

# Tutorial zu Kapitel WT:III

## III. Dokumentensprachen

- HTML
- CSS
- XML-Grundlagen
  - XML-Dokumentstruktur
  - Document Type Definition (DTD)
  - Namespaces
- XML-Schema
- XSL-Familie
  - XPath
  - XSLT

Die nach folgenden Erklärungen basieren auf [W3C HTML 5](#). Vereinfachungen an verschiedenen Stellen sollen das grundsätzliche Verständnis erleichtern.

# HTML

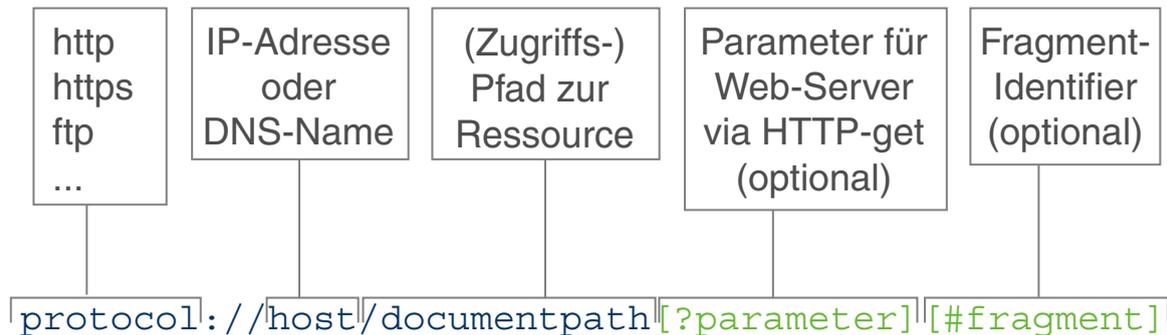
## Identifizierung von Ressourcen im WWW

- URI [\[RFC 3986\]](#)

Identifiziert eindeutig eine Informationsressource im WWW, unabhängig davon, ob es sich um ein Hypermedia-Dokument handelt.

- Spezialfall der URI: URL [\[RFC 1738, RFC 3986\]](#)

Identifiziert über eine eindeutige Adresse den Ort (*Location*) einer Informationsressource im WWW.



### Typisches Beispiel:

`http://www.cs.upb.de/fachgebiete/fg-kleine-buening/lehre/webis.html`

# HTML

## Ausprägungen von URLs (vereinfacht)

- ❑ Absolute URLs [\[RFC 3986\]](#)  
enthalten Protokoll, Host, Dokumentenpfad, Parameter (optional) und Fragment (optional).
- ❑ Relative URLs [\[RFC 3986\]](#)  
enthalten Dokumentenpfad, Parameter (optional) und Fragment (optional).  
Der Dokumentenpfad darf nicht leer sein. Der Dokumentenpfad kann absolut sein (beginnend mit "/") oder relativ.
- ❑ Relative URLs mit gleicher Ressource [\[RFC 3986\]](#)  
enthalten nur ein Fragment.

(Eine Unterscheidung zwischen den zwei relativen URL-Typen ist sinnvoll, da z.B. in einem Browser für eine Referenz auf das gleiche Dokument kein erneutes Laden des Dokumentes erforderlich ist, wenn die Basis-URL die des Dokumentes ist.)

→ Relative URLs müssen für einen Zugriff auf die Ressource zu absoluten URLs vervollständigt werden.

# HTML

## Basis-URL von HTML-Dokumenten

1. Falls im Header des HTML-Dokumentes kein `base` Element mit einem `href` Attribut auftritt, so ist die URL des HTML-Dokumentes die Basis-URL.
2. Falls im Header des HTML-Dokumentes `base` Elemente mit einem `href` Attribut auftreten, so führe folgende Schritte aus:
  - (a) Verwende die URL aus dem `href` Attribut des ersten `base` Elementes im HTML-Dokument.
  - (b) Ist diese URL eine absolute URL, so verwende sie (nach Entfernung von Fragmenten) als Basis-URL.
  - (c) Ist diese URL eine relative URL, so vervollständige sie mit Hilfe der URL des HTML-Dokumentes zu einer absoluten URL und verwende diese (nach Entfernung von Fragmenten) als Basis-URL.

# HTML

## Vervollständigung von relativen URLs mit Hilfe einer Basis URL

1. Falls die relative URL einen absoluten Dokumentenpfad enthält, so hänge die relative URL an Protokoll- und Host-Angabe aus der Basis-URL.

Die so entstandene absolute URL ist die gesuchte URL.

2. Falls die relative URL einen relativen Dokumentenpfad enthält, so entferne eine (eventuelle) Datei-angabe am Ende des Dokumentenpfades der Basis-URL sowie vorhandene Parameter in der Basis-URL. Hänge die relative URL an den Rest der Basis-URL an. (Im relativen Pfad eventuell vorhandene Pfadsegmente "." und ".." müssen durch passendes Reduzieren des Pfades eliminiert werden.)

Die so entstandene absolute URL ist die gesuchte URL.

Eine absolute URL muss im Dokumentenpfad keine Datei-angabe enthalten. Fehlt eine Datei-angabe, so entscheidet der Web Server, welche Ressource zurückgegeben wird, wenn vorhanden. Ebenso kann ein Anfrage mit einer URL vom Web Server auf eine andere URL umgeleitet werden. Ist der Zugriff auf eine Ressource mit einer angegebenen URL nicht möglich, so wird vom Web Server eine Fehlermeldung (i.d.R. Code 404) zurück gegeben.)