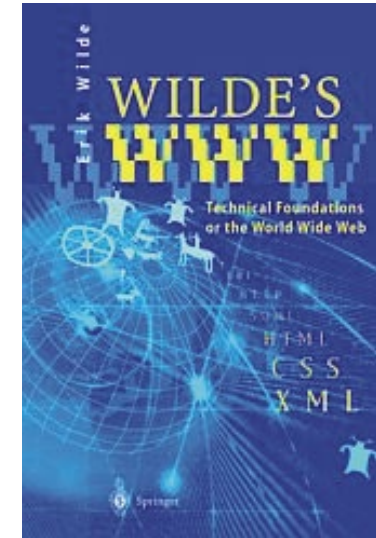


Prof. Dr.-Ing. Georg J. Schneider

Multimedia und Medieninformatik  
Fachbereich Informatik  
Hochschule Trier

## Literatur

Wilde's WWW, Technical Foundations  
of the World Wide Web  
Erik Wilde, Springer



Internetworking: Technische  
Grundlagen und Anwendungen  
Christoph Meinel,  
Harald Sack, Springer

# Elemente eines verteilten Hypermediasystems

- Identifikationsschema
  - Wo sind die Daten
  - Wie kann auf die Daten zugegriffen werden
- Wie werden die Daten dem Benutzer übermittelt?
- Dokumentenformate
  - Inhalt
  - Semantik
  - Präsentation

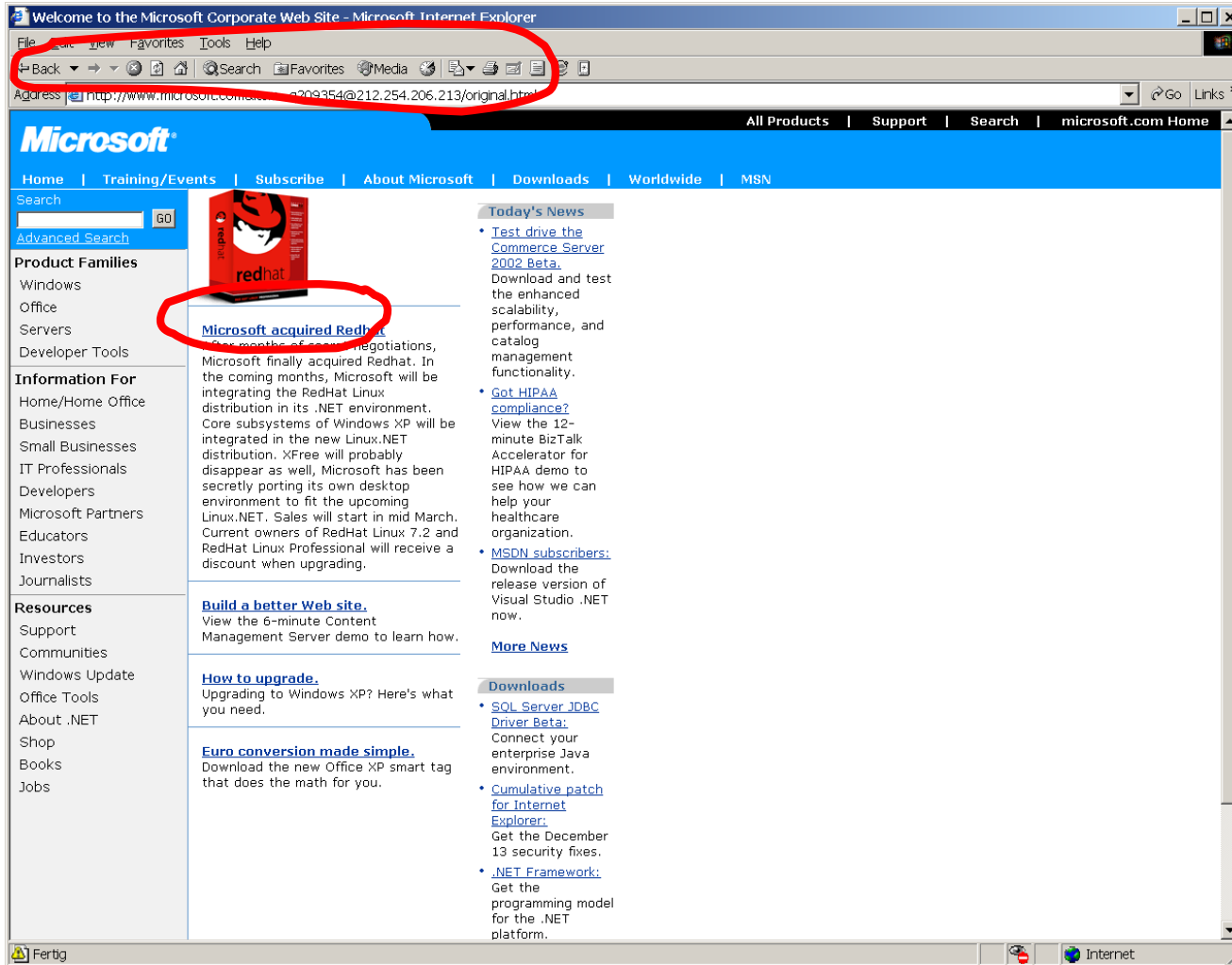
# Elemente des Web

- Identifikationsschema
  - Uniform Resource Identifier, URI
    - Verwendung als Uniform Resource Name, URN
    - Verwendung als Uniform Resource Locator, URL
- Transferprotokoll
  - Hypertext Transfer Protocol, HTTP
    - ASCII-kodiertes Request-Reply Protokoll über TCP/IP
- Dokumentformat
  - Hypertext Markup Language, HTML
    - Abstrakte Sprachspezifikation mit mehreren Ausprägungen

# Elemente des Web

- **Identifikationsschema**
  - Uniform Resource Identifier, URI
    - Verwendung als Uniform Resource Name, URN
    - Verwendung als Uniform Resource Locator, URL
- **Transferprotokoll**
  - Hypertext Transfer Protocol, HTTP
    - ASCII-kodiertes Request-Reply Protokoll über TCP/IP
- **Dokumentformat**
  - Hypertext Markup Language, HTML
    - Abstrakte Sprachspezifikation mit mehreren Ausprägungen

# Beispiel: URL



# „Unglaubliche Nachrichten ...“

**...Microsoft hat den Linux-Distributor Red Hat gekauft...**

**[http://www.microsoft.com&item=q209354@212.254.206.213/1338825GHU\\_98.asp](http://www.microsoft.com&item=q209354@212.254.206.213/1338825GHU_98.asp)**

# Uniforme Identifikation

- Informationsressourcen müssen identifizierbar sein
  - per Name
  - per Adresse bzw. Lokation
- jede Ressource im Internet soll identifizierbar sein
  - Web-Ressourcen, FTP-Ressourcen, News-Ressourcen, Mailboxes, Directories, interaktive Dienste
- Identifikationsschema soll
  - erweiterbar,
  - vollständig und
  - als Zeichenkette darstellbar sein.



# Uniform Resource Identifier (URI)

Syntax für alle Identifikatoren [RFC1630] 6/1994, überarbeitet als [RFC3986] 1/2005

**URI = scheme ":" hier-part [ "?" query ] [ "#" fragment ]**

- **<scheme>**  
bezeichnet Namensschema für diesen URI
- **<scheme-specific-part>**  
enthält aktuelle Identifikation entsprechend des **scheme**
- URIs können sein:
  - Namen: Uniform Resource Name (URN)
  - Lokationen/Adressen: Uniform Resource Locator (URL)

# Uniform Resource Identifier (URI)

## Beispiele

- `ftp://ftp.is.co.za/rfc/rfc1808.txt`
- `http://www.ietf.org/rfc/rfc2396.txt`
- `ldap://[2001:db8::7]/c=GB?objectClass?one`
- `mailto:John.Doe@example.com`
- `news:comp.infosystems.www.servers.unix`
- `tel:+1-816-555-1212`
- `telnet://192.0.2.16:80/`
- `urn:oasis:names:specification:docbook:dtd:xml:4.1.2`

# Uniform Resource Identifier (URI)

[illegible]

```
authority = [ userinfo "@" ] host [ ":" port ]
```

# Sonderzeichen

- Hierarchie-Zeichen /  
dies hat nichts mit dem / im Dateisystem zu tun
- Fragment Delimiter #  
identifiziert ein Fragment innerhalb einer Ressource
- Query Delimiter ?  
grenzt den Anfragestring vom Namen der Ressource ab
- Abtrennung vom **scheme** :
- Escape-Zeichen %  
zum Beispiel genutzt, um Leerzeichen zu kodieren:  
**name%20nachname**

# Sonderzeichen

## Reservierte Zeichen

reserved = gen-delims / sub-delims

gen-delims = ":" "/" "?" "#" "[" "]" "@"

sub-delims = "!" "\$" "&" " ' " "(" ")" "\*" "+" ", " ";" "="

# URN Eigenschaften

- global eindeutig (Global Scope and Uniqueness)
- dauerhaft beständig (Persistence)
- skalierbar
- unterstützt bestehende Applikationen
- erweiterbar
- unabhängig
- auflösbar

# Uniform Resource Name (URN)

- Vereinheitlichung jeglicher Namensgebung
- URN [RFC 1737] 12/1994, [RFC 8141] 4/2017  
**<scheme> ::= "urn"**

**<urn> ::= "urn:" <nid> ":" <nss>**

- **nid** = Namespace Identifier
- **nss** = Namespace Specific String

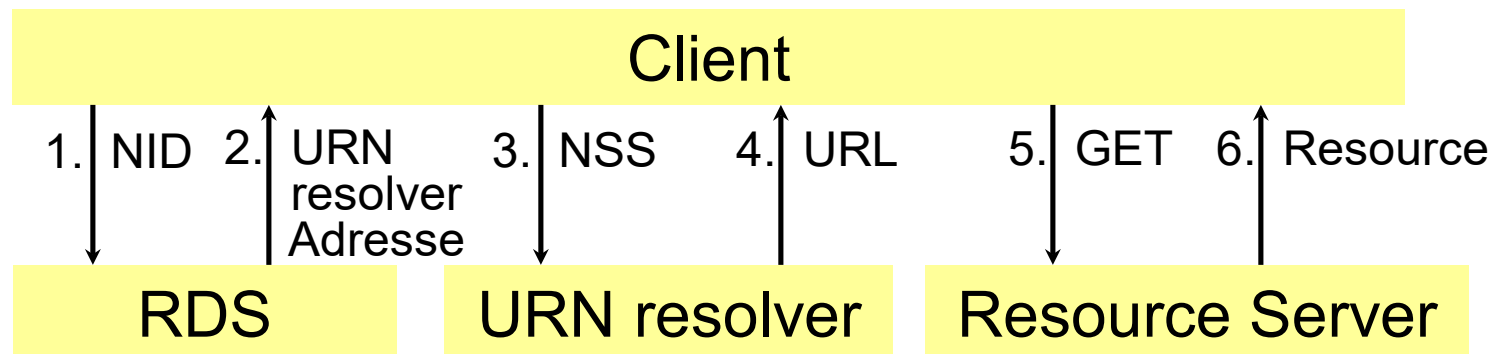
Beispiel

**urn:ISBN:3-540-64285-4**

## URN - Auflösung

- Infrastruktur für URNs, zum Teil noch im experimentellen Stadium
  - Resolver Discovery Service (RDS)
  - Namensdienste, Namensauflösung (URN resolver)
  - Auflösung gegen URL oder URC
  - Weitere Informationen: [RFC 1737, 2276]

**<urn> ::= "urn:" <nid> ":" <nss>**





## URN - Verwendung

- Verwaltung der Namespaces: <http://www.iana.org/assignments/urn-namespaces>
- Namespace für National Bibliography Numbers: "nbn" [RFC3188]
- Die Deutsche Bibliothek <http://www.persistent-identifier.de/> vergibt URNs aus dem Namensraum "**nbn:de**" für:
  - Online-Hochschulschriften (Dissertationen, Habilitationen)
  - Forschungsberichte, online publizierte zeitschriftenartige Reihen
  - Zeitschriftenartikel, Diplom- und Magisterarbeiten, Festschriften

Aufbau: z.B. "**URN:NID:SNID-NISS**" (Subnamespaces, **SNID**)

URNs der Deutschen Nationalbibliothek haben die folgende allgemeine Struktur:  
**urn:nbn:de:[Verbundabk.]:[Sigelnummer]-[Nummer][Prüfziffer]**

<https://nbn-resolving.org>

- **urn:nbn:de:bsz:291-scidok-2295** (Dissertation Georg Schneider)

# Uniform Resource Locator (URL)

- Vereinheitlichung jeglicher Adressangabe
- URL **scheme** Definitionen [RFC1738]
  - **http**    - **https**    - **ftp**    - **news**
  - **nntp**    - **mailto**    - **telnet**    - **ldap** und andere
- darunter **scheme-specific-part** Definitionen der generellen Form  
["//"] [user [ ":"password] "@"] host [ ":"port] [ "/"url-path]
- Definitionen verwaltet die  
Internet Assigned Numbers Authority (IANA)

# HTTP URL I

## Einschränkung des generellen **scheme-specific-part**

```
<http_URL> = "http://" <host> [ ":" <port> ] [<abs_path>]
```

```
<host>      = <A legal Internet host domain name or IP  
              address (in dotted-decimal form),  
              as defined by Section 2.1 of RFC 1123>
```

```
<port>      = *DIGIT
```

```
<abs_path> = "/" [<path>] [ ";" <params> ] [ "?" <query> ] [ "#" <fragment> ]
```

```
<path>      = <fsegment> *( "/" <segment> )
```

# Uniform Resource Locator (URL)

## Relative URLs [RFC 1808]

Beispiel:

Netz Pfad `//de.wikipedia.org/styles.css`

Absoluter Pfad `/aufgaben/uebung1/`

Relativer Pfad `../images/image.png`

# HTTP URL II

## Beispiele für die Verwendung des Escape Characters:

`http://www.informatik.hochschule-trier.de/~schneider/frage.png`

`http://www.INFORMATIK.hochschule-trier.de/%7Eschneider/frage.png`

- URLs unterscheiden zwischen Groß- und Kleinschreibung
- Ausnahme: Host Name

# Top Level Domains I

## "country-code" TLDs oder "ccTLDs"

.de	Deutschland
.fr	Frankreich
.jp	Japan

## "generic" TLDs oder "gTLDs "

- Sponsored
- Un-sponsored

## Top Level Domains II

### "generic" TLDs oder "gTLDs"

<b>.com</b>	kommerzielle Anbieter*
<b>.net</b>	Netzwerk Anbieter*
<b>.org</b>	Organisationen*
<b>.edu</b>	Ausbildung
<b>.gov</b>	Verwaltung
<b>.mil</b>	US Militär
<b>.int</b>	Internationale Verbände

### unsponsored

**.info,**  
**.biz**  
**.name**  
**.pro**

### sponsored

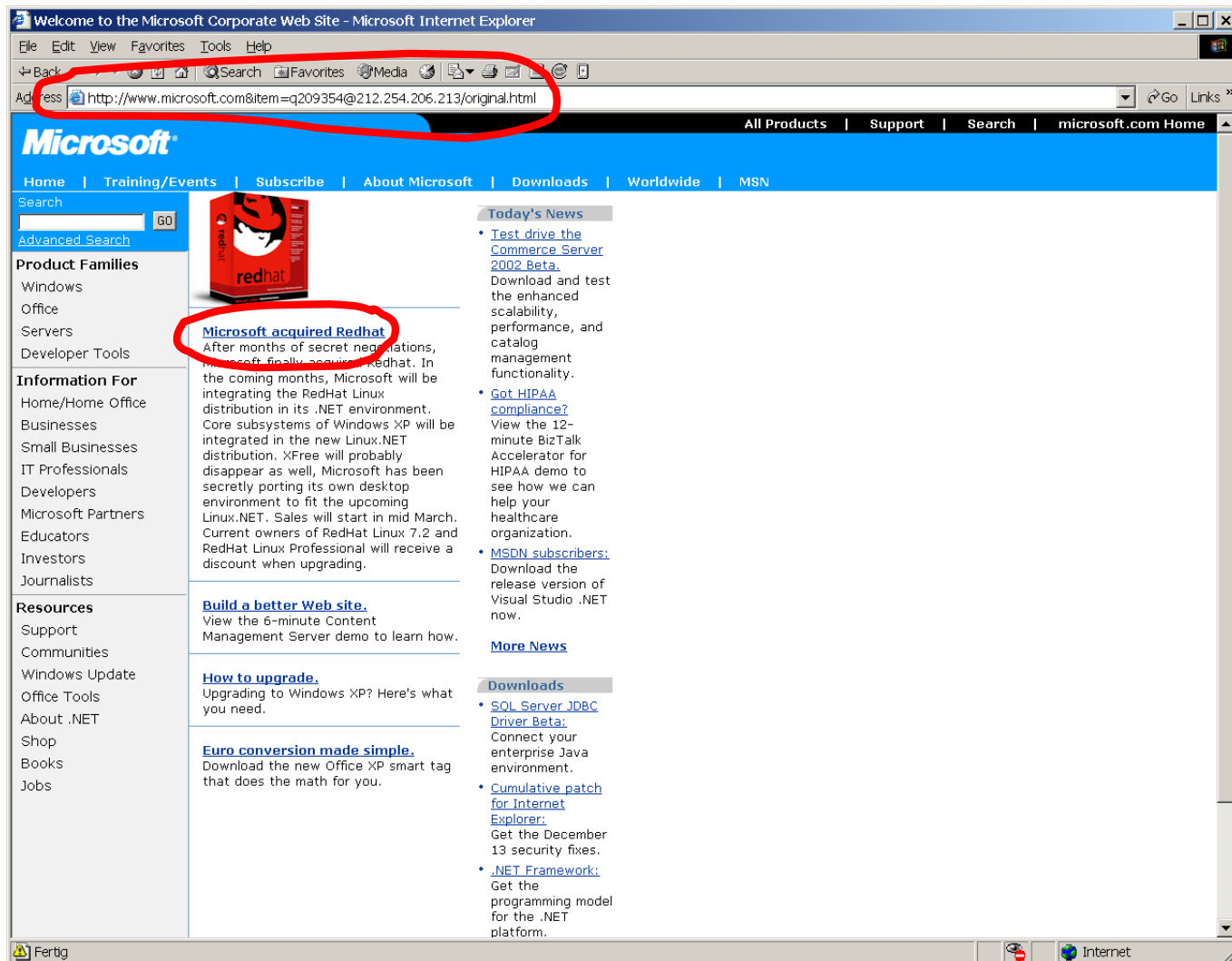
**.aero**  
**.coop**  
**.museum**

\* uneingeschränkte Registrierung

<https://www.iana.org/domains/root/db>

[https://en.wikipedia.org/wiki/List\\_of\\_Internet\\_top-level\\_domains](https://en.wikipedia.org/wiki/List_of_Internet_top-level_domains)

# Beispiel: URL





# „Unglaubliche Nachrichten hinter falschen URLs“

**...Ein anderer meldet, Microsoft habe den Linux-Distributor  
Red Hat gekauft**

**([http://www.microsoft.com&item=q209354@212.254.206.213/  
1338825GHU\\_98.asp](http://www.microsoft.com&item=q209354@212.254.206.213/1338825GHU_98.asp)). ...“**

Heise Newsticker (<http://www.ix.de>), Meldung vom 10.02.2002 11:31