

Имена: Христо Спасов

ФН: 62278

Имена: Таня Желева

ФН: 62288

Начална година: 2018

Програма: бакалавър, (СИ)

Kypc: 3

Тема: Уеб архив (wayback machine)

Дата: 2021-06-01

Предмет: w16pri SI final

имейл: hristo.b.spasov@gmail.com

имейл: tanya.n.zheleva@gmail.com

преподавател: доц. д-р Милен Петров

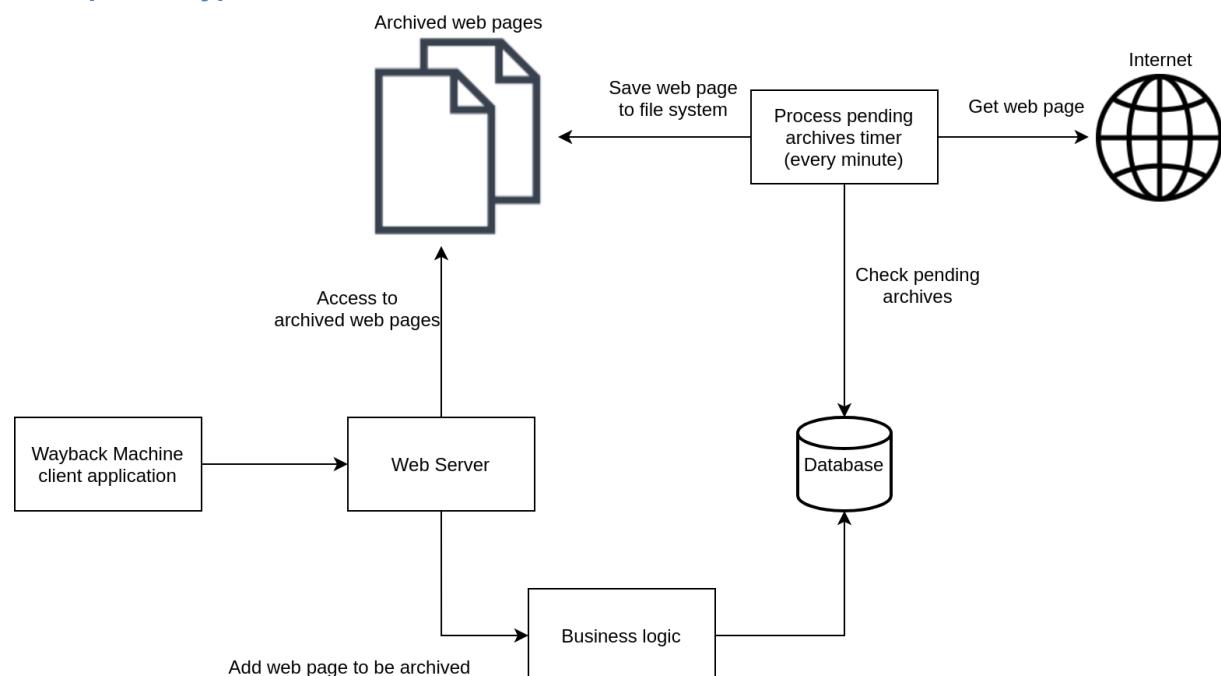
ТЕМА: Уеб архив (wayback machine)

1. Условие

Да се направи система, подобна на „<https://archive.org/web/>“, която по зададен URL прави архив на дадена страница. След архивиране може да се влезе и да се навигира по страници и да се търси архив за страница. Ако има - всяко от архивираните се показва в календар или списък.

2. Въведение

2.1. Архитектура



3. Теория

4. Използвани технологии

- Системата е проектирана да работи на операционна система **Ubuntu 20.04**.
 - За обработка на заявки се използва уеб сървър **Apache2**.
 - Периодичното изпълнение на скриптове се осъществява от **systemd timer**.

- Базата от данни е MySQL.
- Сървърната логика е изградена с езика PHP
- За изграждане на фронт енд са използвани технологиите HTML, CSS и JavaScript

5. Инсталация и настройки

Инструкциите за инсталация са предвидени за операционна система Ubuntu 20.04.

1. Свалете изходния код на системата.

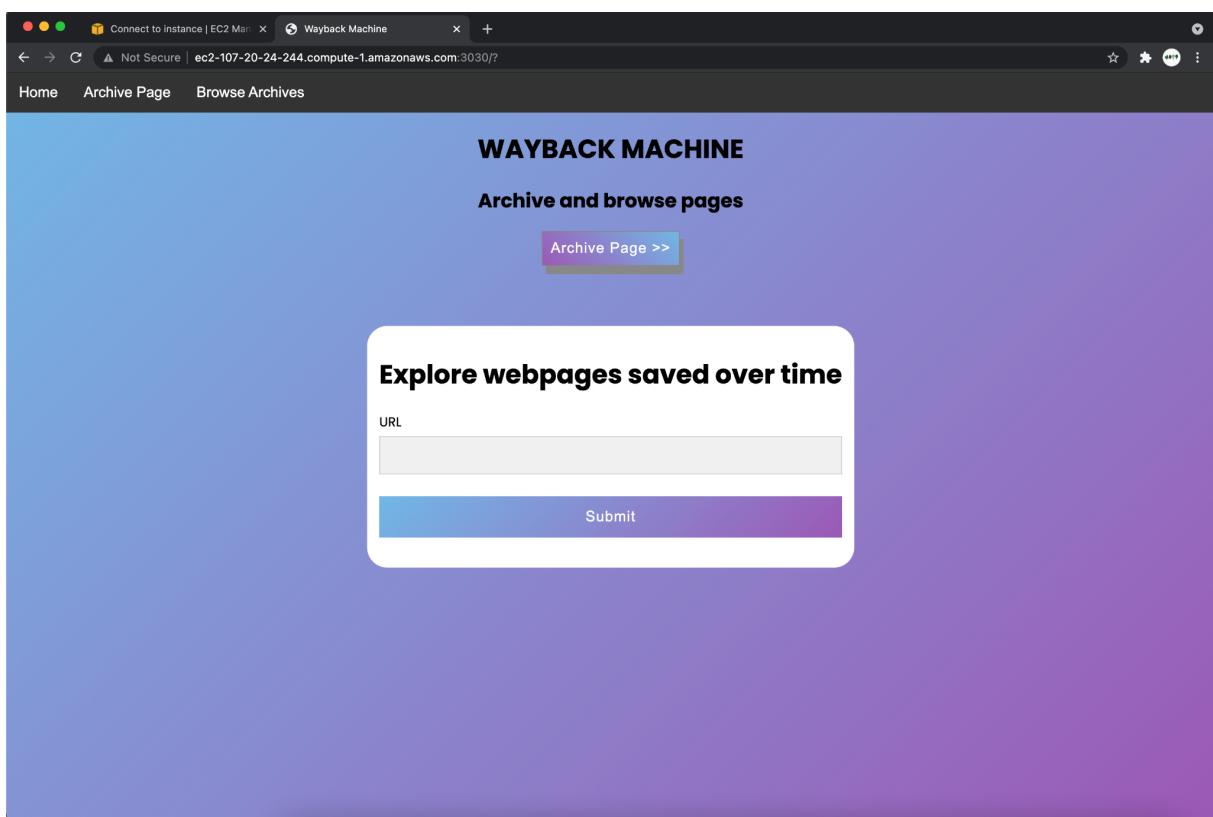
2. Отворете директория wbmachine.

```
$ cd ./wbmachine
```

3. Изпълнете със супер потребител инсталационния скрипт:

```
$ sudo ./setup/install.sh
```

4. Отворете localhost:3030. Ако видите началната страница на Wayback Machine, значи системата е успешно инсталирана



5.1. Деинсталация на системата

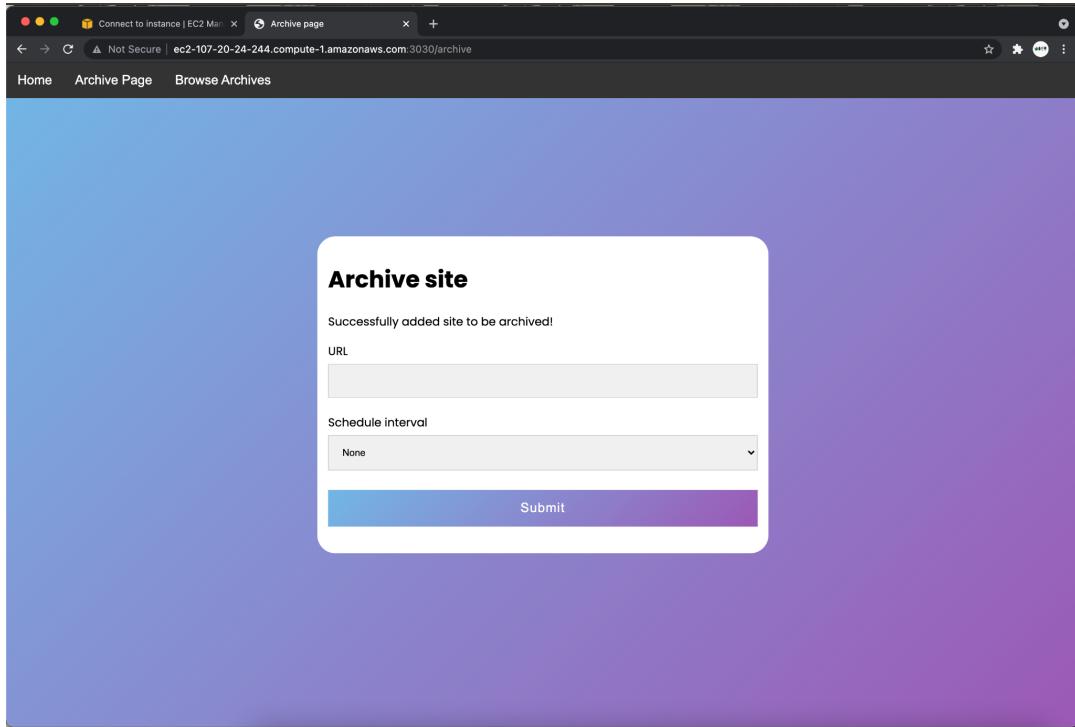
Можете да деинсталирате системата и да изтриете всички данни от нея със скрипта за деинсталация:

```
$ sudo ./setup/uninstall.sh
```

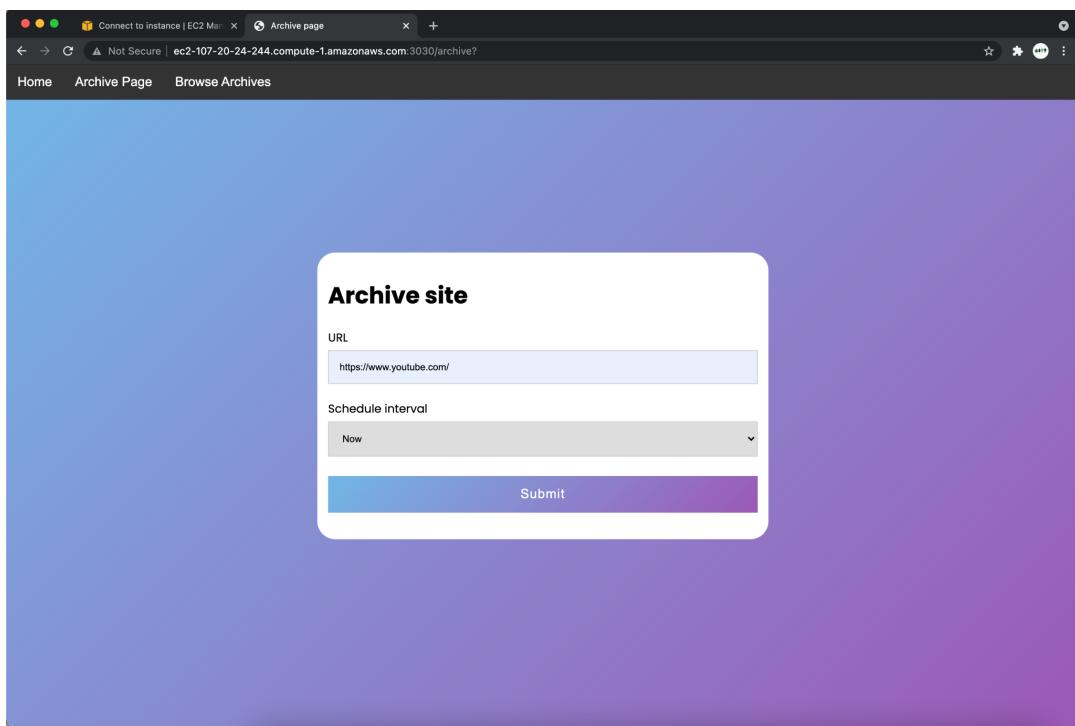
6. Кратко ръководство на потребителя

6.1. Архивиране на страница

1. Отворете страницата /archive:



2. В полето URL въведете адреса на страницата, която желаете да архивирате.
За Schedule Interval изберете Now.

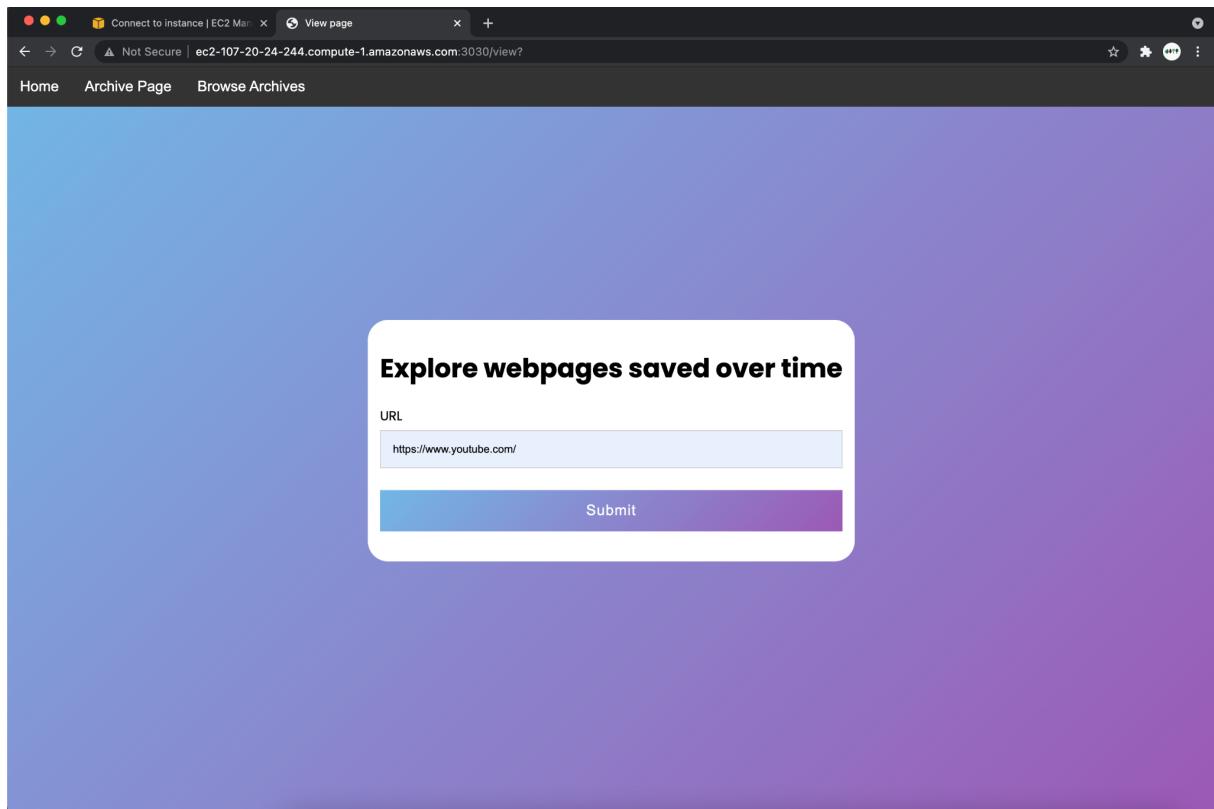


Натиснете бутона Submit.

Системата е планирала зададеното архивиране и след няколко минути ще можете да отворите архивираната страница.

6.2. Отваряне на архивирана страница

1. Отворете страницата /view. В полето URL въведете адреса на архивираната страница, която желаете да отворите.



2. Ако в системата тази страница е добавена, ще ви се отвори списък с моменти на архивиране. Изберете момент от миналото, за да видите страницата в тогавашния ѝ вид.

Explore webpages saved over time

URL
https://www.youtube.com/

Submit

2021-06-12 16:00:21
2021-06-12 16:04:03
2021-06-12 16:20:42
2021-06-14 16:46:47

Displayed webpage: https://www.youtube.com/

Current timestamp: 2021-06-14 16:46:47

Select different archive 2021-06-14 16:46:47

YouTube

Search

SIGN IN

YouTubeTV
\$30 to save, 3 months to explore. Get \$10 off per month for 3 months to watch this summer.

CLAIM OFFER

New users only. Cancel anytime. Terms apply.

12-year-old jumps off overpass, kills driver

1:38 CBS News 4.5M views • 3 years ago

Make THIS Shot, Win \$300,000

15:04 MiBeast 21M views • 1 day ago

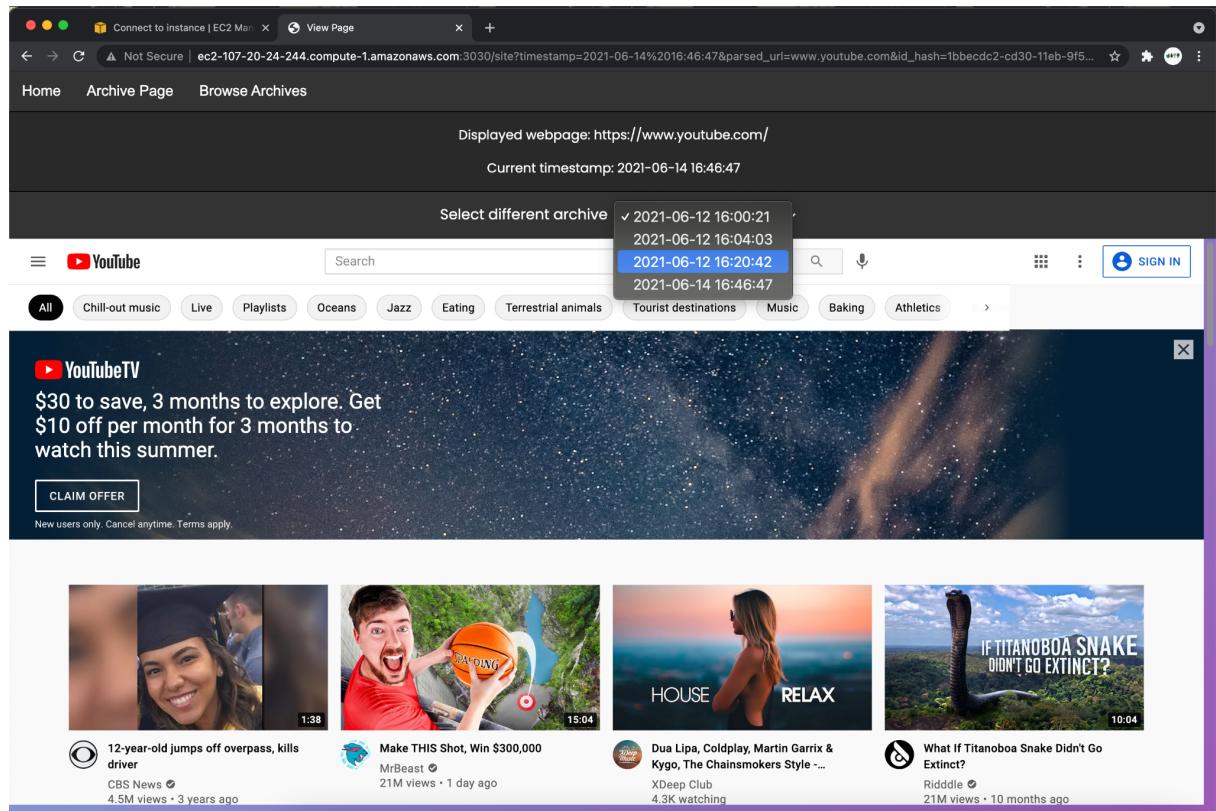
Dua Lipa, Coldplay, Martin Garrix & Kygo, The Chainsmokers Style -----

XDeep Club 4.3K watching

IF TITANOBAA SNAKE DIDN'T GO EXTINCT?

10:04 Riddlle 21M views • 10 months ago

3. След отваряне на архивирана страница, можете да изберете друг момент на архивиране чрез опцията "Select different archive"



7. Примерни данни

8. Описание на програмния код

8.1. Описание на структурата на проекта

- **doc** - диаграми и друга документация
- **setup** - директория с инсталационни неща
 - **db_create.sql** - създава базата
 - **db_drop.sql** - трябва базата
 - **db_fill.sql** - попълва базата с данни
 - **install.sh** - прави setup на системата
 - **uninstall.sh** - деинсталира системата
- **server** - директория с бизнес логика
 - **modules** - директория с различни модули, които са php файлове
 - **index.php** - входната точка за сървъра, тук са route-овете
- **views** - директория с front end неща
 - **components** - преизползвани UI компоненти
 - **pages** - цели страници
 - **public** - директория, която е публична за клиента
 - **js** - js файлове
 - **css** - css файлове
 - **layout.html** - UI layout-a
- **config** - конфигурационни файлове
- **scripts** - директория със скриптове

9. Приноси на студента, ограничения и възможности за бъдещо разширение

10. Какво научих

- Git Feature Branch Workflow
- php ini
- Дефиниране на константи в PHP
- require, include и use в PHP
- PHP функциите print_r и error_log, чрез които удобно може да се прави TRACE
- PHP функцията parse_url
- PHP функцията за проверка дали променлива е зададена: empty.
Разликата между empty, isset
- Как се взимат параметри при POST заявка: от \$_POST
- Как се взимат параметри при GET заявка: от \$_GET
- \$_SERVER - информация за заявката и за сървъра - хедъри, протокол, HTTP метод, порт и други
- Apache VirtualHost конфигурация
- PHP алтернативен синтаксис за контролни структури
- Употреба на wget - опциите за сваляне на всички ресурси от страницата и конвертиране на линковете за тях да са линкове към локално свалените ресурси

11. Използвани източници

[1] **Wget - The non-interactive network downloader.** [<https://linux.die.net/man/1/wget>]

[2] **systemd.timer — Timer unit configuration**

[<https://www.freedesktop.org/software/systemd/man/systemd.timer.html>]

[3] **PHP Manual** [<https://www.php.net/manual/en/>]

[4] **VirtualHost Examples, Apache HTTP Server Version 2.4**

[<https://httpd.apache.org/docs/2.4/vhosts/examples.html>]

Предал (подпис):

/фн, имена, спец., група/

Приел (подпис):

/доц. Милен Петров/