

# Protokół HTTP

## Coś o HTTP:

- Początki protokołu HTTP (ang. Hypertext Transfer Protocol) sięgają 1990 roku. Od wersji 1.0 pozwala on na przesyłanie poza czystym tekstem także różnego rodzaju danych o charakterze binarnym (obrazy, dźwięki, animacje). Obecnie obowiązuje wersja HTTP/1.1
- Istotą WWW jest struktura hipertekstowa. Hipertekst rozumiany jest jako tekst z możliwością tworzenia odnośników do innych treści/dokumentów
- Na WWW składają się, poza protokołem HTTP, który służy do przesyłania informacji także: HTML (ang. Hypertext Markup Language) czyli język tworzenia dokumentów oraz URI (ang. Universal Resource Identifier) czyli sposób jednoznacznego identyfikowania zasobów (w HTTP używane są URL – UR Locators)

## Ogólny schemat internetowego URL:

```
<scheme>://<user>:<password>@<host>:<port>/<url-path>;<params>?<query>#<fragment>
```

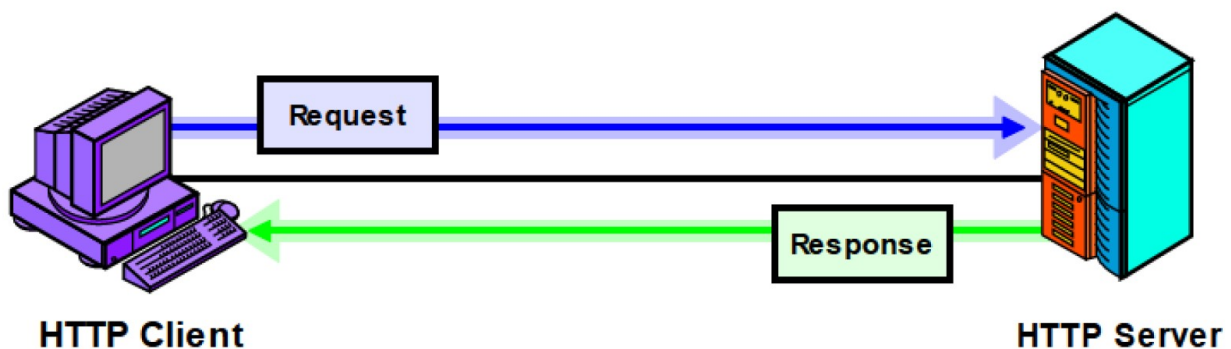
## Składnia URL w HTTP:

```
http://<user>:<password>@<host>:<port>/<urlpath>?<query>#<bookmark>
```

## Wyjaśnienie:

- url-path – identyfikator zasobu na serwerze HTTP; ma tu znaczenie wielkość liter
- query – zapytanie przetwarzane przez serwer HTTP, pozwala to na interakcję użytkownika ze stroną (alternatywa to metody POST)
- bookmark – zakładka do określonego fragmentu dokumentu

HTTP pracuje w architekturze **klient-serwer**, definiując żądania i odpowiedzi na nie



HTTP wykorzystuje do przesyłania wiadomości protokół **TCP** z nawiązywaniem indywidualnych połączeń. HTTP/1.1 podtrzymuje połączenia tworząc sesje.

## Ogólna struktura wiadomości HTTP:

<linia początkowa>  
<nagłówek wiadomości>  
<pusta linia>  
[treść wiadomości]  
[zakończenie]

### Omówienie żądania:

- Linia żądania – zawiera metodę, żądany zasób i wersję protokołu HTTP
- Ogólny nagłówek – data, godzina, ustawienia dotyczące buforowania treści, połączenie (czy ma być podtrzymane), kodowania, śledzenia trasy wiadomości (opcja Via), ostrzeżenia
- Nagłówek żądania – specyficzny dla metody, zawiera zwykle akceptowane przez klienta rodzaje treści, kodowanie (dyrektywy accept), autoryzację, zwracanie treści tylko gdy została zmodyfikowana (if-modified), odnośnik jaki doprowadził do żądanej treści (! – pole referer), oprogramowanie klienta (user-agent), fragment zasobu (range), adres e-mail użytkownika (o ludzka naiwności...)
- Treść wiadomości

### żądanie:

GET /index.html HTTP/1.1	<b>Request Line</b>	<b>HTTP Request</b>
Date: Thu, 20 May 2004 21:12:55 GMT Connection: close	<b>General Headers</b>	
Host: www.myfavoriteamazingsite.com From: joebloe@somewebsitesomewhere.com Accept: text/html, text/plain User-Agent: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)	<b>Request Headers</b>	
	<b>Entity Headers</b>	
	<b>Message Body</b>	

### Metody HTTP:

Podstawowe:

- GET – pobierz zasób
- HEAD – to samo co GET ale bez właściwej treści (używana np. po to by sprawdzić czy zasób istnieje)
- POST – przesłanie danych do serwera, URL określa np. nazwę programu który dane ma przetworzyć

Rzadko spotykane:

- OPTIONS – prośba o informacje o dostępnych opcjach transmisji
- PUT – umieszczanie informacji na serwerze pod adresem URL
- DELETE i TRACE – skasuj zasób (genialne!), prośba o przesłanie kopii żądania (dla celów diagnostycznych)

### Omówienie odpowiedzi:

- Linia statusu – informacja zwrotna od serwera (wersja, kod statusu)
- Ogólny nagłówek (jak w żądaniu)
- Nagłówek odpowiedzi – zawiera np. wiek zasobu (age), właściwą jego lokalizację (location), instrukcję by powtórzyć żądanie po x sekundach, wersję serwera, metodę uwierzytelniania
- Nagłówek jednostki – dotyczą przenoszonego przez wiadomość zasobu, np. jego kodowania, obsługiwanych na przyszłość metod, długości, sumy kontrolnej, fragmentu większej całości jaki wiadomość stanowi, typ treści, datę ostatniej modyfikacji, „datę ważności”
- Treść wiadomości

### Odpowiedź:

HTTP/1.1 200 OK	Status Line
Date: Thu, 20 May 2004 21:12:58 GMT Connection: close	General Headers
Server: Apache/1.3.27 Accept-Ranges: bytes	Response Headers
Content-Type: text/html Content-Length: 170 Last-Modified: Tue, 18 May 2004 10:14:49 GMT	Entity Headers
<b>HTTP Response</b>	
<html> <head> <title>Welcome to the Amazing Site!</title> </head> <body> <p>This site is under construction. Please come back later. Sorry!</p> </body> </html>	Message Body

### Kody odpowiedzi serwera:

- 1xx – wiadomość informacyjna np. 101 – zgoda na zmianę wersji protokołu),
- 2xx – sukces np. 200 – żądanie wykonane, 201 – stworzono zasób (wynik metody PUT), 203 – informacja ze źródła nieautorytatywnego,
- 3xx – przekierowanie np. 303 – użyj innego URL, 305 – użyj proxy
- 4xx – błąd po stronie klienta np. 400 – złe żądanie, 401 – brak praw dostępu (autoryzacji), 402 – wymagana opłata , 404 – brak zasobu, 408 – minął czas oczekiwania na reakcję klienta
- 5xx – błąd po stronie serwera np. 500 – błąd wewnętrzny serwera, 501 – nie zaimplementowano, 503 – usługa niedostępna

