Entwicklung webbasierter Anwendungen

7. Kapitel: Webserver

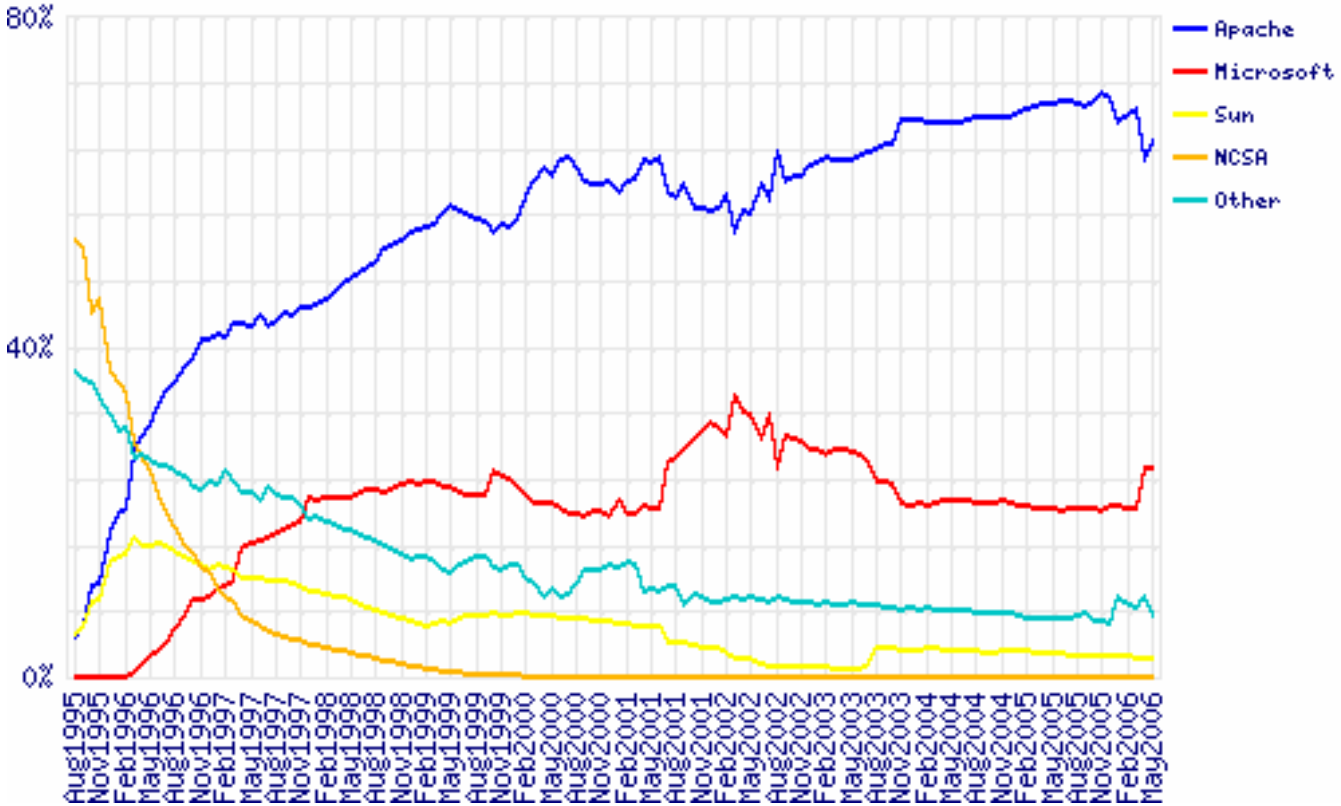
Quellenhinweis:
Viele Folien dieser Vorlesung entstammen der gleichnamigen Vorlesung von Prof. B. Kreling

Einleitung

- Was ist ein Webserver?
  - ⇨ eine (spezielle) Software
  - ⇨ übermittelt auf Anfrage Daten mit dem HTTP-Protokoll
- Was braucht ein Webserver?
  - ⇨ TCP/IP-Unterstützung

    (vom Betriebssystem; darauf setzt das TCP/IP-Protokoll auf)
  - ⇨ Internet-Zugang (sinnvoll, aber für die Funktion nicht nötig)
- Was ist zu tun?
  - ⇨ Installation
  - ⇨ Zuordnung der "öffentlichen" Daten / Verzeichnisse
  - ⇨ Anbindung an andere Software (Skripte, Office,…)
  - ⇨ Konfiguration der Berechtigungen

Verfügbare Webserver (05/2006)



Quelle:

NETCRAFT

http://news.netcraft.com/

Developer	April 2006	Percent	May 2006	Percent	Change
Apache	50588433	62.72	52819517	64.76	2.04
Microsoft	20343656	25.22	20764239	25.46	0.24
Sun	1907503	2.36	1917950	2.35	-0.01
Zeus	563381	0.70	550437	0.67	-0.03

Verfügbare Webserver

- Apache http://httpd.apache.org
  - ⇨ defacto-Standard im Web. OpenSource-Produkt und Freeware.
    für *UNIX-Plattformen* und für *MS Windows/DOS* verfügbar.
- Microsofts Internet Information Server (IIS)
  - ⇨ Kommerzieller Webserver für Windows Server
- SunONE http://wwws.sun.com
  - ⇨ Kommerzieller Webserver, ehemals Netscape Enterprise Server
- NCSA (National Center for Supercomputing Applications)
  - ⇨ Universität von Illinois, Champaign-Urbana. Neben dem CERN eine der ursprünglichen Entwicklungsstätten des WWW.
    Hier wurde die erste Version des Web-Browsers Mosaic entwickelt.
- Zeus http://www.zeus.com/
  - ⇨ Kommerzieller Webserver; für Sun, Mac, Linux, HP UX, IBM AIX …
    – aber nicht für Windows

Allgemeines

- Konfiguration erfolgt je nach Webserver entweder Dialog-gesteuert, oder über Konfig-Datei
- Webserver läuft entweder als Anwendung oder als Prozess im Hintergrund (Dienst)
- Verwendung des Webservers ist auch lokal (ohne Internetzugang) möglich (z.B. zu Testzwecken)
- Manche Webserver unterstützen „Virtual Hosts“, d.h. mehrere Web-Zugänge werden auf einem Server realisiert
- Die Konfiguration der Webserver unterscheidet sich. Details in der jeweiligen Dokumentation.

Grundeinstellungen allgemein (1)

- **IP-Adresse und Hostnamen des Servers**
  - ⇨ Für lokalen Betrieb: 127.0.0.1 oder localhost.

    Test: Im Web-Browser (nach Start des Webservers) *http://127.0.0.1/* oder *http://localhost/* aufrufen.
- **Port des Servers**
  - ⇨ normalerweise Port 80
- **HTTP-Wurzelverzeichnis für HTML-Dateien**
  - ⇨ Pfadname (je nach Syntax Ihres Betriebssystems) unterhalb dessen sich die lokalen HTML-Dateien befinden. z.B. c:\www\myhtml unter Windows
- **Default-HTML-Dateiname für Verzeichnisse**
  - ⇨ index.html oder index.htm

Grundeinstellungen allgemein (2)

- **Physisches Verzeichnis für CGI-Scripts**
  - ⇨ normalerweise *cgi-bin*

    Pfadname (je nach Syntax Ihres Betriebssystems) mit ausführbaren CGI-Scripts

    z.B. c:\www\cgi-bin
- **Virtuelles Verzeichnis für CGI-Scripts**
  - ⇨ normalerweise */cgi-bin*

    Pfadname zu den CGI-Scripts für WWW-Zugriffe

    Zugriff über *http://localhost/cgi-bin/myCGI.pl*
- **Pfad zu Perl-Interpreter und anderen Interpretern**
  - ⇨ z.B. c:\programme\perl\bin\perl.exe unter Windows

Grundeinstellungen allgemein (3)

- **Log-Dateien**
  - ⇨ Protokollierung der Zugriffe: *access.log*
  - ⇨ Fehlerprotokollierung: *error.log*
- **Timeouts**
  - ⇨ für Senden und Empfangen: 60 (= eine Minute)
  - ⇨ Die Angaben erfolgen in der Regel in Sekunden
- **MIME-Typen**
  - ⇨ Dateiformate, die der Webserver kennt und an den aufrufenden Web-Browser überträgt
  - ⇨ Andere Dateitypen sendet der Server nicht korrekt bzw. mit dem eingestellten Standard-MIME-Typ (text/plain)

Apache

- im April 1995 erstmals in einer Version 0.6.2 publiziert
- Open-Source-Entwicklung (steht jedem kostenlos zur Verfügung)
- 1999 Gründung der Apache Software Foundation
- heute weltweit am häufigsten eingesetzte Webserver
  - ⇨ oft auch noch in der alten Version: 1.3.35
  - ⇨ aktuelle Version (05.2006): Apache 2.2.2

- Einfache Installation im Paket: XAMPP
  - ⇨ Apache distribution enthält MySQL, PHP, Perl uvm.
  - ⇨ verfügbar für Linux, Windows, Mac, Solaris
  - ⇨ http://www.apachefriends.org/de/index.html

XAMPP

The Apache Software Foundation
http://www.apache.org/

Apache Grundeinstellungen

Konfig-Datei `httpd.conf` im Verzeichnis *…\xampp\appache\conf*

- Wurzelverzeichnis der Apache-Installation
  - ⇨ `ServerRoot "C:/programme/xampp/apache"`
- Port, über den der Server kommuniziert (Standard)
  - ⇨ `Listen 80`
- eMail-Adresse des Administrators für Probleme
  - ⇨ `ServerAdmin name@fbi.h-da.de`
- Wurzelverzeichnis für Dokumente (Alias kann auf anderes Verzeichnis zeigen)
  - ⇨ `DocumentRoot "C:/Programme/xampp/apache/htdocs"`
- mögliche Dateinamen der Startseite

```
<IfModule mod_dir.c>
    DirectoryIndex index.html index.php
</IfModule>
```

www.covalent.net bietet dazu Comanche, einen (teuren) Konfigurator mit GUI

Log-Dateien

- Log-Datei für Fehlermeldungen

```
ErrorLog logs/error.log
```

- Log-Datei für Zugriffe

```
CustomLog logs/access.log common
```

- Woher kamen die Verweise ?

```
CustomLog logs/referer.log referer
```

- Welche Browser wurden verwendet ?

```
CustomLog logs/agent.log agent
```

- Alternativ: alles zusammen

```
CustomLog logs/access.log common
```

- eigenes Logging-Format definieren

```
LogFormat "Formatstring" Name
```

Alias-Verzeichnisse

- Alias definiert die Abbildung URL ⇨ Verzeichnis
  - ⇨ Dokumente können in anderen Verzeichnissen abgelegt werden als mit DocumentRoot festgelegt wurde
- ScriptAlias definiert die Abbildung URL ⇨ Verzeichnis für Server-Skripte
  - ⇨ d.h. für Dateien, die nicht zum Client gesendet sondern im Server ausgeführt werden
-
```
<IfModule mod_alias.c>
    Alias /manual/ "C:/Dokumentation/"
    Alias /user/ "C:/Benutzerverzeichnisse/"
    ScriptAlias /cgi-bin/ "C:/cgi-Skripte/"
    ScriptAlias /php/ "C:/php/"
</IfModule>
```

Benutzer-Verzeichnisse

- Gliederung der Dokumente nach Benutzern
  - ⇨ z.B. für persönliche Homepages
- Aufruf der Startseite mit ~Username
  - ⇨ z.B. http://www.fbi.h-da.de/~r.hahn
-
```
<IfModule mod_userdir.c>
    UserDir "C:/Benutzerverzeichnisse/"
</IfModule>
```
- Benutzer-Verzeichnisse werden i.a. von den Benutzern selbst gepflegt
  - ⇨ eventuell mit eigenen Berechtigungs-Dateien (`.htaccess`)
  - ⇨ Zugriffsrechte gut überlegen und steuern mit `AllowOverride`

MIME-Types

- HTTP-Header kennzeichnet das beigefügte Dokument mit dessen MIME-Type
  - ⇨ Multipurpose Internet Mail Extension
- Server ermittelt MIME-Type aus Datei-Endung
  - ⇨ Zuordnung gängiger Typen in /conf/mime.types

    ```
    TypesConfig conf/mime.types
    ```
  - ⇨ zusätzliche Definitionen in /conf/httpd.conf

    ```
    AddType application/vnd.ms-excel .csv
    ```
  - ⇨ Standardvorgabe, falls kein MIME-Type ermittelt werden kann

    ```
    DefaultType text/plain
    ```
    oder
    ```
    DefaultType application/octet-stream
    ```
- Browser entscheidet, wie das Dokument dargestellt wird
  - ⇨ was nicht angezeigt werden kann, wird zum Download angeboten
  - ⇨ Netscape verwendet den übermittelten MIME-Type,

    Internet Explorer verwendet die Datei-Endung

Zusätzliche Features mittels Options

- CGI-Skripte ausserhalb des ScriptAlias erlaubt

  `ExecCGI`
- Verknüpfungen zu anderen Verzeichnissen folgen

  `FollowSymLinks` `SymLinksIfOwnerMatch`
- HTML-Vorverarbeitung mittels Server-Side Includes (SSI)

  `Includes` `IncludesNOEXEC`
- Inhaltsverzeichnis zeigen, wenn Indexdatei (z.B. index.html) fehlt

  `Indexes`
- Browser und Server verständigen sich über MIME-Type, Sprache und Codierung (content negotiation); "unscharfe" Dateiendung

  `MultiViews`
- Standardeinstellung: alles außer MultiViews

  `All`

Options x y z setzt neu für dieses Verzeichnis;
Options +x -y akkumuliert mit vererbten Options

Zugriffsschutz in httpd.conf

- Standardeinstellung sehr restriktiv

```
<Directory />
    Options FollowSymLinks          keine sonstigen Options
    AllowOverride None              .htaccess wirkungslos
</Directory>
```

- Wurzelverzeichnis der Dokumente gezielt öffnen

```
<Directory "C:/Programme/Apache/htdocs">
    AllowOverride All               .htaccess wirksam
    Allow from all                  Wurzel freigeben
</Directory>
```

- Dateiname für verzeichnisspezifischen Schutz

```
AccessFileName .htaccess            definieren
<Files ~ "^\.ht">                   beginnt mit .ht
    Deny from all                   und schützen
</Files>
```

Verzeichnis-bezogener Schutz wird an Unter-verzeichnisse vererbt

alles andere muss dann explizit freigegeben werden

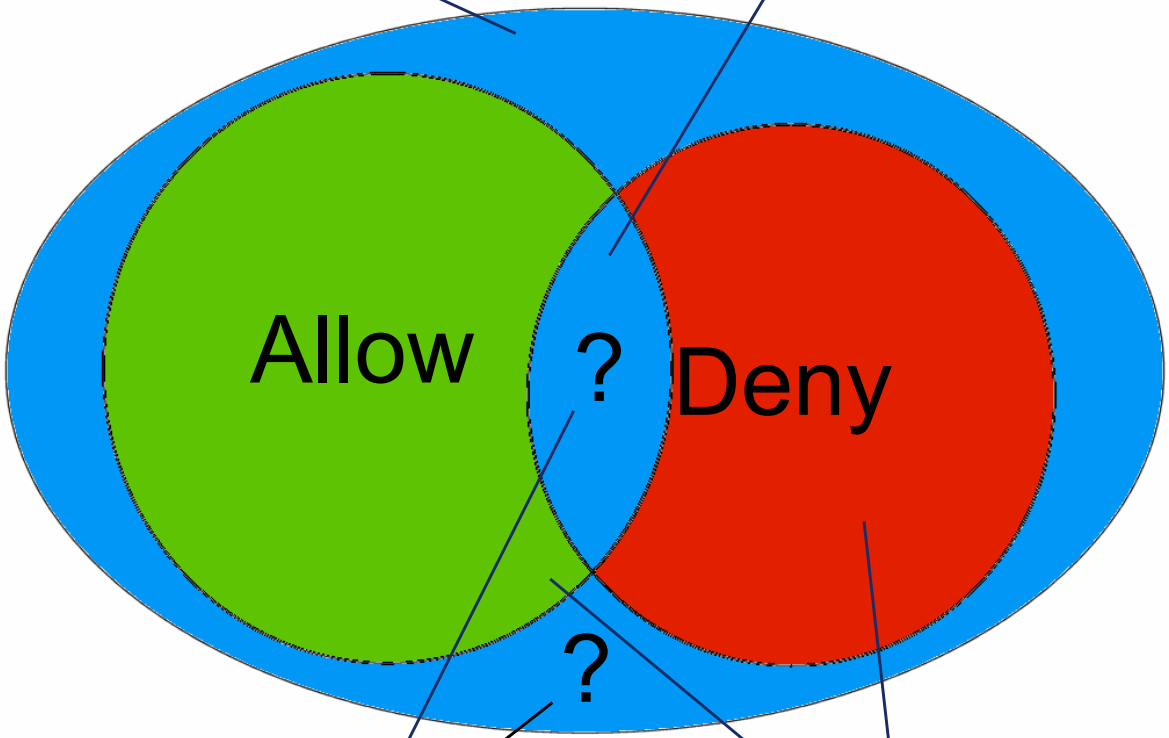
Zugriffsschutz mit .htaccess-Dateien

- Wunsch: Der Benutzer kann die Zugriffsberechtigung selbst pflegen, ohne einen Restart des Webservers zu benötigen
- Lösung: Webserver liest im Benutzerverzeichnis eine Datei mit Zugriffsberechtigungen: *.htaccess*
- Der Name der Dateien muss in httpd.conf festgelegt werden:
  `AccessFileName .htaccess`
- Die Verwendung von lokalen Berechtigungsdateien muss in httpd.conf freigegeben sein mit
  `AllowOverride All`
- .htaccess bezieht sich auf das Verzeichnis, in dem es steht
- Die Berechtigung wird durch 2 "Mengen" beschrieben:
  `deny` und `allow`
- die Auswertungsreihenfolge macht einen Unterschied: `Order allow,deny` oder `Order deny,allow`

Zugriffsberechtigungen als Mengen

weder in Deny noch in Allow enthalten

sowohl in Deny als auch in Allow enthalten



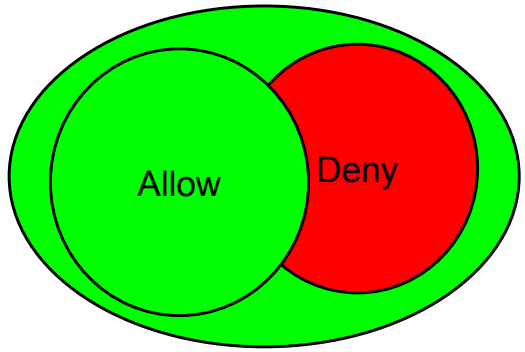
Absicht unklar "Zweifelsfall"

hier ist die Absicht klar

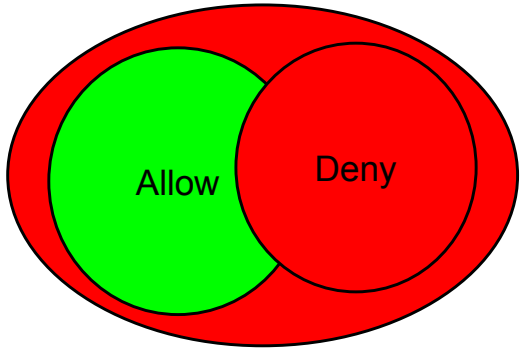
Lösung: Regelung über die Reihenfolge: "order"

Zugriffsschutz mit .htaccess-Dateien - Order

- Order Deny,Allow
  - ⇨ Die Deny Anweisung wird zuerst ausgewertet. Zugang wird als Default erlaubt. Jeglicher Zugriff der nicht in Deny enthalten ist, oder in Allow enthalten ist, darf zugreifen.



- Order Allow,Deny
  - ⇨ Die Allow Anweisung wird zuerst ausgewertet. Zugang wird als Default abgelehnt. Jeglicher Zugriff der nicht in Allow enthalten ist, oder in Deny enthalten ist, wird abgelehnt.



Regel: Order XXX, YYY bedeutet:
Wenn Zuordnung eindeutig ist XXX, im Zweifelsfall YYY

Zugriffsschutz mit .htaccess-Dateien - Beispiele

a) "alle" - offen für die ganze Welt

```
Order Allow,Deny
Allow from all
```

b) "niemand" – im Web nicht verfügbar

```
Order Deny,Allow
Deny from all
```

c) "alle außer…"

```
Order Allow, Deny
Allow from all
Deny from 141.100
```

d) "niemand, außer…"

```
bestimmte IP-Adressen
Order Deny,Allow
Deny from all
Allow from 141.100
```

e) "niemand, außer mit Passwort"

```
AuthType basic
AuthName "Anzeigetext"
AuthUserFile /dir/pw.sec
Require valid-user
```

an sicherem Ort ablegen!
erzeugen mit
htpasswd.exe –c pw.sec user

Sicherheit von Apache-Passwörtern

- Basic Authentification wird quasi im Klartext übertragen und die HTML-Antwort ebenfalls
- Passwort wird im Browser gespeichert und bei jeder Seitenanforderung an denselben Server übermittelt
  - ⇨ notwendig, weil HTTP zustandslos ist
- Username/Passwort ist im Server nur durch schwache Verschlüsselung geschützt
  - ➔ nicht dasselbe Passwort für Website und Bankkonto verwenden !
- Webserver erkennt keine Einbruchsversuche durch Ausprobieren (mehrfach falsches Passwort eingegeben)
  - ⇨ Betriebssysteme erkennen dies üblicherweise
  - ⇨ allenfalls in Log-Datei erkennbar

Zugriffsrechte des Servers selbst

- Sicherheitsmaßnahme, falls Zugriffsrechte nicht sauber und vollständig definiert sein sollten
  - ⇨ soll den Server selbst schützen, weniger die Dokumente
- Apache wird normalerweise vom User "system" (root) gestartet
- Apache startet Child-Prozesse, die die Requests beantworten
- Child-Prozesse können eingeschränkte Zugriffsrechte haben
  - ⇨ User und Group z.B. so konfigurieren, dass
    - nur die freigegebenen Verzeichnisse lesbar sind
    - nur das Nötigste via CGI schreibbar ist

Zusammenfassung

- Grundlagen
  - ⇨ Was ist ein Webserver?
  - ⇨ Webserver am Markt
- Grundeinstellungen
  - ⇨ IP-Adresse & Ports, Log-Dateien
  - ⇨ Verzeichnisse für Dokumente, Skripte, User
  - ⇨ Besondere Dateinamen (index.html, .htaccess,...)
  - ⇨ Pfade zu Skripten und ausführbaren Programmen (z.B. Perl)
  - ⇨ MIME-Typen und Dateiendungen
- Apache
  - ⇨ Grundeinstellungen
  - ⇨ Zugriffsberechtigungen und Zugriffsschutz
- Server Side Includes