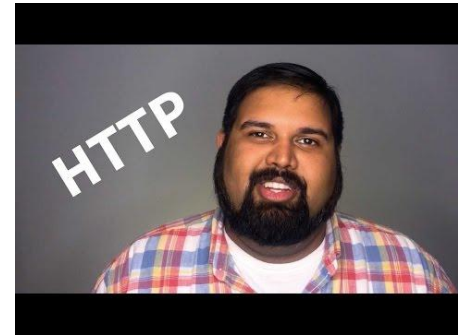
HTTP - Hyper Text Transfer Protocol

- Am Anfang nur HTML Dokumente an Clients gesendet
- Request - Response Zyklus
 - URL im Browser eintippen
 - HTTP Request über das Internet
 - HTTP Response



HTTP - Hyper Text Transfer Protocol

- Interaktion zwischen Browser und Server erfolgt mittels:
der 'Messenger des Web'
- Protokoll auf Anwendungsschicht, basiert auf TCP/IP
- Interaktion erfolgt lediglich in Form eines einfachen Frage/Antwort-Verfahrens
Client macht einen **Request**
Server gibt **Response**
- **Protokoll ist zustandslos**
Server 'weiß' nicht ob Client schon einmal gefragt hat.
Es entstehen einfache Kommunikationszyklen.



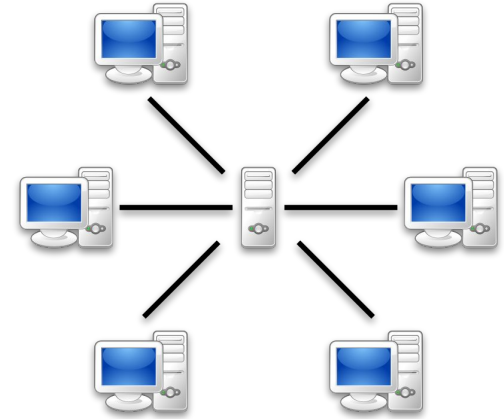
Client - Server Architektur

Was ist ein Client?

- Eine Maschine oder ein Programm, das Anfragen/Request stellen kann.

Was ist ein Server?

- “Serve other programs called clients”
- Ein Server kann gleichzeitig mit mehreren Clients agieren
- Unterschiedliche Servertypen
 - Web Server - Apache: HTTP requests
 - Datenbank Server
- Ein Server beherbergt Web Ressourcen, hosted Web Applikationen, speichert Userdaten und Programme



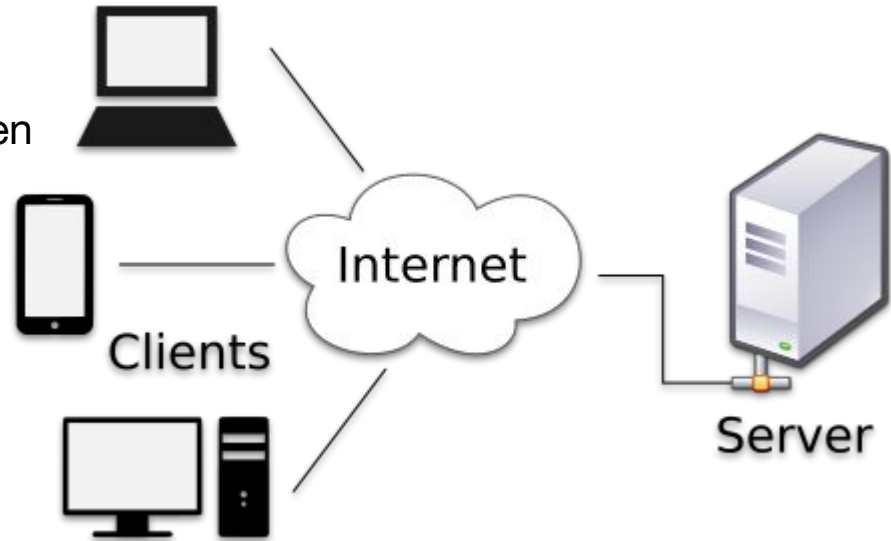
Client - Server Architektur

Benutzer fordert über den Browser (Client) ein Dokument im Web an

Browser kontaktiert den in der URL spezifizierten Server und fordert gewünschtes Dokument an

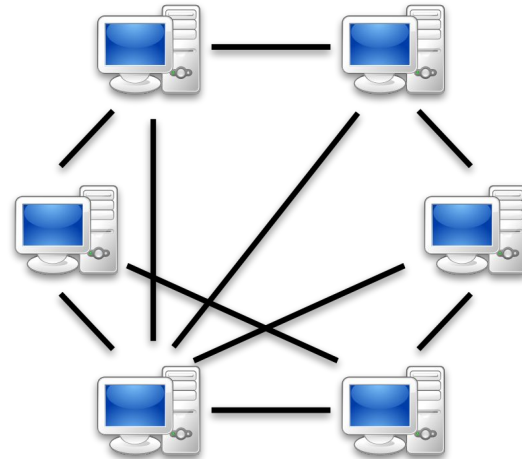
Server greift auf sein lokales Filesystem zu und sendet (in URL) spezifizierte Datei als Anfrage an den Browser

Browser	{	empfängt	Dokument
		interpretiert	
		stellt dar	
		analysiert	
		weist zu	HTML Struktur
			CSS-Anweisungen



Peer-To-Peer Model

- Alternative zu Client - Server Architektur (zentralisiert)
- Peer-To-Peer ist dezentralisiert
- Computer sind sowohl Client, als auch Server
- Beispiele: Torrents oder Skype



HTTP - Hyper Text Transfer Protocol

GET Anfordern von Server-Daten

POST Datenübertragung an Server

HTTP message sind plain text

Startline: GET /http/1.1

Header

Body

Trying 143.50.30.45
Connected to gams.uni-graz.at

Verbindungsaufbau
zum Server

GET / http/1.1
Host: gams.uni-graz.at

HTTP-Anfrage

```
HTTP/1.1 200 OK
Date: Mon, 23 May 2005 22:38:34 GMT
Content-Type: text/html; charset=UTF-8
Content-Encoding: UTF-8
Content-Length: 138
Last-Modified: Wed, 08 Jan 2003 23:11:55 GMT
Server: Apache/1.3.3.7 (Unix) (Red-Hat/Linux)
ETag: "3f80f-1b6-3e1cb03b"
Accept-Ranges: bytes
Connection: close
```

Serverantwort: Header

```
<html>
<head>
  <title>An Example Page</title>
</head>
<body>
  Hello World, this is a very simple HTML document.
</body>
</html>
```

Serverantwort: Body

Verbindungsende

HTTP Methods

GET:	URI + Server Anforderung.
POST:	schickt unbegrenzte Mengen an Daten zur weiteren Verarbeitung zum Server.
HEAD:	wie GET, aber nur um die Metadaten zurück zu bekommen.
PUT:	Hochladen auf Webserver, Resource ersetzen oder anlegen
PATCH:	Ändert ein bestehendes Dokument ohne dieses wie bei PUT vollständig zu ersetzen.
DELETE:	löscht die angegebene Ressource auf dem Server.
TRACE:	Überprüft, ob und wie die Anfrage auf dem Weg zum Server verändert worden ist.
OPTIONS:	liefert eine Liste der vom Server unterstützten Methoden und Merkmale.
CONNECT:	wird von Proxyservern implementiert, um SSL-Tunnel zur Verfügung zu stellen.

URI - URL - URN

URI - URL - URN

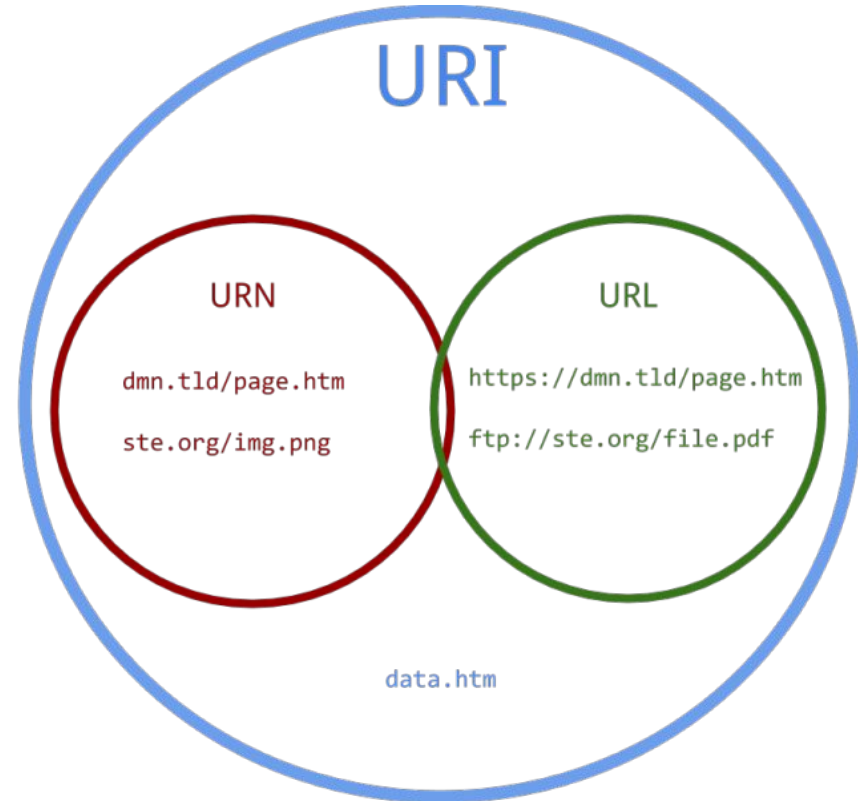
Zur weltweit eindeutigen Identifikation und Adressierung von Dokumenten im Web dienen...

URI - Uniform Resource Identifier

URN - Uniform Resource Name

URL - Uniform Resource Locator

Verbreitetste Ausprägung der URIs sind URLs



Anforderungen an URIs

Universalität

Jede Ressource im Internet soll unabhängig vom Informationsdienst ansprechbar sein.

Eindeutigkeit

Jede Ressource muss weltweit eindeutig identifizierbar sein

Erweiterbarkeit

Fixierbarkeit

URI sollen nicht nur auf elektronischem Weg austauschbar, sondern auch manuell bearbeitbar oder druckbar sein

Bestandteile eines URI

Präfix

Spezifiziert Informationsdienst/-Protokoll

Suffix

Spezifiziert Lage und Identifikation

ftp://max:muster@ftp.example.com

https://open.hpi.de

http://open.hpi.de

rtsp://webradio.com/stream

URI vs. URL vs. URN

URL: `ftp://ftp.is.co.za/rfc/rfc1808.txt`

URL: `http://www.ietf.org/rfc/rfc2396.txt`

URL: `ldap://[2001:db8::7]/c=GB?objectClass?one`

URL?: `mailto:John.Doe@example.com`

URL?: `news:comp.infosystems.www.servers.unix`

URL: `telnet://192.0.2.16:80/`

URN (keine URL): `urn:oasis:names:specification:docbook:dtd:xml:4.1.2`

URN (keine URL): `tel:+1-816-555-1212`

Zur Zeit werden URNs kaum unterstützt.

Eines der wenigen Beispiele:

Deutsche Nationalbibliothek: <http://nbn-resolving.de>

URI: Zusammenfassung

- Ist eine Zeichenkette zur Identifikation von Ressourcen im Internet.
- Identifikation erfolgt über Name **URN**, über Ort **URL** oder beides zusammen
- URL: **Protokoll (Prefix)** + **Domain (Sufix)**
 - Hinter Domain steckt eine IP-Adresse 111.222.33.444
 - <http://mywebsite.com:80/register.php>
 - :80 ... ist ein Port: Verbindungspunkte für TCP/IP
- **Path**
<http://mywebsite.com/products/register.php>
- **Query Strings**: um Parameter zu übergeben
<http://mywebsite.com/signup?fn=Fasching&Dienstag>
- **Fragments**: zur Adressierung auf einer Webseite
<http://mywebsite.com/data#Bier>



URI's, URL's and URN's
URL explained!

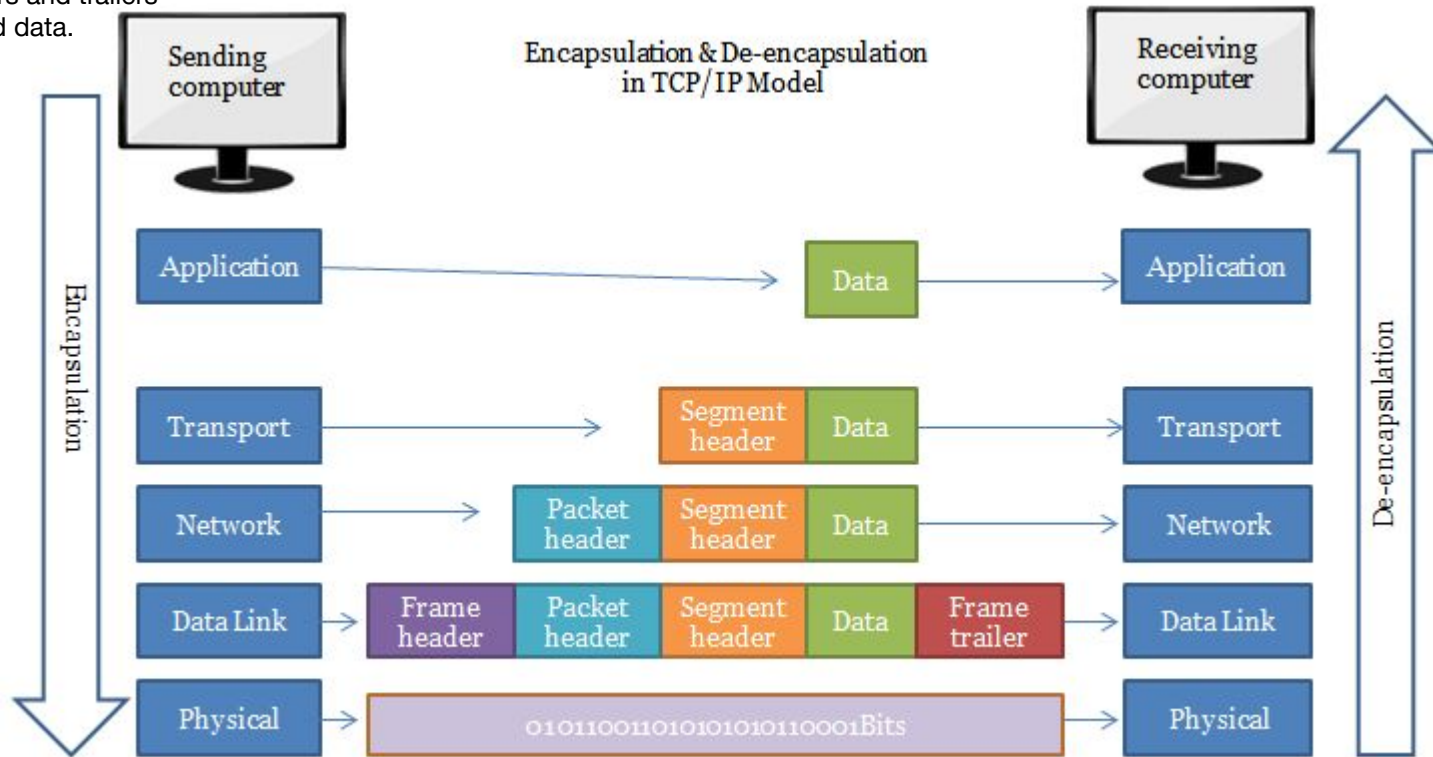
OSI - Open Systems Interconnection Model



Um zum Katzenbild zu kommen müssen wir die Schichten des
OSI - Open Systems Interconnection Model durchgehen

7	Anwendungsschicht	(application layer)	SMTP, FTP, HTTP ... Web
6	Darstellungsschicht	(presentation layer)	Format Data, Encryption, SSL
5	Sitzungsschicht	(session layer)	Start & Stop Session, API
4	Transportschicht	(transport layer)	TCP, UDP: Pakete transportieren
3	Vermittlungsschicht	(network layer)	IP Address: Switching und Routing
2	Sicherungsschicht	(data link)	MAC Address, Ethernet
1	Bitübertragungsschicht	(physical layer)	Cable, Network cards, Hubs

Encapsulation is the process of adding headers and trailers around data.



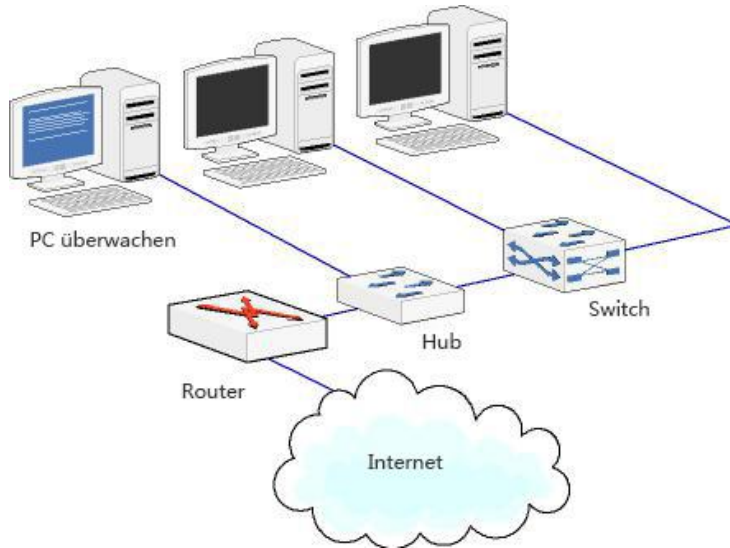
1. Bitübertragungsschicht (physical layer) -

Umwandlung der Bits in ein zum Medium passendes Signal und physikalische Übertragung

- Wir rufen <https://jvns.ca/cat.png> auf, dann muss es einen physikalischen Weg dorthin geben.
- Der Inhalt ist binär kodiert.
- Am *physical layer* wird geregelt wie die *rohen Bits* über ein Kabel als Signal von einer Netzwerkkarte zur anderen gesendet werden kann.
- *Bit Synchronization, Cabling Standards*

1. Bitübertragungsschicht (physical layer)
2. **Sicherungsschicht (data link) -**

Segmentierung der Pakete in Frames und Hinzufügen von Prüfsummen.

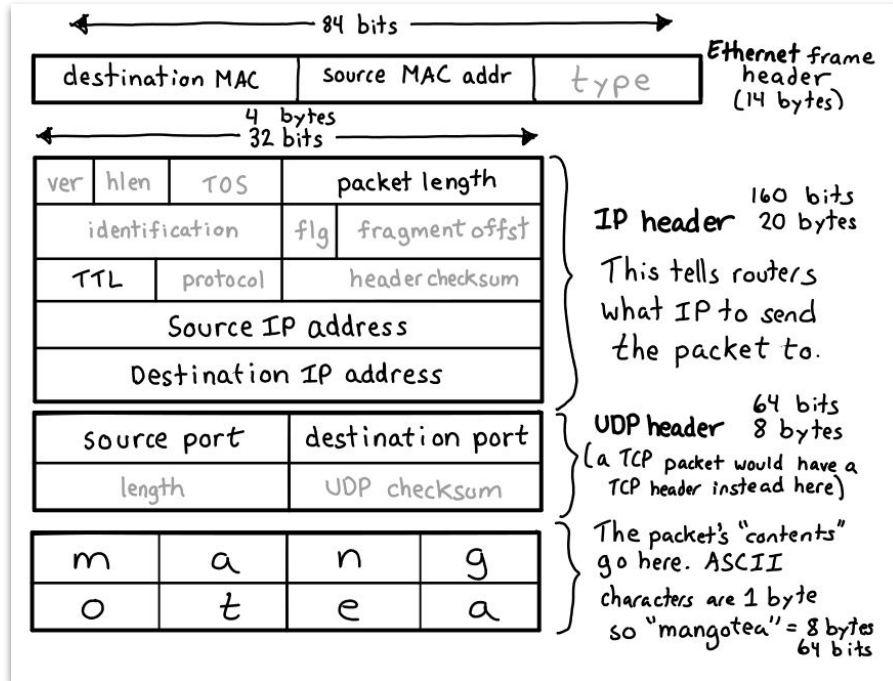


- Liefert alles um den Transfer von Daten zwischen Netzwerken zu ermöglichen und stellt Werkzeuge zur Verfügung damit Fehler auf dem *physical layer* korrigiert werden können.
- *Media Access Control* (MAC) - Adressen geben der Hardware in einem Netzwerk die Adressen. Es wird geregelt wer im Netzwerk kommunizieren darf.
- **Ethernet** ist eine Familie an Netzwerk-Technologien, die in LAN, MAN (Metropolitan) und WAN genutzt werden.

https://en.wikipedia.org/wiki/Data_link_layer

<https://community.fs.com/de/blog/whats-the-difference-hub-vs-switch-vs-router.html>

1. Bitübertragungsschicht (physical layer)
2. Sicherungsschicht (data link)
3. **Vermittlungsschicht (network layer) - Routing der Datenpakete zum nächsten Knoten.**



- Hier wird geregelt wie die Pakete (die Daten) über das Internet übermittelt werden.
- Der Standard für diese Kommunikation ist das **Internet Protocol (IP)**
- **IP** bringt ein Paket zum richtigen Rechner im Internet
- IP-Adresse
- Die Abbildung zeigt ein *UDP packet* mit dem Inhalt "mangotea"
- UDP ... User Datagram Protocol
- TTL ... Time to Live

<https://jvns.ca/networking-zine.pdf>

<https://searchnetworking.techtarget.com/definition/Network-layer>

1. Bitübertragungsschicht (physical layer)
2. Sicherungsschicht (data link)
3. Vermittlungsschicht (network layer)
4. **Transportschicht (transport layer)**



Zuordnung der Datenpakete zu einer Anwendung.

1. **Domain Name System (DNS)**
Namensauflösung von IP/Domainnamen
2. **Socket**
Vom Betriebssystem bereitgestelltes Objekt, das als Kommunikationsendpunkt dient.
3. **Transmission Control Protocol TCP (TCP/IP)**
Transport Protokoll ("Regelwerk") im Internet
4. **Hypertext Transfer Protocol (HTTP) - Request**
Protokoll um das Bild anzufordern
5. **HTTP - Response**
6. **Clean up**

1. Bitübertragungsschicht (physical layer)
2. Sicherungsschicht (data link)
3. Vermittlungsschicht (network layer)
4. Transportschicht (transport layer)
5. **Sitzungsschicht (session layer)**

Steuerung der Verbindungen und des Datenaustauschs.

- Authentication
- Authorization
- *“Damit beim Streamen alle Daten richtig zusammen kommen”*
- Regelwerk für *Sessions*
 - Request
 - Response
- z.B. Remote procedure calls (RPCs).
Damit Funktionen auf anderen Rechnern aufgerufen werden können

1. Bitübertragungsschicht (physical layer)
2. Sicherungsschicht (data link)
3. Vermittlungsschicht (network layer)
4. Transportschicht (transport layer)
5. Sitzungsschicht (session layer)
6. **Darstellungsschicht (presentation layer)**

Umwandlung der systemabhängigen Daten in ein unabhängiges Format.

- Verschlüsselung/Entschlüsselung
 - SSL
- Kodierung/Dekodierung
 - ASCII etc.

1. Bitübertragungsschicht (physical layer)
2. Sicherungsschicht (data link)
3. Vermittlungsschicht (network layer)
4. Transportschicht (transport layer)
5. Sitzungsschicht (session layer)
6. Darstellungsschicht (presentation layer)
7. **Anwendungsschicht (application layer)**



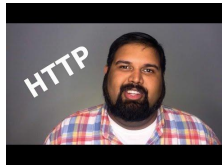
Client ... Maschine oder ein Programm, das Anfragen/Request stellen kann.

Server ... “Serve other programs called clients”

- ein Server - mehreren Clients
- Web Server - Apache
- Datenbank Server
- Ort für Web Ressourcen, hosted Web Applikationen, speichert Userdaten und Programme

Funktionen für Anwendungen sowie die Dateneingabe und -ausgabe.

- **Hyper Text Transfer Protocol (HTTP)**
ermöglicht die Request/Response kommunikation zwischen Server und Client
- **Protokoll ist zustandslos**
Server ‘weiß’ nicht ob Client schon einmal gefragt hat.
Es entstehen einfache Kommunikationszyklen.
- “Hier ist das Web”



Webtechnologien

Web-Technologien

URL	Uniform Resource Locator	- “Ort” eines Dokuments?
HTML	Hypertext Markup Language	- Struktur der Dokumente
CSS	Cascading Style Sheets	- Darstellung, Layout der Dokumente
HTTP	Hypertext Transfer Protocol	- Übertragung der Dokumente
Web-Programming		- Frontend: was der User sieht - Backend: im Hintergrund (Datenbanken)

Frontend and Backend



Ist HTML eine Programmiersprache?!

