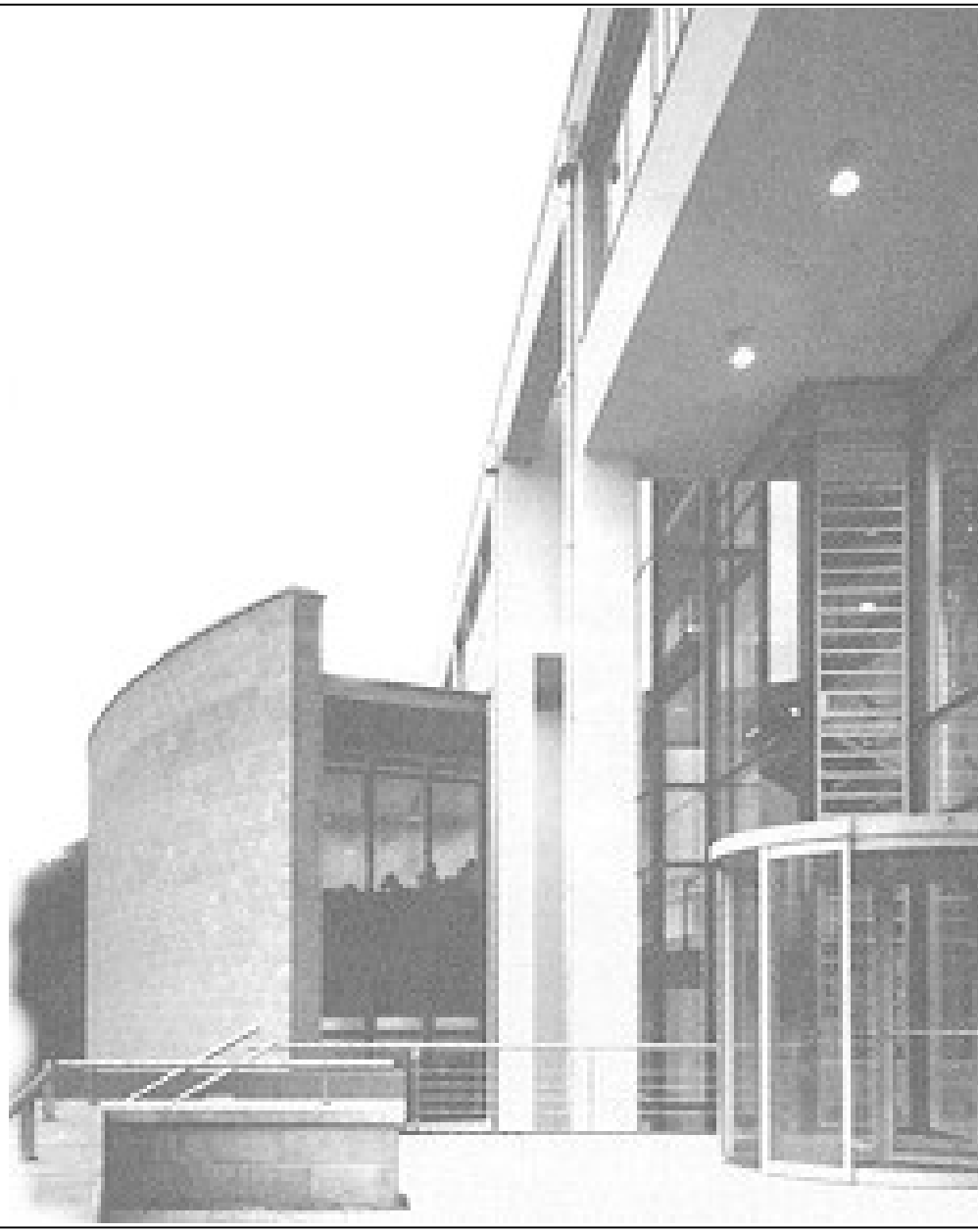
# IT-Infrastruktur

WS 2010/11

**Hans-Georg Eßer**
**Dipl.-Math., Dipl.-Inform.**

Foliensatz C (11.09.2010)



# Datenformate und Wandlung
## (von DTaus bis XML)

## (Teil 3)

# Heutiges Programm

- Kommunikationsformate
- Applikationsformate

# Kommunikation

- Verschiedene Modelle
  - Nachricht: Sender → Empfänger
  - Dialog
  - Gruppenkommunikation
- Klartext vs. „binäre“ Nachrichten

# Begriff „Kommunikationsformat“

- Format/Standard für Kommunikation
  - Netzwerkprotokolle
- Dateiformat für den Austausch von Informationen
  - PDF für Dokumente
  - spezialisierte Formate, etwa DTaus im Bankenbereich

# Netzwerkprotokolle (1/3)

## Eine kleine Auswahl

- HTTP (Webserver)
- FTP (FTP-Fileserver)
- SMTP (Mail-Versand)
- POP3, IMAP (Mail-Empfang)
- IRC (Chatten)
- SSL (Verschlüsselung)
- Time Protocol (Zeitabfrage)

# Netzwerkprotokolle (2/3)

## Grundlagen / Allgemeines

- meist Kommunikation zwischen zwei Seiten
- typisch: Client-Server-Prinzip
- verbindungslos
  - A schickt Anfrage an B
  - B schickt Antwort an A
  - Ende
- verbindungsorientiert
  - A und B bauen Verbindung auf
  - A und B kommunizieren über die Verbindung, bis sie geschlossen wird

# Netzwerkprotokolle (3/3)

- Kommunikation folgt einem festen, vorgegebenen Protokoll
- Client und Server müssen sich exakt an das Protokoll halten, damit Kommunikation gelingt
- wie im „richtigen Leben“
  - A: „Wie möchten Sie bezahlen?“ –
    B: „mit Kreditkarte“ – *erfolgreich*
  - A: „Wie möchten Sie bezahlen?“ –
    B: „Heute ist es sehr warm“ – *nicht erfolgreich*

# HTTP (1/8)

- HTTP: **H**yper**t**ext **T**ransfer **P**rotocol
- Client: Webbrowser (Firefox, IE, Opera etc.)
- Server: Webserver (Apache, Microsoft IIS etc.)
- verbindungsorientierte Kommunikation (TCP)
- Anfrage: HTTP-Request
- Antwort: HTTP-Response

# HTTP (2/8)

## HTTP-Requests

- GET (Daten: Header + Dokument holen)
- POST (Daten an den Server senden, Formulare)
- HEAD (nur Header holen)

## HTTP GET

- Anfrage:

```
GET /ii-ws2010/fom-itis-datenformate-01.pdf HTTP/1.1
Host: fom.hgesser.de
```

- Antwort:

```
HTTP/1.1 200 OK
Date: Fri, 10 Sep 2010 10:33:23 GMT
Server: Apache/2.2.3 (Debian) PHP/4.4.4-8+etch6
Last-Modified: Fri, 03 Sep 2010 13:40:06 GMT
ETag: "3c630f7e-aec13-118b9980"
Accept-Ranges: bytes
Content-Length: 715795
Content-Type: application/pdf

%PDF-1.4
[...]
```

# HTTP (4/8)

- Angabe des Hosts

  `Host: fom.hgesser.de`

  war früher (HTTP 1.0) nicht nötig und nicht möglich
- Moderne Webserver hosten mehrere Domains unter derselben IP-Adresse (z. B. fom.hgesser.de, hm.hgesser.de, hgesser.com) – ohne Angabe des Hostnamens keine Unterscheidung möglich

# HTTP (5/8)

## **HTTP HEAD:** holt nur den Header

- Anfrage:

```
HEAD /ii-ws2010/fom-itis-datenformate-01.pdf HTTP/1.1
Host: fom.hgesser.de
```

- Antwort:

```
HTTP/1.1 200 OK
Date: Fri, 10 Sep 2010 10:37:00 GMT
Server: Apache/2.2.3 (Debian) PHP/4.4.4-8+etch6
Last-Modified: Fri, 03 Sep 2010 13:40:06 GMT
ETag: "3c630f7e-aec13-118b9980"
Accept-Ranges: bytes
Content-Length: 715795
Content-Type: application/pdf
```

# HTTP (6/8)

## **HTTP POST:** überträgt Zusatzdaten

- Anfrage:

```
POST /search.html HTTP/1.1
Host: fom.hgesser.de
Content-Type: application/x-www-form-urlencoded
Content-Length: 20

search=Infrastruktur
```

- Antwort:

```
HTTP/1.1 200 OK
Date: Fri, 10 Sep 2010 10:37:00 GMT
Server: Apache/2.2.3 (Debian) PHP/4.4.4-8+etch6
Last-Modified: Fri, 03 Sep 2010 13:40:06 GMT
[...]
```

# HTTP (7/8)

- Verbindung bleibt nach GET/HEAD/POST geöffnet
  - Client kann sofort weitere Anfragen schicken (z. B. für Bilder auf der HTML-Seite)
  - reduziert den Aufwand für Verbindungsauf- und -abbau (TCP-Verbindungen)
- Bei Fehler (z. B. falsche Syntax in Anfrage) wird Verbindung geschlossen. Rückmeldung:

```
HTTP/1.1 400 Bad Request
```

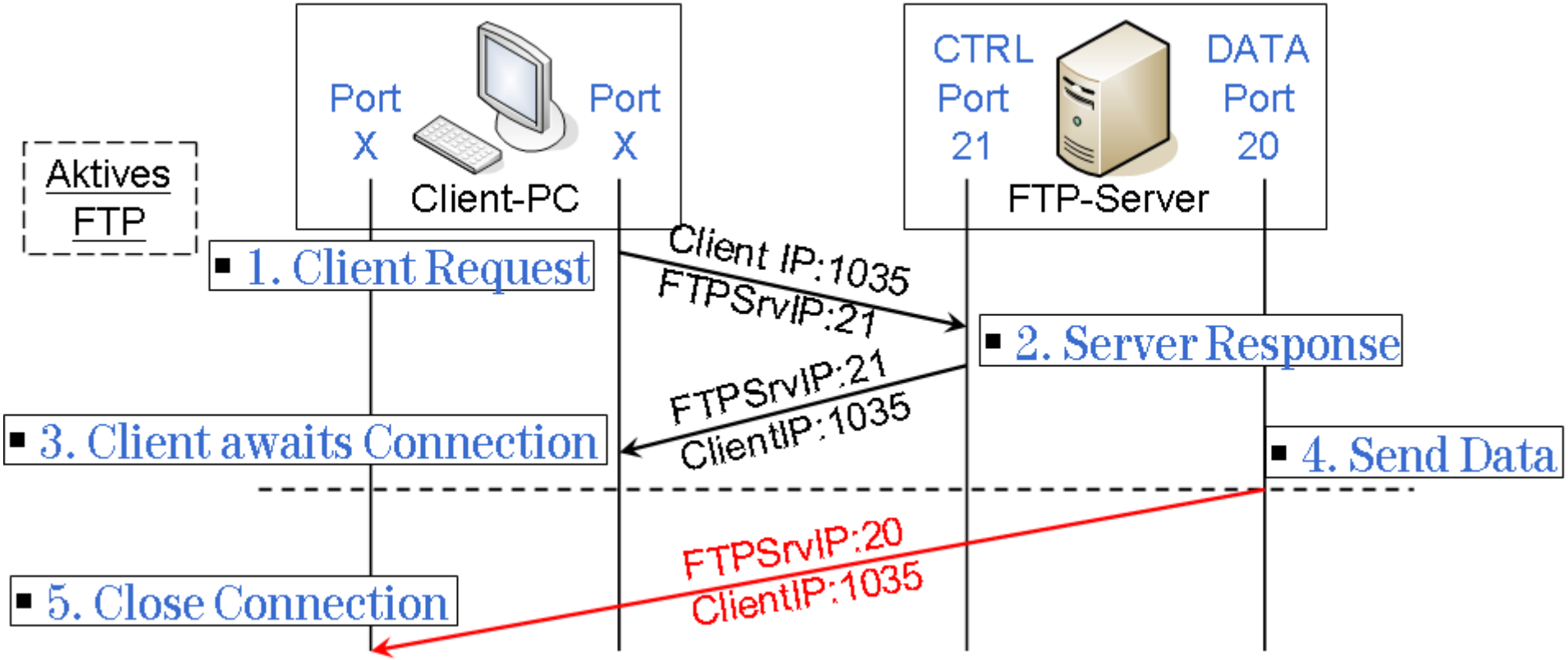
# HTTP (8/8)

- HTTP Status Codes
  - immer in der ersten Zeile der Antwort
  - 200 OK
  - 301 Moved Permanently
  - 304 Not Modified
  - 400 Bad Request
  - 401 Unauthorized
  - 403 Forbidden
  - 404 Not Found
  - 500 Internal Server Error
  - 503 Service Unavailable
  - weitere: http://de.wikipedia.org/wiki/HTTP-Statuscode
- Antwort steht nur im Header!

# FTP (1/3)

- **FTP: F**ile **T**ransfer **P**rotocol
- Client: FTP-Client (z. B. Webbrowser oder spezielles FTP-Programm)
- Server: FTP-Server
- verbindungsorientierte Kommunikation (TCP) mit zwei Kanälen
  - Kontrollkanal (Port 21)
  - und Datenkanal (Port 20)
- aktives FTP (Direktverbindung) vs. passives FTP (Firewall)

# FTP (2/3)




Quelle Bild: http://de.wikipedia.org/wiki/File_Transfer_Protocol

# FTP (3/3)

- Authentifizierung (Username, Passwort)
- alternativ: „anonymous FTP“
- verschiedene Übertragungsvarianten
  - ASCII
  - Binary

# Mail-Versand mit SMTP (1/2)

- **SMTP: S**imple **M**ail **T**ransfer **P**rotocol
- Kommunikation im Textmodus, SMTP-Port 25
- Kommandos:
  - HELO („Begrüßung“)
  - MAIL FROM: *Adresse*
  - RCPT TO: *Adresse*
  - DATA
  - dann die Nachricht (erst die Mail-Header)

# Mail-Versand mit SMTP (2/2)

```
esser@netbook:~$ telnet hgesser.de smtp
Trying 217.160.135.96...
Connected to hgesser.de.
Escape character is '^]'.
220 hgesser.de ESMTP Exim 3.36 #1 Sat, 11 Sep 2010 00:28:31 +0200
helo absender.de
250 hgesser.de Hello mnch-123.pool.mediaways.net [93.133.124.239]
mail from: h.g.esser@gmx.de
250 <h.g.esser@gmx.de> is syntactically correct
rcpt to: esser@hgesser.de
250 <esser@hgesser.de> verified
data
354 Enter message, ending with "." on a line by itself
Hallo, das ist eine Nachricht.
.
250 OK id=1OuC5Z-0005cW-00
quit
221 s15337257.onlinehome-server.info closing connection
Connection closed by foreign host.
```

# Daytime Protocol

- veraltetes (aber angenehm einfaches) Protokoll, mit dem man nach der Uhrzeit fragen kann
- Kontakt zu Port 13 (daytime) eines Servers, der einen Daytime-Service anbietet
- Server liefert Zeit als lesbaren Text

```
esser@netbook:~$ netcat hgesser.de daytime
Sat Sep 11 00:23:55 2010
```

- nicht mit NTP verwechseln ...

# Austauschformate

## Und wieder: Eine kleine Auswahl

- RSS (News-Feeds)
- DTaus (Banküberweisungen)
- PDF/PostScript
- CSV (Tabellendaten)
- EPS (Vektorgrafik-Austauschformat)
- RTF (Textdatei-Austauschformat)

# RSS (1/3)

- **RSS: R**eally **S**imple **S**yndication
- RSS-Feed: XML-Format, enthält Titel, Kurzfassungen und Links zu Nachrichten
- typisch bei Blogs, News-Seiten u. ä.
- Idee:
  - News-Feed abonnieren
  - regelmäßig Infos zu neuen Beiträgen erhalten
  - Bei Klick auf Link den vollen Artikel lesen

# RSS (2/3)

```
<?xml version="1.0" encoding="utf-8"?>
<rss version="2.0">
  <channel>
    <title>IT-Infrastruktur News</title>
    <link>http://fom.hgesser.de/</link>
    <description>Infos zu IT-Infrastruktur / FOM</description>
    <language>de-de</language>
    <copyright>Hans-Georg Eßer</copyright>
    <pubDate>Sat, 11 Sep 2010 2:43:19</pubDate>

    <item>
      <title>Folien zu RSS fertig</title>
      <description>Kurze Zusammenfassung: RSS ...</description>
      <link>http://fom.hgesser.de/rss-folien/</link>
      <author>Hans-Georg Eßer</author>
      <guid>http://fom.hgesser.de/id00012</guid>
      <pubDate>Sat, 11 Sep 2010 2:43:19</pubDate>
    </item>

    <item> ...  </item>
  </channel>
</rss>
```

# RSS (3/3)

IT-Infrastruktur News - Mozilla Developer Preview 3.7 Alpha 5
File Edit View History Bookmarks Tools Help
http://hgesser.de/test.rss
Google
Meistbesuchte Se...
Getting Started
Latest Headlines
IT-Infrastruktur News
IT-Infrastruktur News
Infos zu IT-Infrastruktur / FOM
Folien zu RSS fertig
Kurze Zusammenfassung: RSS ...
Mehr zu RSS
Noch mehr zu RSS aufgeschrieben! Die Beschreibung darf ruhig ein paar Zeilen lang sein. Manche schreiben den kompletten Artikel mit rein.
Done
3.0.1 [29.07.2010]

# Banking: DTaus (1/7)

- Datenträgeraustauschverfahren
  - ursprünglich: mit Disketten („Datenträger-...“)
  - auch: online
  - Kunde ↔ Bank und Bank ↔ Bank
- Datei besteht aus
  - A-Satz: „Datenträgervorsatz“ (Header), 128 Byte
  - C-Sätze: Buchungen („Zahlungsaustausch“), $187 + 29 \cdot n$ Byte
  - E-Satz: „Datenträgernachsatz“ (Prüfsumme), 128 Byte

# Banking: DTaus (2/7)

## A-Satz

| | | | |
|---|---|---|---|
| 1 | 0 | 4 Zeichen | Länge des Datensatzes, immer "0128" |
| 2 | 4 | 1 Zeichen | Datensatz-Typ, immer 'A' |
| 3 | 5 | 2 Zeichen | Art der Transaktionen<br>"LB" für Lastschriften Bankseitig<br>"LK" für Lastschriften Kundenseitig<br>"GB" für Gutschriften Bankseitig<br>"GK" für Gutschriften Kundenseitig |
| 4 | 7 | 8 Zeichen | Bankleitzahl des Auftraggebers |
| 5 | 15 | 8 Zeichen | CST, "00000000", nur belegt, wenn Diskettenabsender Kreditinstitut |
| 6 | 23 | 27 Zeichen | Name des Auftraggebers |
| 7 | 50 | 6 Zeichen | aktuelles Datum im Format DDMMJJ |
| 8 | 56 | 4 Zeichen | CST, "  " (Blanks) |
| 9 | 60 | 10 Zeichen | Kontonummer des Auftraggebers |
| 10 | 70 | 10 Zeichen | Optionale Referenznummer |
| 11a | 80 | 15 Zeichen | Reserviert, 15 Blanks |
| 11b | 95 | 8 Zeichen | Optionales Ausführungsdatum, DDMMJJJJ. |
| 11c | 103 | 24 Zeichen | Reserviert, 24 Blanks |
| 12 | 127 | 1 Zeichen | Währungskennzeichen: " " = DM, "1" = Euro |
| | -- | 128 Zeichen | |

Quelle: Dokumentation zum Programm *dtaus* (http://www.infodrom.org/projects/dtaus/)

# Banking: DTaus (3/7)

## C-Satz (1/3)

```
1    0   4 Zeichen Länge des Datensatzes, 187 + x * 29 (x..Anzahl
                   Erweiterungsteile)
2    4   1 Zeichen Datensatz-Typ, immer 'C'
3    5   8 Zeichen Bankleitzahl des Auftraggebers (optional)
4   13   8 Zeichen Bankleitzahl des Kunden
5   21  10 Zeichen Kontonummer des Kunden
6   31  13 Zeichen 1. Zeichen        "0"
                   2. - 12. Zeichen  interne Kundennummer oder Nullen
                   13. Zeichen       "0"
                   Die interne Nummer wird vom erstbeauftragten
                   Institut zum endbegünstigten Institut
                   weitergeleitet.  Die Weitergabe der internenen
                   Nummer an den Überweisungsempfänger ist der
                   Zahlstelle freigestellt.
```

Quelle dieser und der folgenden zwei Seiten:
Dokumentation zum Programm *dtaus* (http://www.infodrom.org/projects/dtaus/)

# Banking: DTaus (4/7)

## C-Satz (2/3)

```
 7a  44   2 Zeichen Art der Transaktion
 7b  46   3 Zeichen ------- "----------
                    "04000" Lastschrift des Abbuchungsauftragsverfahren
                    "05000" Lastschrift des Einzugsermächtigungsverfahren
                    "05005" Lastschrift aus Verfügung im elec. Cash-System
                    "05006" Wie 05005 mit ausländischen Karten
                    "05015" Lastschrift aus Verfügung im elec. Cash-
                            System - POZ
                    "51000" Überweisungs-Gutschrift
                    "53000" Überweisung Lohn/Gehalt/Rente
                    "54XXJ" VL mit Sparzulage
                            XX = 00 oder %-Satz der Sparzulage
                            J = Endziffer des Jahres für diese Leistung
                    "56000" Überweisung öffentlicher Kassen
 8   49   1 Zeichen Reserviert, " " (Blank)
 9   50  11 Zeichen Betrag
10   61   8 Zeichen Bankleitzahl des Auftraggebers
11   69  10 Zeichen Kontonummer des Auftraggebers
```

# Banking: DTaus (5/7)

## C-Satz (3/3)

```
12   79  11 Zeichen Betrag in Euro einschließlich
                    Nachkommastellen, nur belegt, wenn Euro als
                    Währung angegeben wurde (A12, C17a), sonst Nullen
13   90   3 Zeichen Reserviert, 3 Blanks
14a  93  27 Zeichen Name des Kunden
14b 120   8 Zeichen Reserviert, 8 Blanks
     -- 128 Zeichen
15  128  27 Zeichen Name des Auftraggebers
16  155  27 Zeichen Verwendungszweck
17a 182   1 Zeichen Währungskennzeichen: " " = DM, "1" = Euro
17b 183   2 Zeichen Reserviert, 2 Blanks
18  185   2 Zeichen Anzahl der Erweiterungsdatensätze, "00" bis "15"
19  187   2 Zeichen Typ (1. Erweiterungsdatensatz)
                      "01" Name des Kunden
                      "02" Verwendungszweck
                      "03" Name des Auftraggebers
20  189  27 Zeichen Beschreibung gemäß Typ
21  216   2 Zeichen wie C19, oder Blanks,  (2. Erweiterungsdatensatz)
22  218  27 Zeichen wie C20, oder Blanks
23  245  11 Zeichen 11 Blanks
     -- Ende des ersten Erweiterungsdatensatzes
     -- 256 Zeichen
```

# Banking: DTaus (6/7)

- Daten erzeugen mit Tool *dtaus* (http://www.infodrom.org/projects/dtaus/)
- Umwandeln mit

```
dtaus -dtaus -c control.txt
```

- Ergebnis in Datei *dtaus0.txt*

control.txt:

```
BEGIN {
  Art   GK
  Name  Martin Schulze
  Konto 123545
  BLZ   2004002
  Ausfuehrung 23.12.2001
  Euro
}

{
  Transaktion Gutschrift
  Name   Martha Schulze
  Konto  98832
  BLZ    2004003
  Betrag 20.00
  Zweck  Gebuehr Wohnheimnetz
  Text   Anschluss u. 11+12.97
}
```

# Banking: DTaus (7/7)

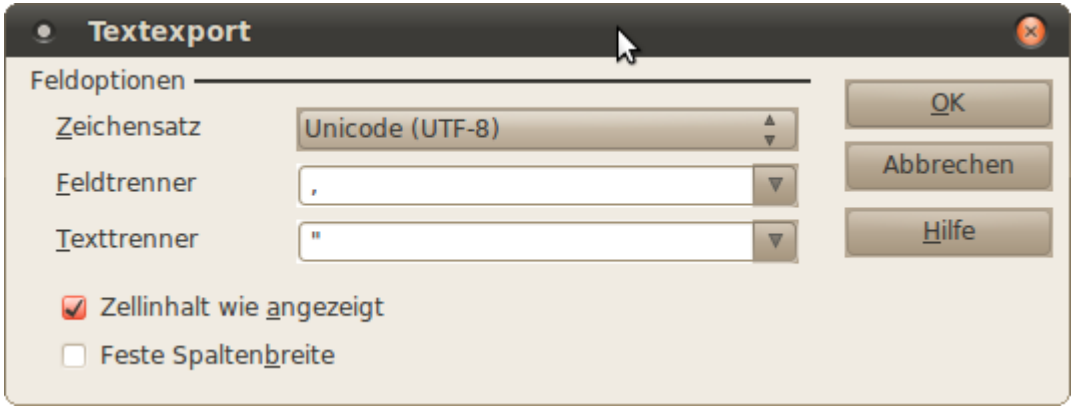
```
00000000  30 31 32 38 41 47 4b 30  32 30 30 34 30 30 32 30  |0128AGK020040020|
00000010  30 30 30 30 30 30 30 4d  41 52 54 49 4e 20 53 43  |0000000MARTIN SC|
00000020  48 55 4c 5a 45 20 20 20  20 20 20 20 20 20 20 20  |HULZE           |
00000030  20 20 31 30 30 39 31 30  20 20 20 20 30 30 30 30  |  100910    0000|
00000040  31 32 33 35 34 35 30 30  30 30 30 30 30 30 30 30  |1235450000000000|
00000050  20 20 20 20 20 20 20 20  20 20 20 20 20 20 20 32  |               2|
00000060  33 31 32 32 30 30 31 20  20 20 20 20 20 20 20 20  |3122001         |
00000070  20 20 20 20 20 20 20 20  20 20 20 20 20 20 20 31  |               1|
00000080  30 32 31 36 43 30 32 30  30 34 30 30 32 30 32 30  |0216C02004002020|
00000090  30 34 30 30 33 30 30 30  30 30 39 38 38 33 32 30  |0400300000988320|
000000a0  30 30 30 30 30 30 30 30  30 30 30 30 35 31 30 30  |0000000000005100|
000000b0  30 20 30 30 30 30 30 30  30 30 30 30 30 30 32 30  |0 00000000000020|
000000c0  30 34 30 30 32 30 30 30  30 31 32 33 35 34 35 30  |0400200001235450|
000000d0  30 30 30 30 30 30 32 30  30 30 20 20 20 4d 41 52  |0000002000   MAR|
000000e0  54 48 41 20 53 43 48 55  4c 5a 45 20 20 20 20 20  |THA SCHULZE     |
000000f0  20 20 20 20 20 20 20 20  20 20 20 20 20 20 20 20  |                |
00000100  4d 41 52 54 49 4e 20 53  43 48 55 4c 5a 45 20 20  |MARTIN SCHULZE  |
00000110  20 20 20 20 20 20 20 20  20 20 20 47 45 42 55 45  |           GEBUE|
00000120  48 52 20 57 4f 48 4e 48  45 49 4d 4e 45 54 5a 20  |HR WOHNHEIMNETZ |
00000130  20 20 20 20 20 20 31 20  20 30 31 30 32 41 4e 53  |      1  0102ANS|
00000140  43 48 4c 55 53 53 20 55  2e 20 31 31 2b 31 32 2e  |CHLUSS U. 11+12.|
00000150  39 37 20 20 20 20 20 20  20 20 20 20 20 20 20 20  |97              |
00000160  20 20 20 20 20 20 20 20  20 20 20 20 20 20 20 20  |                |
```

# CSV (1/2)

- **CSV: C**omma-**S**eparated **V**alues
- einfache Textdatei, die Tabellen in folgendem Aufbau speichert:
  - jede Zeile der Datei entspricht einer Tabellenzeile
  - Felder (Spalten) sind durch ein Trennsymbol (oft: ein Komma, alternativ: Semikolon, Tabulator, Leerzeichen) getrennt
  - Felder zusätzlich in Anführungszeichen (um Kommata im Feld zu erlauben)

# CSV (2/2)

| | A | B |
|---|---|---|
| 1 | Eins | Vier |
| 2 | Zwei | Fünf |
| 3 | Drei | Sechs |
| 4 | | |




Ergebnis des Exports ins CSV-Format:

```
"Eins","Vier"
"Zwei","Fünf"
"Drei","Sechs"
```

# EPS

- **EPS: E**ncapsulated **P**ost**S**cript
- im Prinzip eine normale (einseitige) PostScript-Datei, aber mit einer zusätzlichen „Bounding Box“
- EPS-Dateien kann man in andere Dokumente integrieren (wie Bilddateien)

```
%!PS-Adobe-3.0 EPSF-3.0
%%Title: rob-aktiv.eps
%%Creator: (ImageMagick)
%%CreationDate: Thu May 13 09:43:18 2004
%%BoundingBox: 0 0 168 86
%LanguageLevel: 1
%%Pages: 1
```

unten links: (0, 0)
oben rechts: (168, 86)
(Einheit: 1/72 Zoll)

# Applikationsformate

- **CAD**, **C**omputer **A**ided **D**esign
- **GIS**, **G**eo**i**nformations**s**ystem
- **DTP**, **D**esk**t**op **P**ublishing

# CAD (1/4)

**Computer Aided Design**

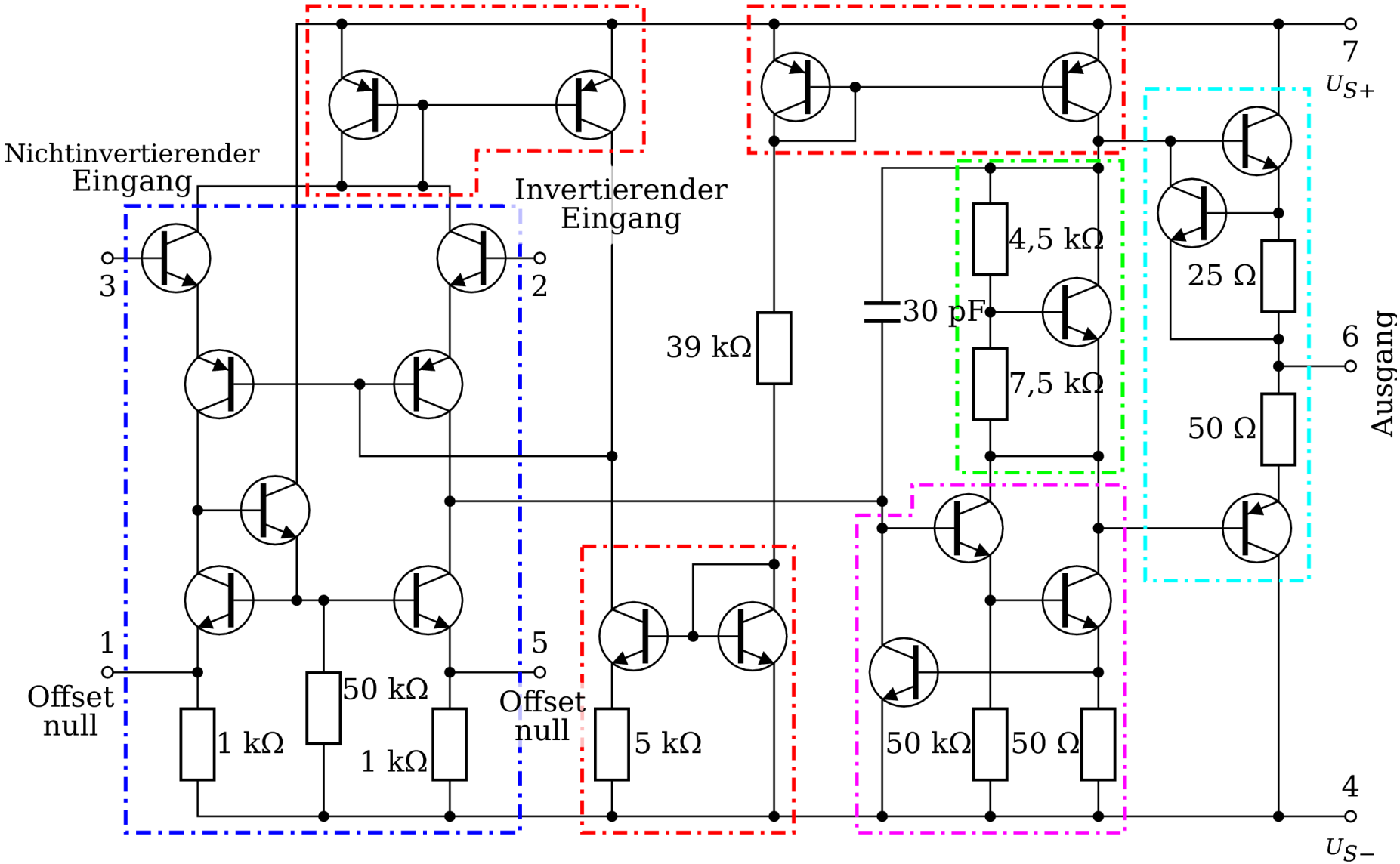
- Mechanische Konstruktion
  - Bauwesen
    - Architektur (CAAD)
    - Holzbau
    - Ingenieurbau
    - Historische Rekonstruktion
    - Städtebau
    - Wasserbau
    - Verkehrswegebau
  - Vermessungswesen
  - Produktdesign
- Maschinenbau
  - Anlagenbau
  - Fahrzeugbau
  - Formen- und Werkzeugbau
  - Antriebstechnik
  - Mechanische Simulation
- Schaltpläne in der Elektrotechnik
- Zahnmedizin
- Schmuck- und Textilindustrie
- Elektronische Schaltungen

Quelle: http://de.wikipedia.org/wiki/CAD

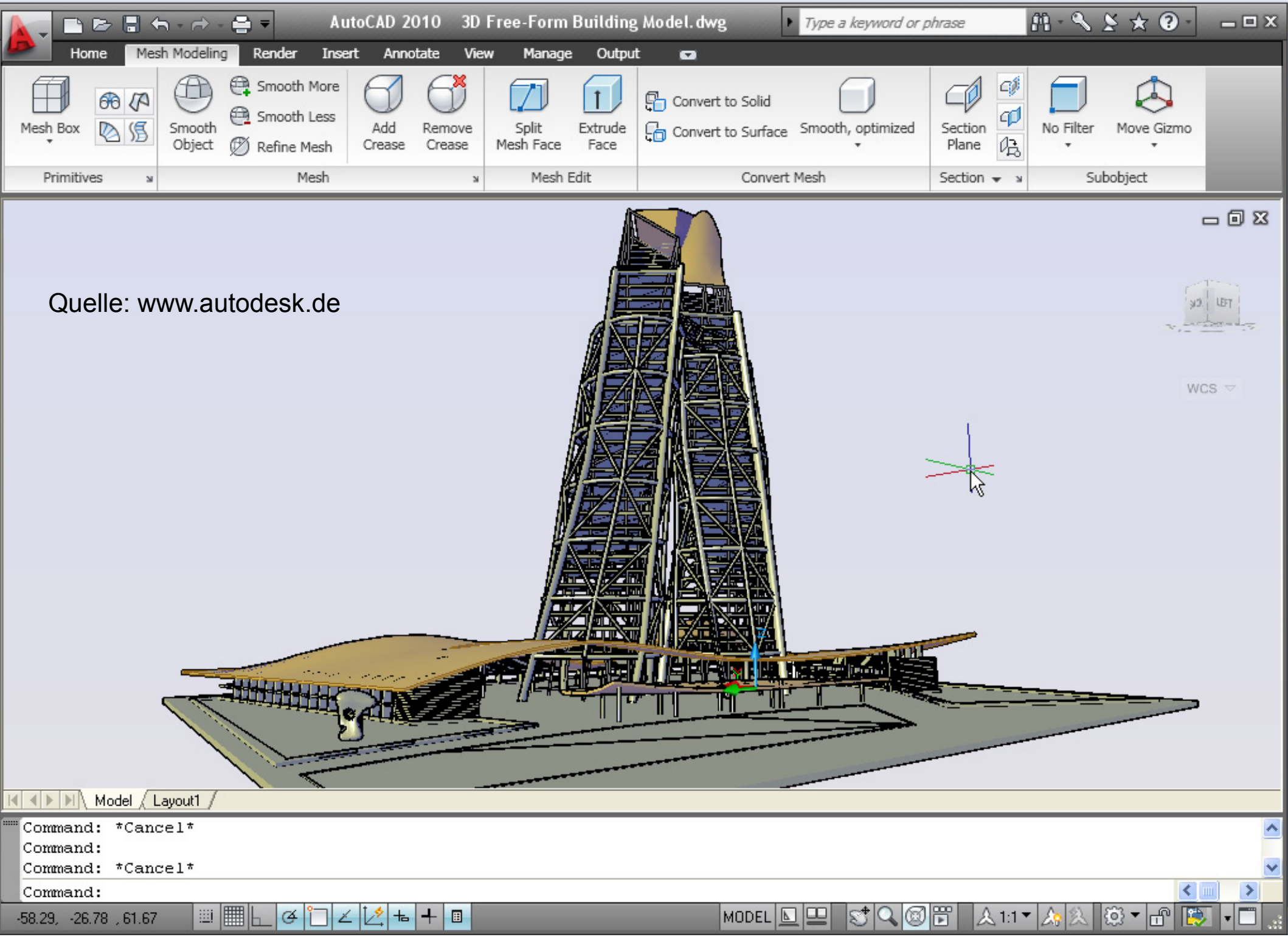
# CAD (2/4)

- Dateiformate: vektorbasiert
- Verwaltung in objektorientierter Datenbank (nicht in einfachen Vektorgrafikdateien)
- Jedes Objekt besteht aus kleineren Teilobjekten
- Objekte haben Eigenschaften
  - Maschinenteile: Anschlüsse, Befestigungen etc.

# CAD (3/4)




Quelle: http://de.wikipedia.org/wiki/Schaltplan


AutoCAD 2010 3D Free-Form Building Model.dwg
Type a keyword or phrase
Home
Mesh Modeling
Render
Insert
Annotate
View
Manage
Output
Mesh Box
Smooth Object
Smooth More
Smooth Less
Refine Mesh
Add Crease
Remove Crease
Split Mesh Face
Extrude Face
Convert to Solid
Convert to Surface
Smooth, optimized
Section Plane
No Filter
Move Gizmo
Primitives
Mesh
Mesh Edit
Convert Mesh
Section
Subobject
Quelle: www.autodesk.de
WCS
Model
Layout1
Command: *Cancel*
Command:
Command: *Cancel*
Command:
-58.29, -26.78 , 61.67
MODEL
1:1

# GIS (1/3)

- GIS: **G**eo**i**nformations**s**ystem
- Geometriedaten + Sachdaten
- gibt es vektor- und rastergrafikbasiert
- Beispiel: **GPX**, **G**PS E**x**change **F**ormat
  - XML-basiert, speichert
    - GPS-Informationen (Breiten- und Längengrad, Höhe)
    - Zusatzdaten (Zeitpunkt, Ortsbeschreibung etc.)

# GIS (2/3)

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<gpx xmlns="http://www.topografix.com/GPX/1/1"
    creator="byHand" version="1.1"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.topografix.com/GPX/1/1
    http://www.topografix.com/GPX/1/1/gpx.xsd">

  <wpt lat="39.921055008" lon="3.054223107">
    <ele>12.863281</ele>
    <time>2005-05-16T11:49:06Z</time>
    <name>Cala Sant Vicenç - Mallorca</name>
    <sym>City</sym>
  </wpt>
</gpx>
```

lat = latitude = Breite,
lon = longitude = Länge,
ele = elevation = Höhe
wpt = waypoint = Wegpunkt

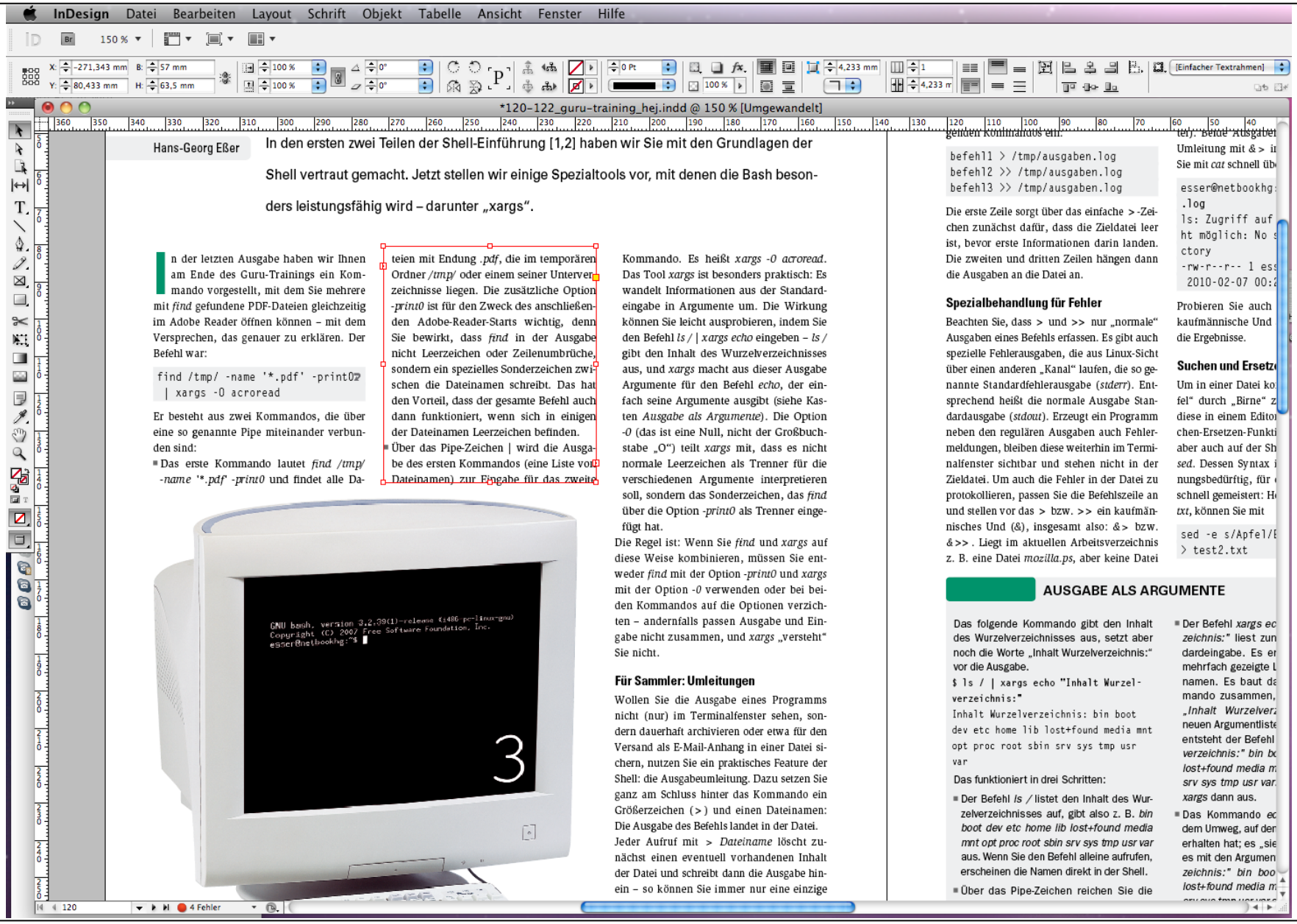
# GIS (3/3)

Quelle: http://www.gpsvisualizer.com/examples/

# Zeitschriftenlayout (1/2)

- Layout-/Satzprogramme (Desktop Publishing)
- ähnlich wie Textverarbeitung, aber rahmenbasiert
- Text fließt von Rahmen zu Rahmen (auch zur nächsten Seite)
- Beispiele: Adobe InDesign, QuarkXPress, Scribus
- Dateiformate proprietär (bei kommerziellen Anwendungen)

# Zeitschriftenlayout (2/2)



# Ausblick

## Nächste Veranstaltung: Freitag, 01.10.

- Ergonomie und Arbeitsschutz (voraussichtlich)