Web Technologies

Emilio Coppa
coppa@diag.uniroma1.it

Sapienza University of Rome

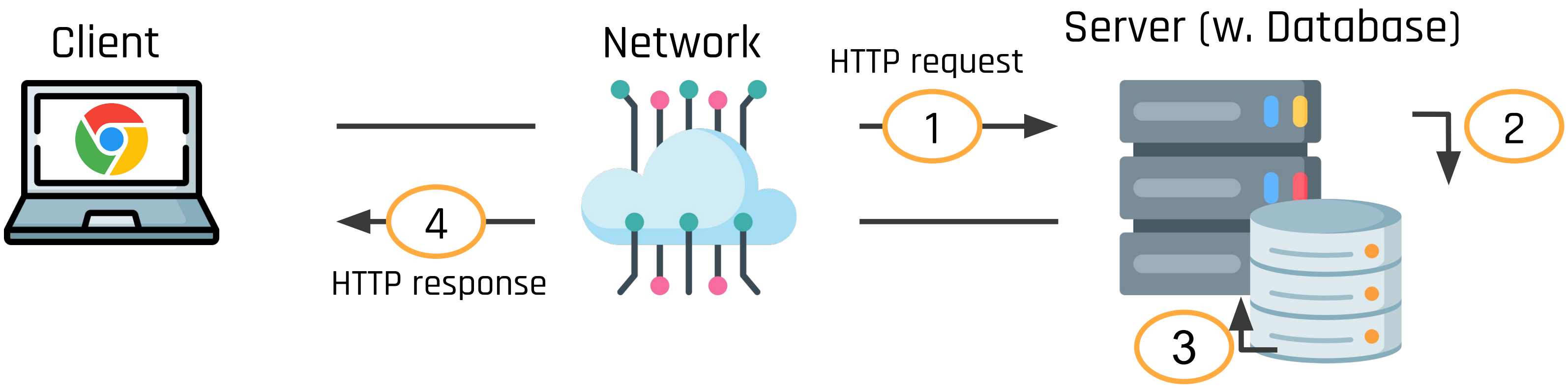
Credits

These slides are based on teaching material originally created by:

- Marco Squarcina (marco.squarcina@tuwien.ac.at), S&P Group, TU WIEN
- Mauro Tempesta (mauro.tempesta@tuwien.ac.at), S&P Group, TU WIEN
- Fabrizio D'Amore (damore@diag.uniroma1.it), Sapienza University of Rome

Introduction to HTTP

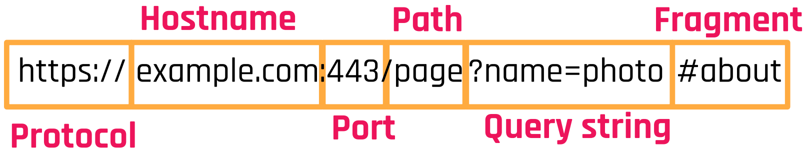
Anatomy of a Typical Web Application



1. The user request a webpage with dynamically generated content
2. The web application queries the database for user's data
3. The data from the database is used to generate page content
4. The page is rendered by the client's browser

Uniform Resource Locator (URL)

URLs are identifiers for documents on the Web



Hostname Path Fragment

https:// example.com :443 /page ?name=photo #about

Protocol Port Query string

- Some elements are optional: port, query string, fragment
- When reserved characters (like space : ? /) need to be used in the URL, they must be URL-encoded:
    - %20 = space
    - %2F = /
    - ...

NOTE: For clarity, we will not URL-encode the attack payloads in the next slides

Example of encoding:

https://example.com/page?name=my%20page

The HTTP Protocol

- HTTP (Hypertext Transfer Protocol) defines the structure of the communication between client and web server
- Properties:
    - **Stateless**: different requests are processed independently from each other
        - Cookies are used to implement stateful applications on top of HTTP
    - **Not encrypted:** HTTP traffic can be read and modified on the network without the communication parties to notice it
    - Default port for HTTP is 80

The HTTPS Protocol

- HTTPS is the secure variant of HTTP:
  - Essentially, HTTP traffic delivered over a TLS connection
  - Default port is 443
- Security properties:
  - **Confidentiality**: content of the traffic cannot be inspected as it travels on the network
  - **Integrity**: content of the traffic cannot be modified as it travels on the network
  - **Authentication**: the client can verify that it is communicating with the expected server

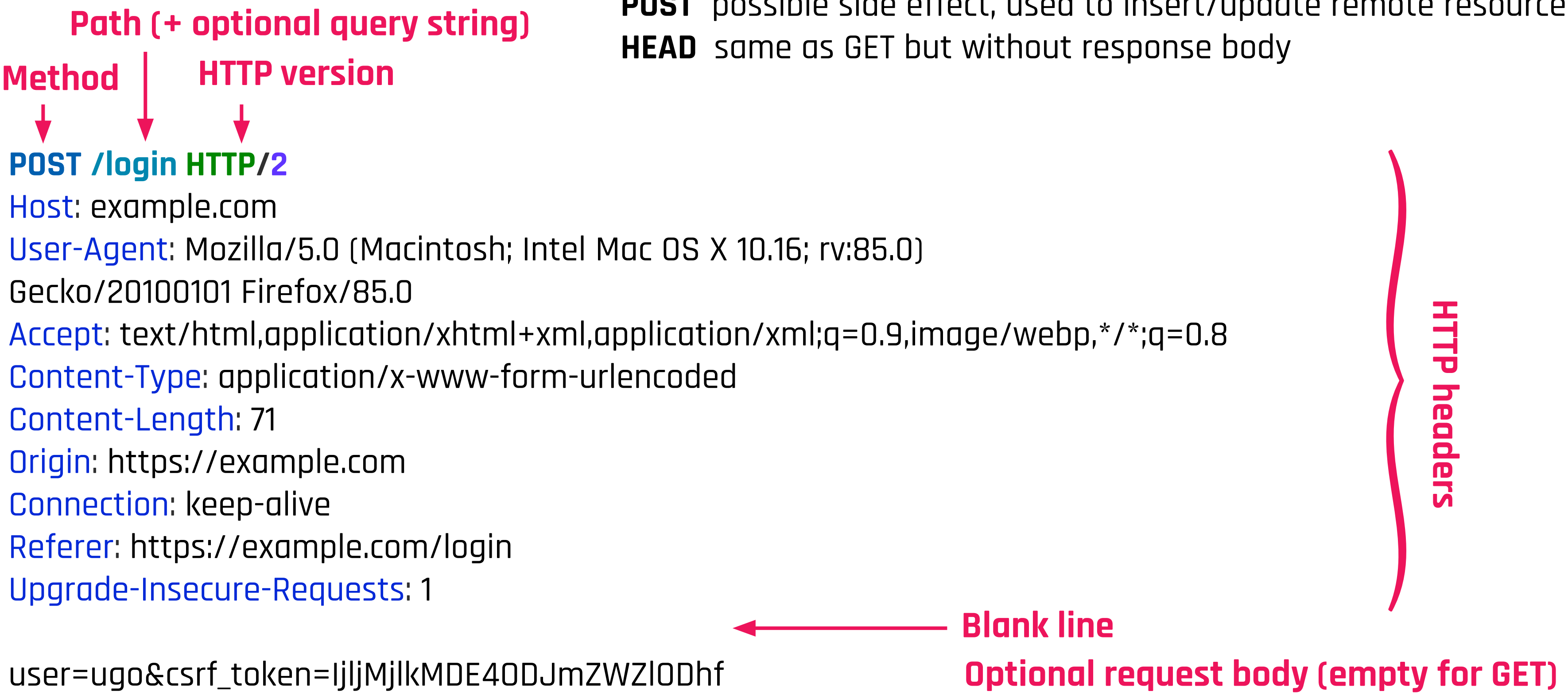
HTTP Request

Most common HTTP Methods:

GET should have no side effects, used to retrieved data

POST possible side effect, used to insert/update remote resources

HEAD same as GET but without response body



Path (+ optional query string)

Method

HTTP version

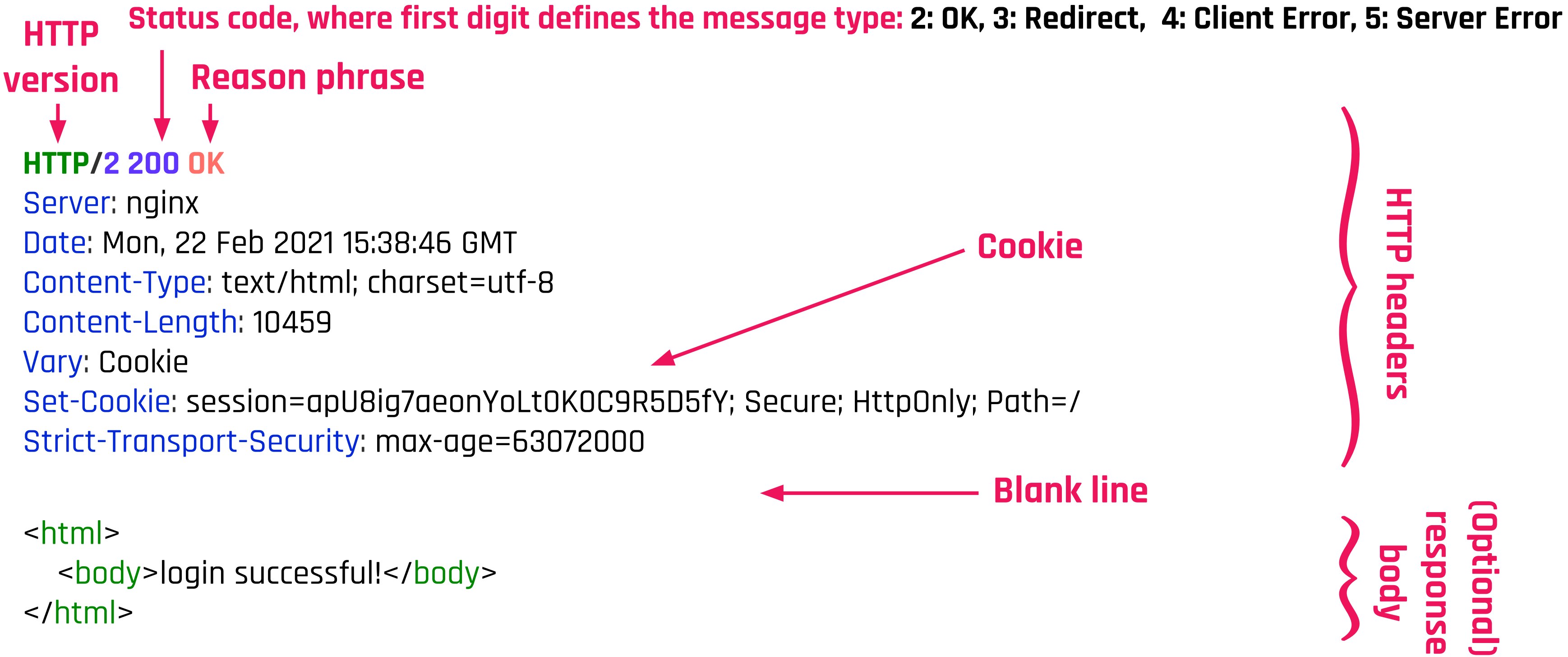
```
POST /login HTTP/2
Host: example.com
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10.16; rv:85.0)
Gecko/20100101 Firefox/85.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Content-Type: application/x-www-form-urlencoded
Content-Length: 71
Origin: https://example.com
Connection: keep-alive
Referer: https://example.com/login
Upgrade-Insecure-Requests: 1

user=ugo&csrf_token=IjljMjlkMDE4ODJmZWZlODhf
```

HTTP headers

Blank line

Optional request body (empty for GET)

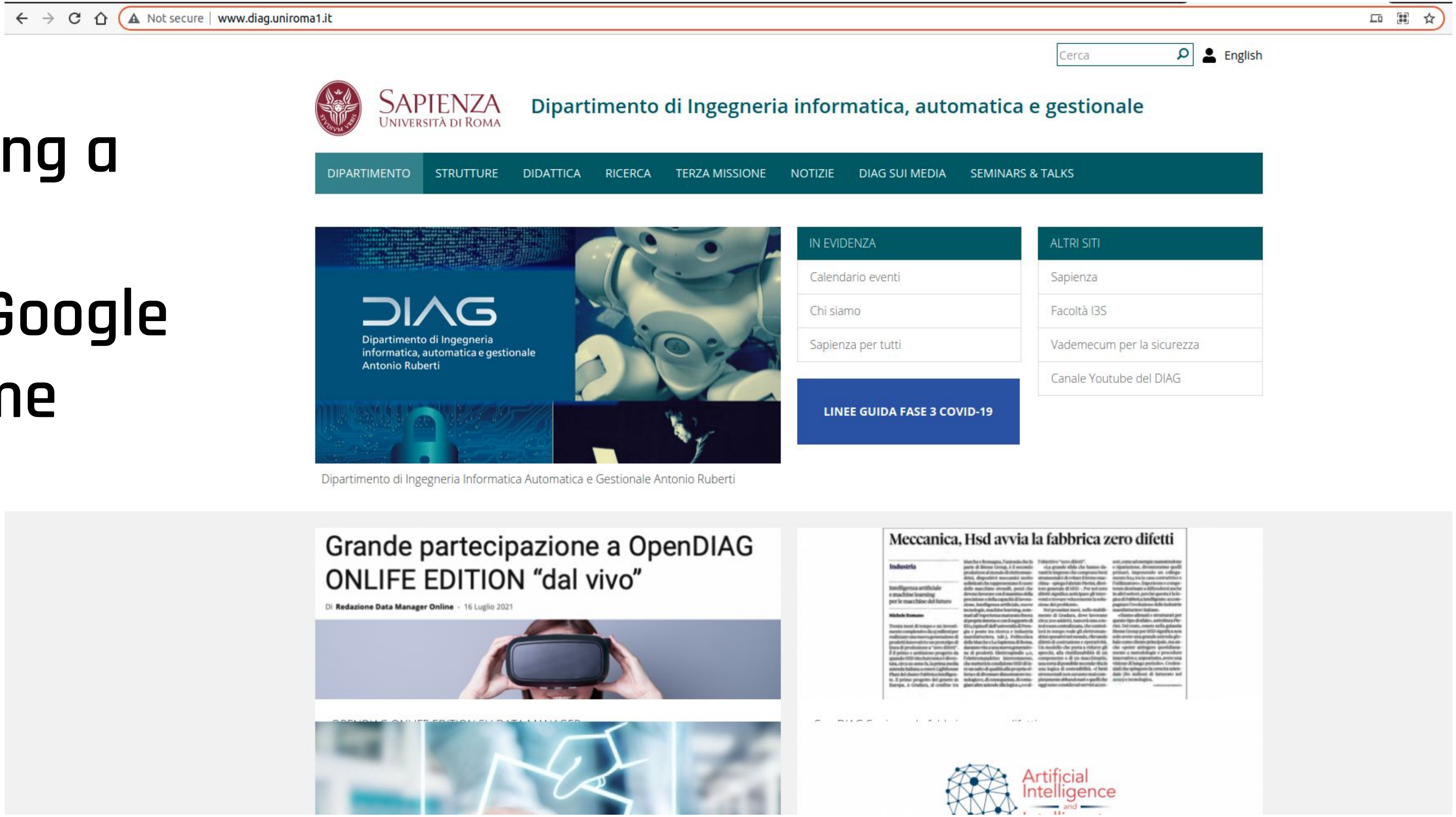
HTTP Response



HTTP version
Status code, where first digit defines the message type: 2: OK, 3: Redirect, 4: Client Error, 5: Server Error
Reason phrase
Cookie
Blank line
HTTP headers
(Optional) response body

```
HTTP/2 200 OK
Server: nginx
Date: Mon, 22 Feb 2021 15:38:46 GMT
Content-Type: text/html; charset=utf-8
Content-Length: 10459
Vary: Cookie
Set-Cookie: session=apU8ig7aeonYoLt0K0C9R5D5fY; Secure; HttpOnly; Path=/
Strict-Transport-Security: max-age=63072000

<html>
  <body>login successful!</body>
</html>
```

Opening a page with Google Chrome



Not secure | www.diag.uniroma1.it

Cerca

English

SAPIENZA UNIVERSITÀ DI ROMA

Dipartimento di Ingegneria informatica, automatica e gestionale

DIPARTIMENTO | STRUTTURE | DIDATTICA | RICERCA | TERZA MISSIONE | NOTIZIE | DIAG SUI MEDIA | SEMINARS & TALKS

DIAG

Dipartimento di Ingegneria informatica, automatica e gestionale Antonio Ruberti

IN EVIDENZA

- Calendario eventi
- Chi siamo
- Sapienza per tutti

LINEE GUIDA FASE 3 COVID-19

ALTRI SITI

- Sapienza
- Facoltà I3S
- Vademecum per la sicurezza
- Canale Youtube del DIAG

Dipartimento di Ingegneria Informatica Automatica e Gestionale Antonio Ruberti

Grande partecipazione a OpenDIAG ONLIFE EDITION "dal vivo"

Di **Redazione Data Manager Online** - 16 Luglio 2021

Meccanica, Hsd avvia la fabbrica zero difetti

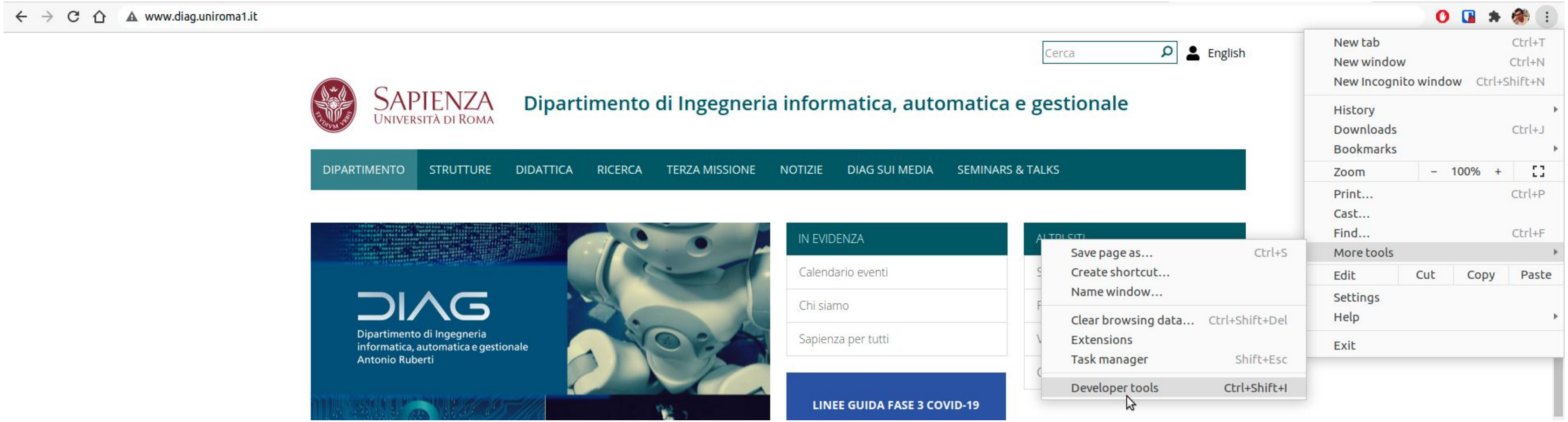
Industria

Intelligenza artificiale e machine learning per le macchine del futuro

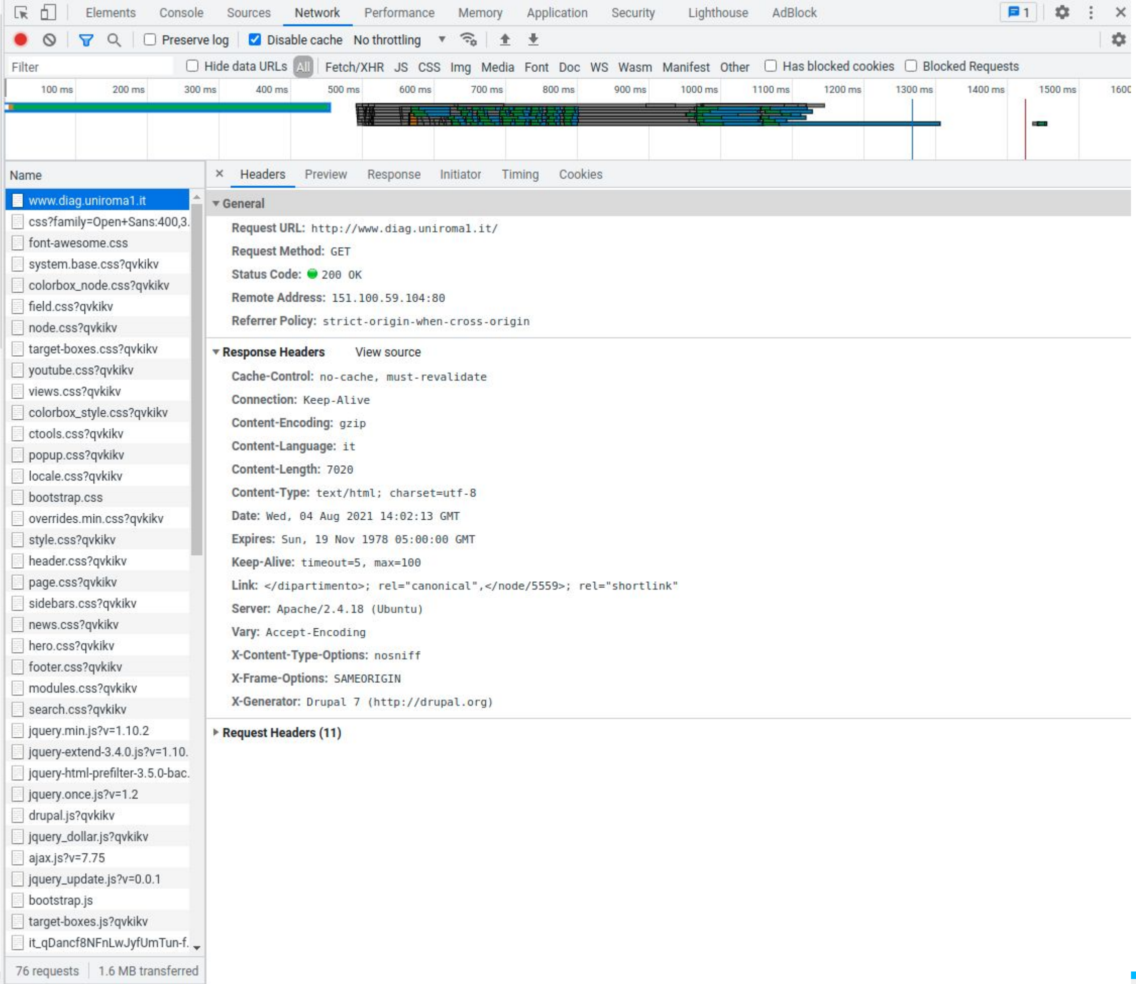
Michele Romano

Artificial Intelligence and

Google Chrome: Developers tools



1. Select Network
2. Refresh the page
3. Choose a request
4. Inspect request and response



☐ Preserve log ☑ Disable cache No throttling

Name

- www.diag.uniroma1.it
- css?family=Open+Sans:400,3.
- font-awesome.css
- system.base.css?qvkikv
- colorbox_node.css?qvkikv
- field.css?qvkikv
- node.css?qvkikv
- target-boxes.css?qvkikv
- youtube.css?qvkikv
- views.css?qvkikv
- colorbox_style.css?qvkikv
- ctools.css?qvkikv
- popup.css?qvkikv
- locale.css?qvkikv
- bootstrap.css
- overrides.min.css?qvkikv
- style.css?qvkikv
- header.css?qvkikv
- page.css?qvkikv
- sidebars.css?qvkikv
- news.css?qvkikv
- hero.css?qvkikv
- footer.css?qvkikv
- modules.css?qvkikv
- search.css?qvkikv
- jquery.min.js?v=1.10.2
- jquery-extend-3.4.0.js?v=1.10.
- jquery-html-prefilter-3.5.0-bac.
- jquery.once.js?v=1.2
- drupal.js?qvkikv
- jquery_dollar.js?qvkikv
- ajax.js?v=7.75
- jquery_update.js?v=0.0.1
- bootstrap.js
- target-boxes.js?qvkikv
- it_qDancf8NFnLwJyfUmTun-f.

76 requests | 1.6 MB transferred

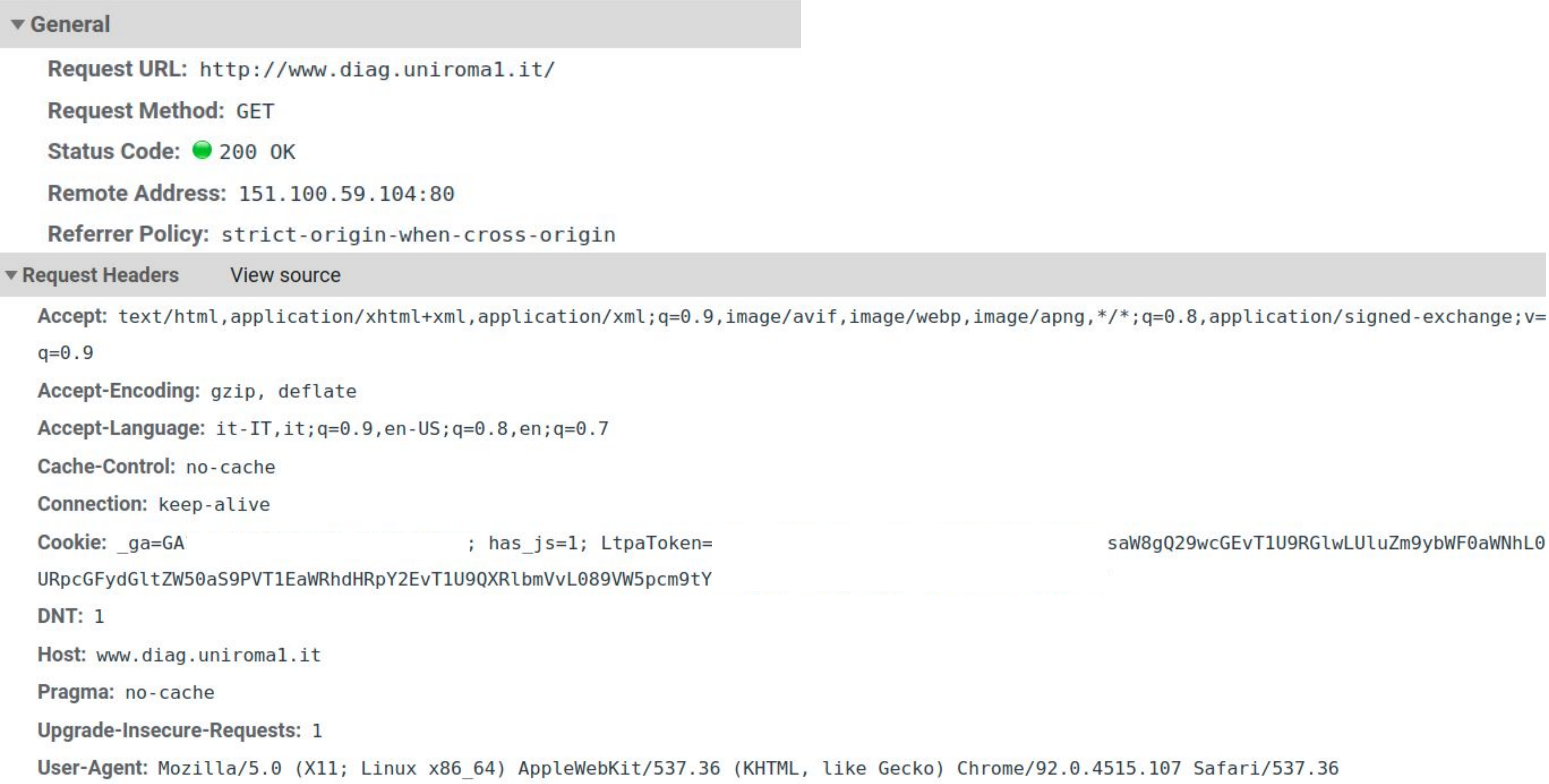
▾ General

Request URL: http://www.diag.uniroma1.it/
Request Method: GET
Status Code: ● 200 OK
Remote Address: 151.100.59.104:80
Referrer Policy: strict-origin-when-cross-origin

▾ Response Headers   View source

Cache-Control: no-cache, must-revalidate
Connection: Keep-Alive
Content-Encoding: gzip
Content-Language: it
Content-Length: 7020
Content-Type: text/html; charset=utf-8
Date: Wed, 04 Aug 2021 14:02:13 GMT
Expires: Sun, 19 Nov 1978 05:00:00 GMT
Keep-Alive: timeout=5, max=100
Link: </dipartimento>; rel="canonical",</node/5559>; rel="shortlink"
Server: Apache/2.4.18 (Ubuntu)
Vary: Accept-Encoding
X-Content-Type-Options: nosniff
X-Frame-Options: SAMEORIGIN
X-Generator: Drupal 7 (http://drupal.org)

▸ Request Headers (11)



▼ General

Request URL: http://www.diag.uniroma1.it/

Request Method: GET

Status Code: ● 200 OK

Remote Address: 151.100.59.104:80

Referrer Policy: strict-origin-when-cross-origin

▼ Request Headers   View source

Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=
q=0.9

Accept-Encoding: gzip, deflate

Accept-Language: it-IT,it;q=0.9,en-US;q=0.8,en;q=0.7

Cache-Control: no-cache

Connection: keep-alive

Cookie: _ga=GA ; has_js=1; LtpaToken= saW8gQ29wcGEvT1U9RGlwLUluZm9ybWF0aWNhL0
URpcGFydGltZW50aS9PVT1EaWRhdHRpY2EvT1U9QXRlbmVvL089VW5pcm9tY

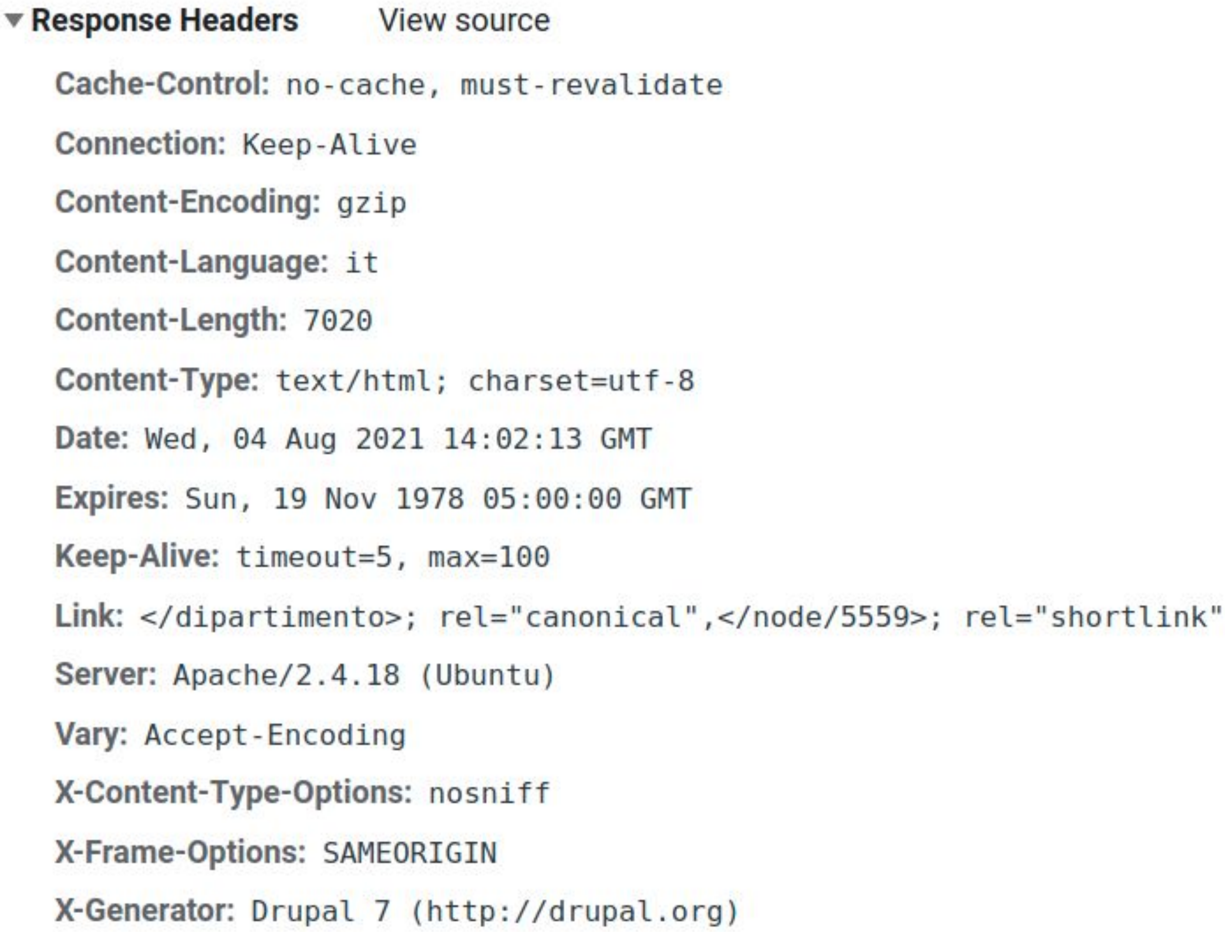
DNT: 1

Host: www.diag.uniroma1.it

Pragma: no-cache

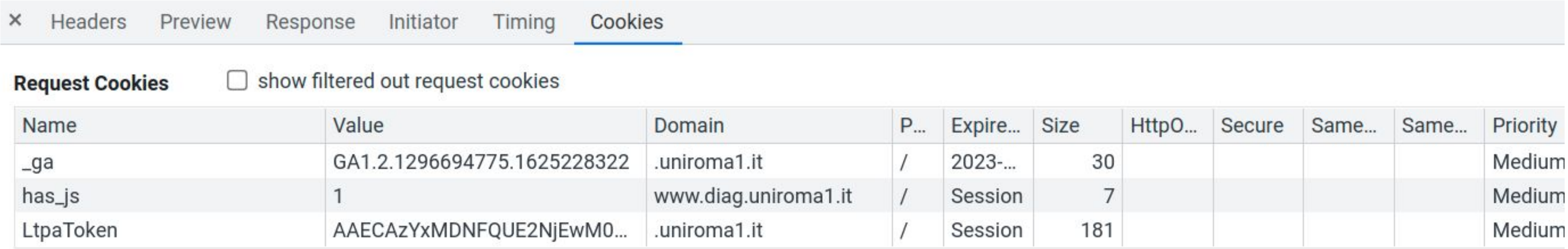
Upgrade-Insecure-Requests: 1

User-Agent: Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.107 Safari/537.36



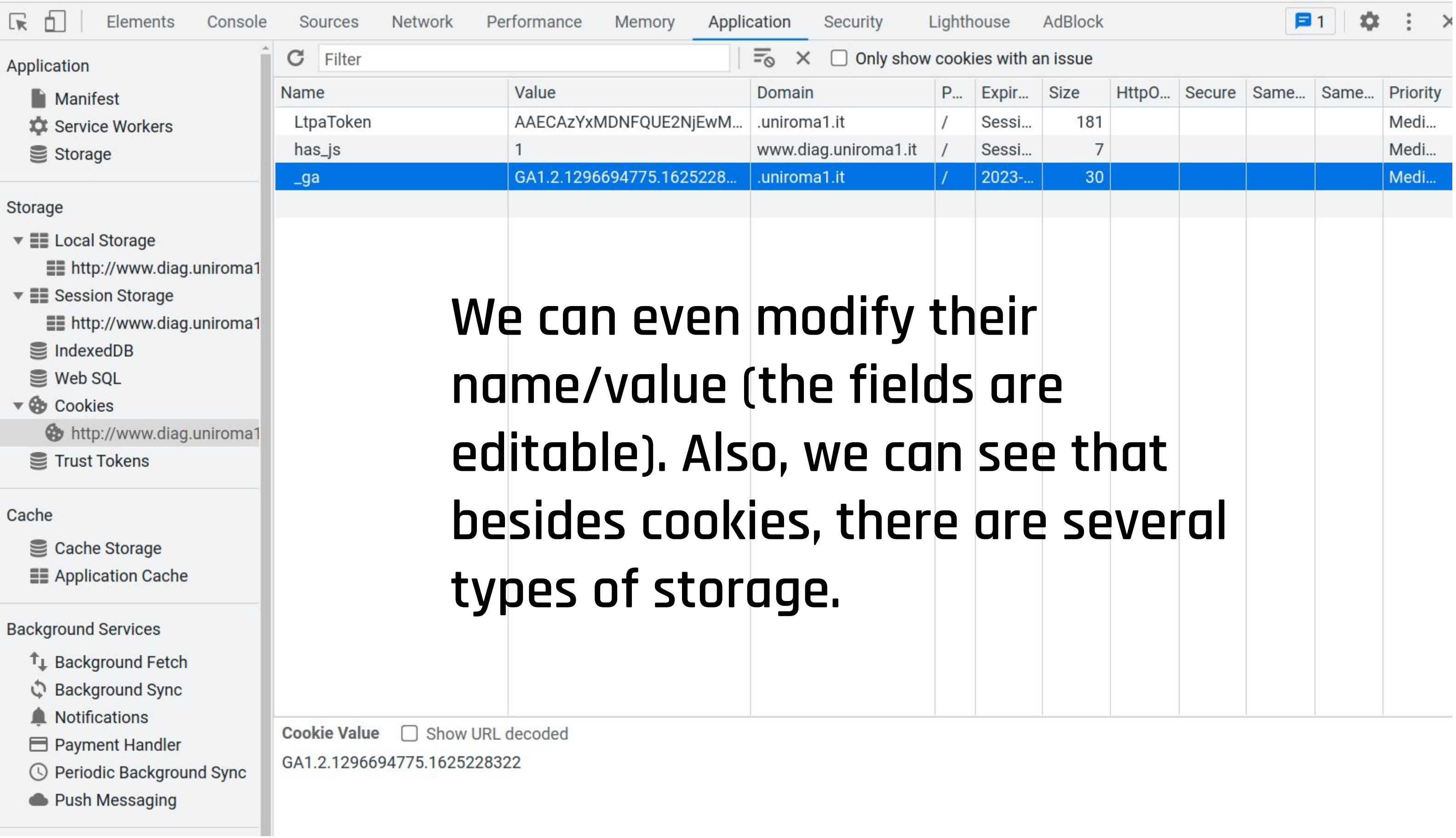
```
Cache-Control: no-cache, must-revalidate
Connection: Keep-Alive
Content-Encoding: gzip
Content-Language: it
Content-Length: 7020
Content-Type: text/html; charset=utf-8
Date: Wed, 04 Aug 2021 14:02:13 GMT
Expires: Sun, 19 Nov 1978 05:00:00 GMT
Keep-Alive: timeout=5, max=100
Link: </dipartimento>; rel="canonical",</node/5559>; rel="shortlink"
Server: Apache/2.4.18 (Ubuntu)
Vary: Accept-Encoding
X-Content-Type-Options: nosniff
X-Frame-Options: SAMEORIGIN
X-Generator: Drupal 7 (http://drupal.org)
```

We can see which cookies are used by the page



Request Cookies ☐ show filtered out request cookies

Name	Value	Domain	P...	Expire...	Size	HttpO...	Secure	Same...	Same...	Priority
_ga	GA1.2.1296694775.1625228322	.uniroma1.it	/	2023-...	30					Medium
has_js	1	www.diag.uniroma1.it	/	Session	7					Medium
LtpaToken	AAECAzYxMDNFQUE2NjEwM0...	.uniroma1.it	/	Session	181					Medium

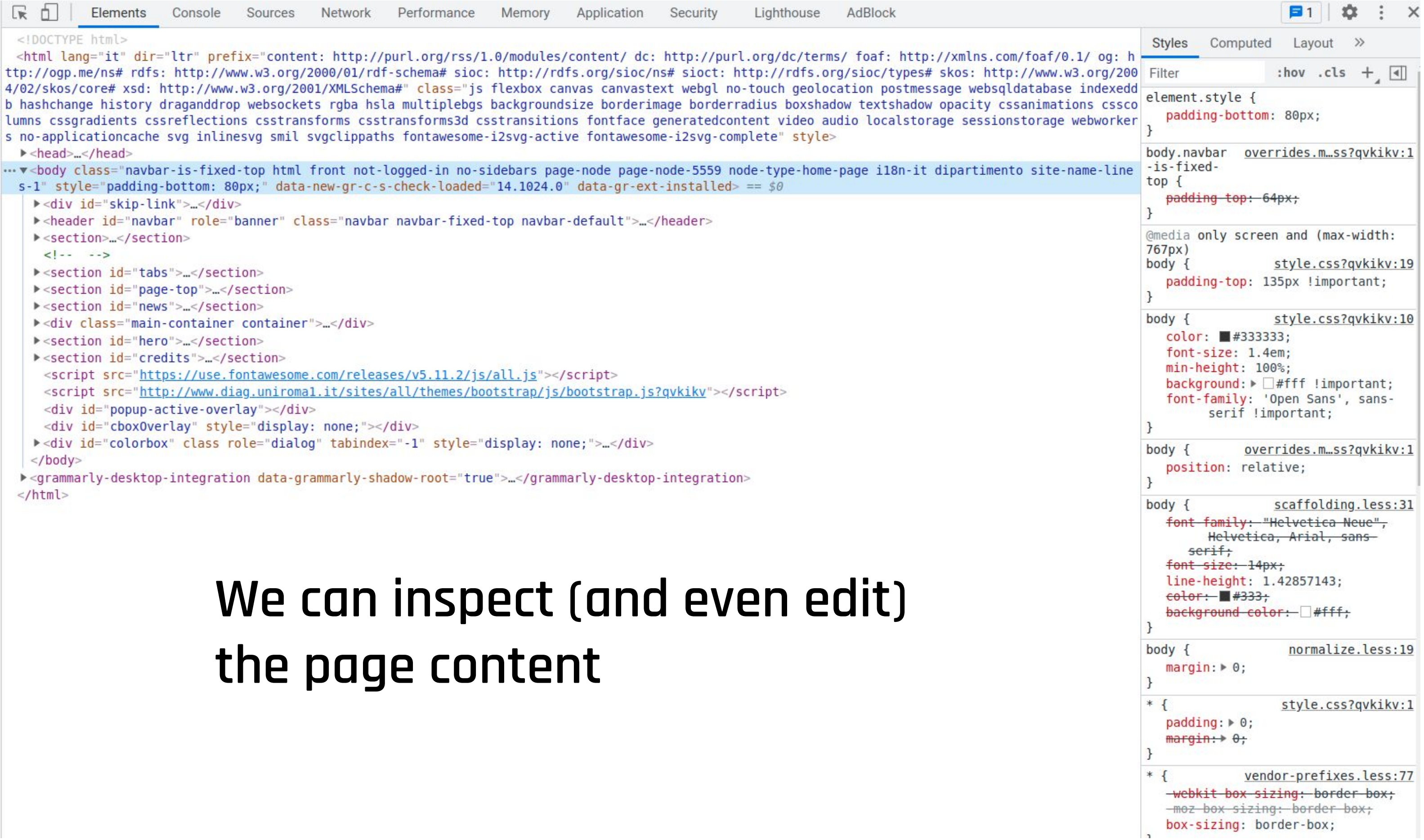


Name	Value	Domain	P...	Expir...	Size	HttpO...	Secure	Same...	Same...	Priority
LtpaToken	AAECAzYxMDNFQUE2NjEwM...	.uniroma1.it	/	Sessi...	181					Medi...
has_js	1	www.diag.uniroma1.it	/	Sessi...	7					Medi...
_ga	GA1.2.1296694775.1625228...	.uniroma1.it	/	2023-...	30					Medi...

Cookie Value □ Show URL decoded

GA1.2.1296694775.1625228322

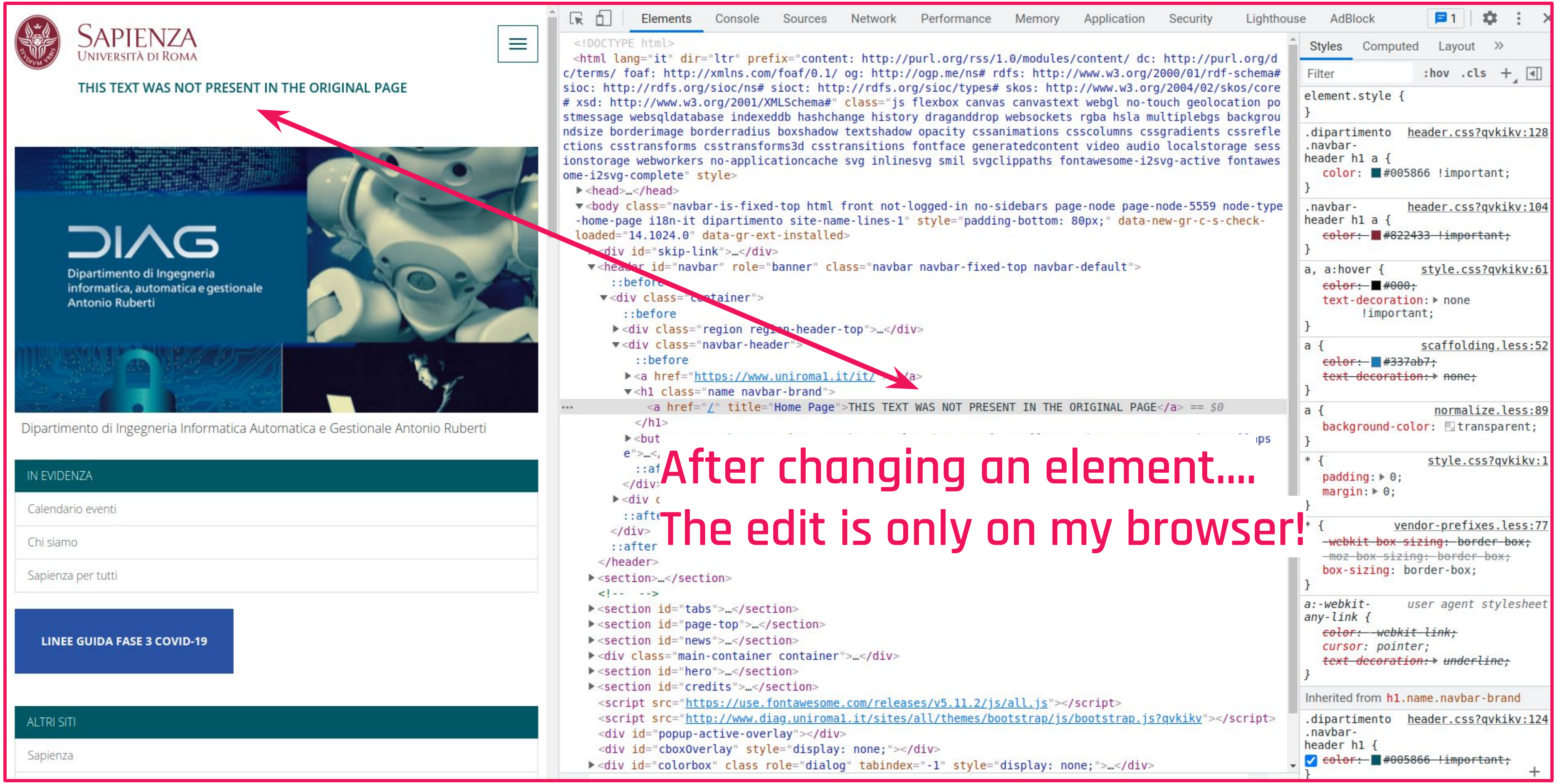
We can even modify their name/value (the fields are editable). Also, we can see that besides cookies, there are several types of storage.



```
<!DOCTYPE html>
<html lang="it" dir="ltr" prefix="content: http://purl.org/rss/1.0/modules/content/ dc: http://purl.org/dc/terms/ foaf: http://xmlns.com/foaf/0.1/ og: http://ogp.me/ns# rdfs: http://www.w3.org/2000/01/rdf-schema# sioc: http://rdfs.org/sioc/ns# sioct: http://rdfs.org/sioc/types# skos: http://www.w3.org/2004/02/skos/core# xsd: http://www.w3.org/2001/XMLSchema#" class="js flexbox canvas canvastext webgl no-touch geolocation postmessage websqldatabase indexeddb hashchange history draganddrop websockets rgba hsla multiplebgs backgroundsize borderimage borderradius boxshadow textshadow opacity cssanimations csscolumns cssgradients cssreflections csstransforms csstransforms3d csstransitions fontface generatedcontent video audio localstorage sessionstorage webworkers no-applicationcache svg inlinesvg smil svgclippaths fontawesome-i2svg-active fontawesome-i2svg-complete" style>
  ▶<head>…</head>
…▼<body class="navbar-is-fixed-top html front not-logged-in no-sidebars page-node page-node-5559 node-type-home-page i18n-it dipartimento site-name-line-1" style="padding-bottom: 80px;" data-new-gr-c-s-check-loaded="14.1024.0" data-gr-ext-installed> == $0
    ▶<div id="skip-link">…</div>
    ▶<header id="navbar" role="banner" class="navbar navbar-fixed-top navbar-default">…</header>
    ▶<section>…</section>
     <!--  -->
    ▶<section id="tabs">…</section>
    ▶<section id="page-top">…</section>
    ▶<section id="news">…</section>
    ▶<div class="main-container container">…</div>
    ▶<section id="hero">…</section>
    ▶<section id="credits">…</section>
     <script src="https://use.fontawesome.com/releases/v5.11.2/js/all.js"></script>
     <script src="http://www.diag.uniroma1.it/sites/all/themes/bootstrap/js/bootstrap.js?qvkikv"></script>
     <div id="popup-active-overlay"></div>
     <div id="cboxOverlay" style="display: none;"></div>
    ▶<div id="colorbox" class role="dialog" tabindex="-1" style="display: none;">…</div>
   </body>
  ▶<grammarly-desktop-integration data-grammarly-shadow-root="true">…</grammarly-desktop-integration>
</html>
```

```
element.style {
  padding-bottom: 80px;
}
body.navbar-is-fixed-top {          overrides.m…ss?qvkikv:1
  ~~padding-top: 64px;~~
}
@media only screen and (max-width: 767px)
body {                              style.css?qvkikv:19
  padding-top: 135px !important;
}
body {                              style.css?qvkikv:10
  color: ■#333333;
  font-size: 1.4em;
  min-height: 100%;
  background: ▶ □#fff !important;
  font-family: 'Open Sans', sans-serif !important;
}
body {                              overrides.m…ss?qvkikv:1
  position: relative;
}
body {                              scaffolding.less:31
  ~~font-family: "Helvetica Neue", Helvetica, Arial, sans-serif;~~
  ~~font-size: 14px;~~
  line-height: 1.42857143;
  ~~color: ■#333;~~
  ~~background-color: □#fff;~~
}
body {                              normalize.less:19
  margin: ▶ 0;
}
* {                                 style.css?qvkikv:1
  padding: ▶ 0;
  ~~margin: ▶ 0;~~
}
* {                                 vendor-prefixes.less:77
  ~~-webkit-box-sizing: border-box;~~
  ~~-moz-box-sizing: border-box;~~
  box-sizing: border-box;
```

We can inspect (and even edit) the page content



THIS TEXT WAS NOT PRESENT IN THE ORIGINAL PAGE

Dipartimento di Ingegneria Informatica Automatica e Gestionale Antonio Ruberti

IN EVIDENZA

Calendario eventi

Chi siamo

Sapienza per tutti

LINEE GUIDA FASE 3 COVID-19

ALTRI SITI

Sapienza

```
<!DOCTYPE html>
<html lang="it" dir="ltr" prefix="content: http://purl.org/rss/1.0/modules/content/ dc: http://purl.org/dc/terms/ foaf: http://xmlns.com/foaf/0.1/ og: http://ogp.me/ns# rdfs: http://www.w3.org/2000/01/rdf-schema# sioc: http://rdfs.org/sioc/ns# sioct: http://rdfs.org/sioc/types# skos: http://www.w3.org/2004/02/skos/core# xsd: http://www.w3.org/2001/XMLSchema#" class="js flexbox canvas canvastext webgl no-touch geolocation postmessage websqldatabase indexeddb hashchange history draganddrop websockets rgba hsla multiplebgs backgroundsize borderimage borderradius boxshadow textshadow opacity cssanimations csscolumns cssgradients cssreflections csstransforms csstransforms3d csstransitions fontface generatedcontent video audio localstorage sessionstorage webworkers no-applicationcache svg inlinesvg smil svgclippaths fontawesome-i2svg-active fontawesome-i2svg-complete" style>
 ▶<head>…</head>
 ▼<body class="navbar-is-fixed-top html front not-logged-in no-sidebars page-node page-node-5559 node-type-home-page i18n-it dipartimento site-name-lines-1" style="padding-bottom: 80px;" data-new-gr-c-s-check-loaded="14.1024.0" data-gr-ext-installed>
   ▶<div id="skip-link">…</div>
   ▼<header id="navbar" role="banner" class="navbar navbar-fixed-top navbar-default">
      ::before
     ▼<div class="container">
        ::before
       ▶<div class="region region-header-top">…</div>
       ▼<div class="navbar-header">
          ::before
         ▶<a href="https://www.uniroma1.it/it/">…</a>
         ▼<h1 class="name navbar-brand">
            <a href="/" title="Home Page">THIS TEXT WAS NOT PRESENT IN THE ORIGINAL PAGE</a> == $0
          </h1>
         ▶<but                                                           ps
          e">…</
          ::af
        </div>
       ▶<div c
        ::afte
      </div>
      ::after
    </header>
   ▶<section>…</section>
    <!--  -->
   ▶<section id="tabs">…</section>
   ▶<section id="page-top">…</section>
   ▶<section id="news">…</section>
   ▶<div class="main-container container">…</div>
   ▶<section id="hero">…</section>
   ▶<section id="credits">…</section>
    <script src="https://use.fontawesome.com/releases/v5.11.2/js/all.js"></script>
    <script src="http://www.diag.uniroma1.it/sites/all/themes/bootstrap/js/bootstrap.js?qvkikv"></script>
    <div id="popup-active-overlay"></div>
    <div id="cboxOverlay" style="display: none;"></div>
   ▶<div id="colorbox" class role="dialog" tabindex="-1" style="display: none;">…</div>
```

```
element.style {
}
.dipartimento          header.css?qvkikv:128
.navbar-
header h1 a {
  color: ■#005866 !important;
}
.navbar-               header.css?qvkikv:104
header h1 a {
  ~~color: ■#822433 !important;~~
}
a, a:hover {           style.css?qvkikv:61
  ~~color: ■#000;~~
  text-decoration: ▶ none
      !important;
}
a {                    scaffolding.less:52
  ~~color: ■#337ab7;~~
  ~~text-decoration: ▶ none;~~
}
a {                    normalize.less:89
  background-color: □transparent;
}
* {                    style.css?qvkikv:1
  padding: ▶ 0;
  margin: ▶ 0;
}
* {                    vendor-prefixes.less:77
  ~~-webkit-box-sizing: border-box;~~
  ~~-moz-box-sizing: border-box;~~
  box-sizing: border-box;
}
a:-webkit-             user agent stylesheet
any-link {
  ~~color: -webkit-link;~~
  cursor: pointer;
  ~~text-decoration: ▶ underline;~~
}
Inherited from h1.name.navbar-brand
.dipartimento          header.css?qvkikv:124
.navbar-
header h1 {
☑ ~~color: ■#005866 !important;~~
}
```

After changing an element....

The edit is only on my browser!

The Languages of the Web: Client-Side

- HTML
  - Defines the structure of the webpage

```
<html>
  <body>
    <p>hello!</p>
  </body>
</html>
```

- CSS
  - Defines the styling of the page

```
p {
  color: red;
}
```

- JavaScript:
  - Allows to add dynamic interactive effects to the webpage (e.g., react to user interactions)

```
let d = window.document;
let p = d.getElementsByTagName('p')[0];
p.addEventListener('click', function () {
  this.style.color = 'blue';
});
```

The Languages of the Web: Server-Side

- Virtually every programming language can be used on the server-side (even C!)

- Most common server-side languages in 2020:
    - **Python**, NodeJS (JavaScript), Java, C#, **PHP**

- The server-side language is used to implement your web application:
    - Session management of users
    - Interaction with the database
    - Generation of the response pages
    - …

Quick and dirty HTTP server

A quick but **unsafe** way of spawning a HTTP server is:

```
> python3 -m http.server 8000

Serving HTTP on 0.0.0.0 port 8000 (http://0.0.0.0:8000/) ...
```

NOTE: the current working directory is the root for the web server

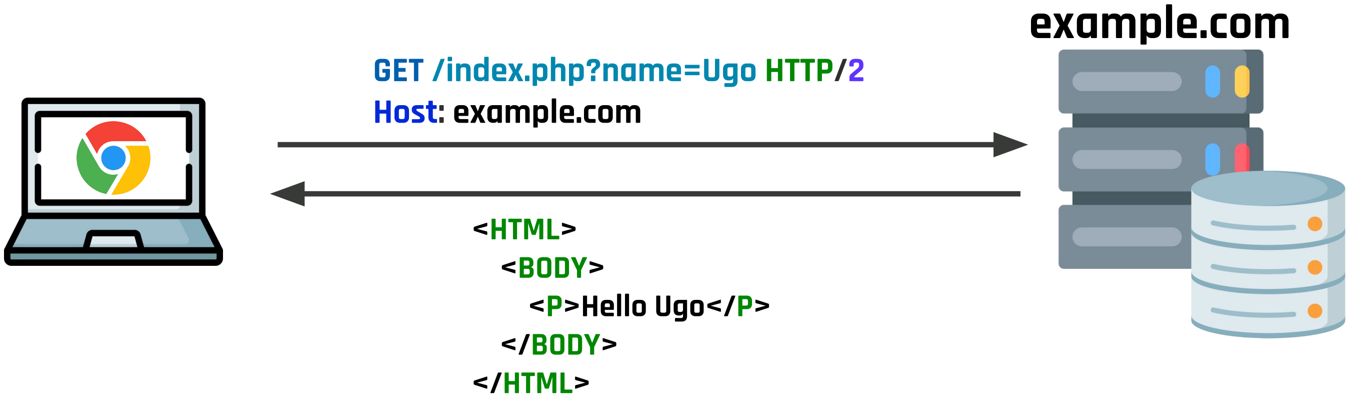
PHP: Hypertext Preprocessor - Basics

- We will use PHP in some of the examples
- It is a server-side scripting language with C-like syntax
- HTML and PHP code can be intermingled in the same file
- Variable names start with **$**
- Command **echo** can be used to print the value of an expression
- The operator **.** denotes string concatenation
- Important global associative arrays (i.e., dictionaries):
  - **$_GET**: parameters provided via the URL query string
  - **$_POST**: parameters provided in the body of a request
  - **$_SESSION**: parameters stored in a PHP session (preserved across multiple requests)

PHP: Hypertext Preprocessor - Example

```
<HTML>
  <BODY>
    <P><?php echo "Hello " . $_GET["name"]; ?></P>
  </BODY>
</HTML>
```

index.php



example.com

```
GET /index.php?name=Ugo HTTP/2
Host: example.com
```

```
<HTML>
  <BODY>
    <P>Hello Ugo</P>
  </BODY>
</HTML>
```

Quick and dirty HTTP+PHP server

A quick but **unsafe** way of spawning a HTTP/PHP server is:

```
> php -S 0.0.0.0:8000
```

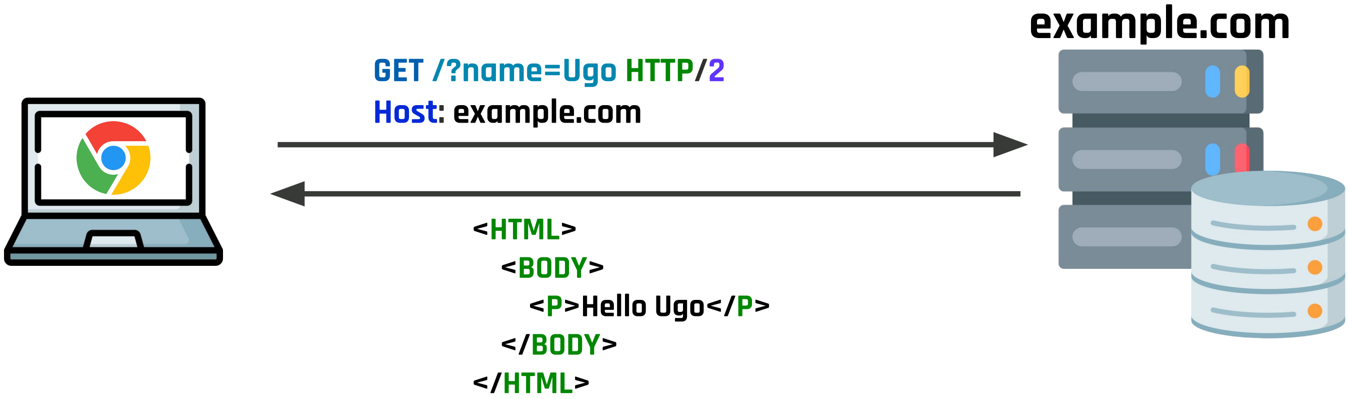
```
[Mon Oct 25 18:42:06 2021] PHP 7.4.3 Development Server (http://0.0.0.0:8000) started
[Mon Oct 25 18:42:28 2021] 127.0.0.1:37502 Accepted
[Mon Oct 25 18:42:28 2021] 127.0.0.1:37502 [200]: GET /
[Mon Oct 25 18:42:28 2021] PHP Notice:  Undefined index: name in index.php on line 3
[Mon Oct 25 18:42:28 2021] 127.0.0.1:37502 Closing
[Mon Oct 25 18:42:37 2021] 127.0.0.1:37506 Accepted
[Mon Oct 25 18:42:37 2021] 127.0.0.1:37504 [200]: GET /?name=ugo
[Mon Oct 25 18:42:37 2021] 127.0.0.1:37504 Closing
```

NOTE: the current working directory is the root for the web server

Quick and dirty HTTP/Python server

```
from flask import Flask, request

app = Flask(__name__)

@app.route("/")
def hello_world():
  return "<html>\n<body>\n<p>Hello %s</p></body></html>"
                    % request.args.get('name')
```

app.py



example.com

```
GET /?name=Ugo HTTP/2
Host: example.com
```

```
<HTML>
  <BODY>
    <P>Hello Ugo</P>
  </BODY>
</HTML>
```

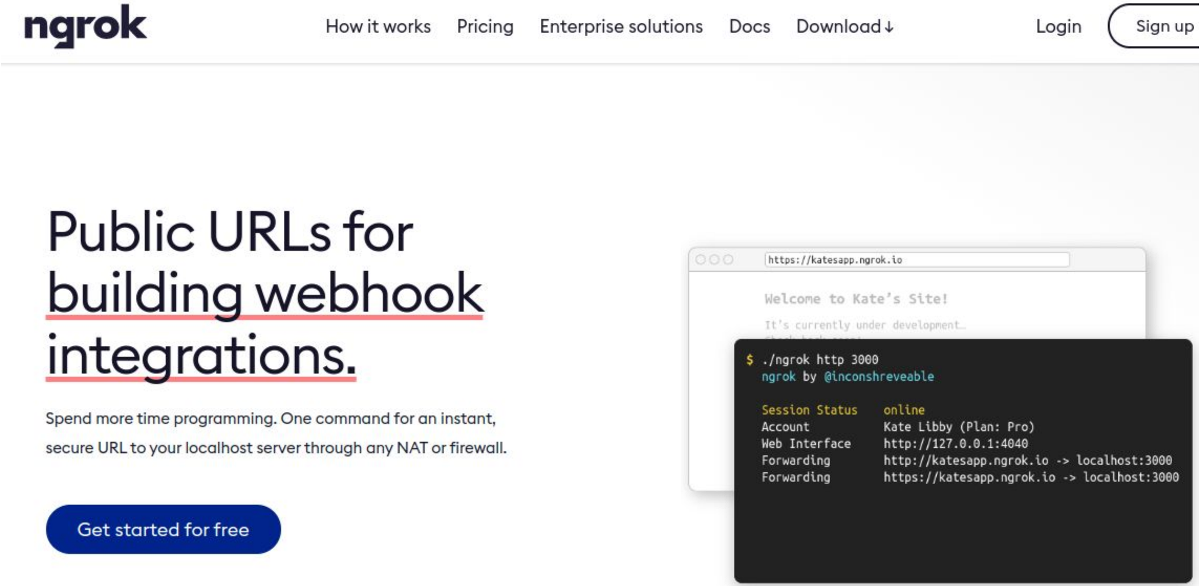
Quick and dirty HTTP/Python server (2)

```
> pip3 install flask
> python3 -m flask run
 * Environment: production
   WARNING: This is a development server. Do not use it in a production deployment.
   Use a production WSGI server instead.
 * Debug mode: off
 * Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
127.0.0.1 - - [25/Oct/2021 18:57:17] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [25/Oct/2021 18:57:17] "GET /favicon.ico HTTP/1.1" 404 -
127.0.0.1 - - [25/Oct/2021 18:57:29] "GET /?name=ugo HTTP/1.1" 200 -
```

How to make our server reachable from the internet?

Assuming that we are just talking about development/CTF deployment… we can use **ngrok** to make our server reachable (possibly even with HTTPS). This will work even without a firewall (port forwarding) and without a (dynamic) domain.

https://ngrok.com/



Public URLs for building webhook integrations.

Spend more time programming. One command for an instant, secure URL to your localhost server through any NAT or firewall.

https://katesapp.ngrok.io

Welcome to Kate's Site!

It's currently under development…

```
$ ./ngrok http 3000
  ngrok by @inconshreveable

  Session Status    online
  Account           Kate Libby (Plan: Pro)
  Web Interface     http://127.0.0.1:4040
  Forwarding        http://katesapp.ngrok.io -> localhost:3000
  Forwarding        https://katesapp.ngrok.io -> localhost:3000
```

ngrok

1. Spawn your local HTTP server on port X

2. Download and install ngrok (available as a snap package!)

3. Register an account on ngrok.com and get the authtoken

4. Configure the authtoken:
   > ngrok authtoken <auth_token>

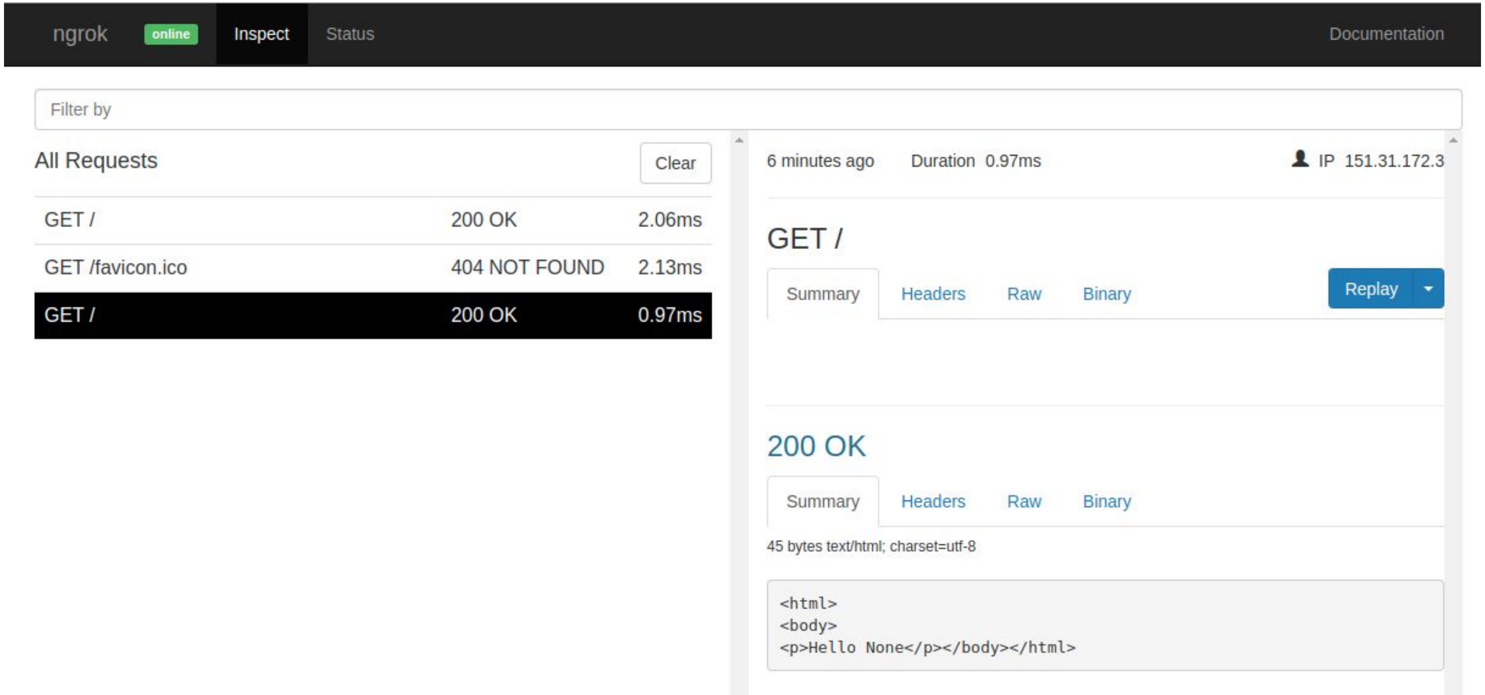
ngrok (2)

5. Run ngrok for http X: > **ngrok http X**

```
Session Status                online
Account                       ercoppa (Plan: Free)
Version                       2.3.40
Region                        United States (us)
Web Interface                 http://127.0.0.1:4040
Forwarding                    http://2781-151-31-172-3.ngrok.io -> http://localhost:5000
Forwarding                    https://2781-151-31-172-3.ngrok.io -> http://localhost:5000

Connections                 ttl     opn     rt1     rt5     p50     p90
                            3       0       0.04    0.01    0.00    0.00
HTTP Requests
-------------
GET /                     200 OK
GET /favicon.ico          404 NOT FOUND
GET /                     200 OK
```

ngrok (3)

6. Get statistics from the ngrok web interface



All Requests

GET /	200 OK	2.06ms
GET /favicon.ico	404 NOT FOUND	2.13ms
GET /	200 OK	0.97ms

6 minutes ago Duration 0.97ms IP 151.31.172.3

GET /

200 OK

45 bytes text/html; charset=utf-8

```
<html>
<body>
<p>Hello None</p></body></html>
```