

3.13 Change detection. Monitoriza cambios en una web



Imagen obtenida de <https://changedetection.io/>

Esta herramienta sirve para...

que se nos notifique cuando una determinada web (o parte de una web) cambie. Es especialmente útil cuando queremos, por ejemplo, enterarnos de una nueva noticia que aparezca en la web del colegio de nuestros hijos/as o cuando estamos esperando una calificación de un proceso selectivo y no queremos estar accediendo continuamente a la página web a comprobarlo. Este servicio se encarga de monitorizar por nosotros/as la web que le indiquemos y avisarnos si hay cambios.

Web de proyecto y otros enlaces de interés

Web del proyecto: <https://changedetection.io/>

Repositorio del proyecto: <https://github.com/dgtlmoon/changedetection.io>

Despliegue

Como en ocasiones anteriores vamos a hacer con docker-compose para ello accedemos al terminal y escribimos

```
cd $HOME
mkdir change-detection
cd change-detection
nano docker-compose.yml
```

y dentro del fichero escribiremos el siguiente contenido

```
version: '3.2'
services:
  changedetection:
    image: ghcr.io/dgtlmoon/changedetection.io
    container_name: changedetection
    hostname: changedetection
    volumes:
      - changedetection-data:/datastore
    ports:
      - 5000:5000
    restart: unless-stopped
volumes:
  changedetection-data:
```

como en ocasiones anteriores, para guardar los cambios pulsaremos control + x y cuando nos pregunte aceptaremos. Una vez volvamos a estar en el terminal, escribiremos `docker compose up -d` para lanzar los servicios ubicados dentro del fichero docker-compose. Veremos algo similar a:

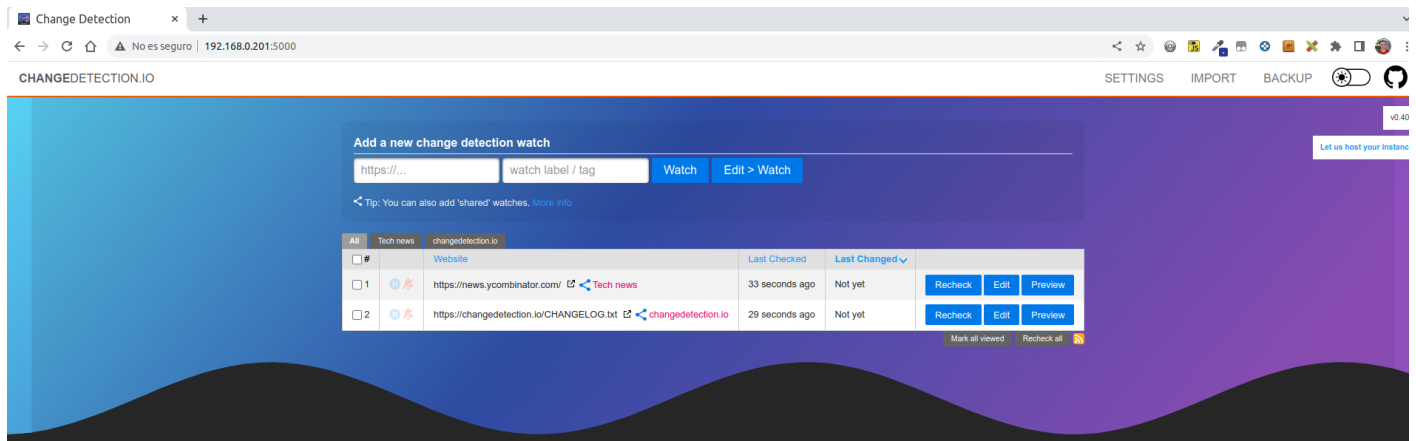
```

pi@mediacenter:~ $ cd $HOME
mkdir change-detection
cd change-detection
nano docker-compose.yml
pi@mediacenter:~/change-detection $ docker-compose up -d
Creating network "change-detection_default" with the default driver
Creating volume "change-detection_changedetection-data" with default driver
Pulling changedetection (ghcr.io/dgtlmoon/changedetection.io)...
latest: Pulling from dgtlmoon/changedetection.io
5731adb3a4ab: Pull complete
6e56a568739f: Pull complete
0f5a571ecf1e: Pull complete
259ed5ef5818: Pull complete
60b87b602a3f: Pull complete
c3491f3c8239: Pull complete
c6238eeafaef: Pull complete
4b9253ce8a82: Pull complete
2deadd51c0f0: Pull complete
1d5e7cf7ed81: Pull complete
84f431c780ea: Pull complete
4f4fb700ef54: Pull complete
Digest: sha256:5d1124db39e447f4f7a00301e531d4c7a230bc2d680e3c3c62f5150e2225a886
Status: Downloaded newer image for ghcr.io/dgtlmoon/changedetection.io:latest
Creating changedetection ... done

```

Elaboración propia

Si accedéis en vuestro navegador a la IP de la raspberry y al puerto que hemos establecido (5000). En mi caso sería `http://192.168.0.201:5000` deberíais ver algo como:



The screenshot shows a web browser window with the URL `192.168.0.201:5000`. The page title is "CHANGEDETECTION.IO". The interface includes a navigation bar with "SETTINGS", "IMPORT", and "BACKUP" options. A "v0.40" version indicator is visible in the top right corner. The main content area features a "Add a new change detection watch" form with a text input for the URL, a "watch label / tag" field, and "Watch" and "Edit > Watch" buttons. Below the form is a table of existing watches:

#	Website	Last Checked	Last Changed	Actions
1	https://news.ycombinator.com/ Tech news	33 seconds ago	Not yet	Recheck Edit Preview
2	https://changedetection.io/CHANGELOG.txt changedetection.io	29 seconds ago	Not yet	Recheck Edit Preview

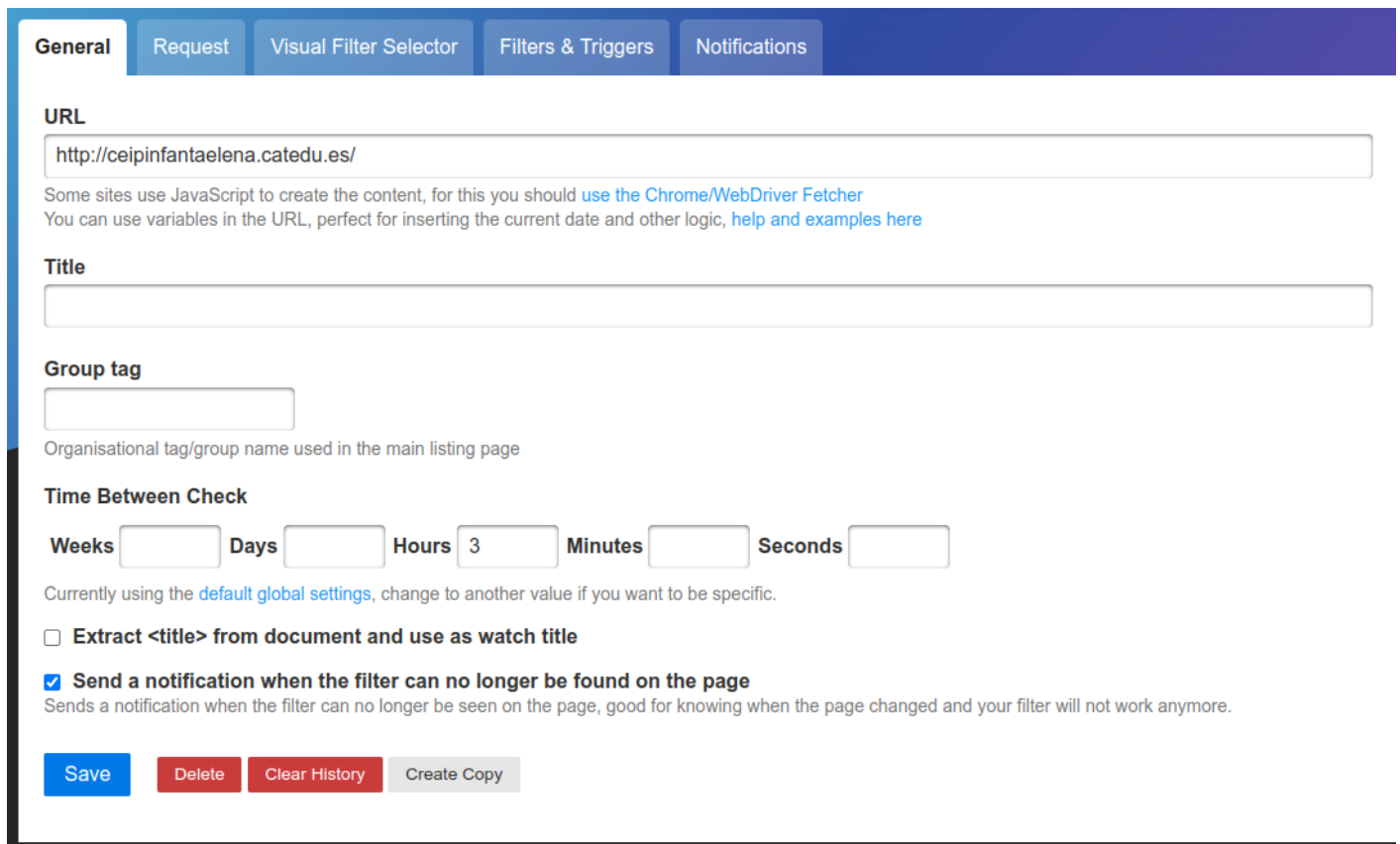
At the bottom of the table, there are "Mark all viewed" and "Recheck all" buttons.

Elaboración propia

Funcionamiento

El funcionamiento es tremendamente sencillo. Por ejemplo, para borrar las 2 páginas webs que vienen monitorizadas por defecto únicamente deberemos marcar los checkbox que aparecen a su izquierda y marcar `delete`.

Si queremos añadir una nueva página a monitorizar es suficiente con introducir una url en el recuadro superior y pulsar `Watch`. Si introducimos una url y pulsamos en `Edit > Watch` entonces tendremos acceso a una pantalla como la siguiente

The screenshot shows a web interface for configuring a filter. At the top, there is a navigation bar with tabs: 'General' (selected), 'Request', 'Visual Filter Selector', 'Filters & Triggers', and 'Notifications'. Below the tabs, the 'General' section contains several fields: a 'URL' field with the value 'http://ceipinfantaelena.catedu.es/', a 'Title' field, and a 'Group tag' field. Below these fields, there is a section for 'Time Between Check' with input boxes for 'Weeks', 'Days', 'Hours' (set to 3), 'Minutes', and 'Seconds'. There are also two checkboxes: 'Extract <title> from document and use as watch title' (unchecked) and 'Send a notification when the filter can no longer be found on the page' (checked). At the bottom, there are four buttons: 'Save' (blue), 'Delete' (red), 'Clear History' (red), and 'Create Copy' (grey).

Elaboración propia

en la cual podremos establecer la frecuencia de comprobaciones o las notificaciones a recibir de esta web en concreto.

Si lo preferimos también podemos acceder a la configuración global en el menú superior en la opción de Settings que nos dará acceso a una configuración general de nuestro servicio como vemos a continuación

General
Notifications
Fetching
Global Filters
API
CAPTCHA & Proxies

Time Between Check

Weeks Days Hours Minutes Seconds

Default time for all watches, when the watch does not have a specific time setting.

Random jitter seconds \pm check

Example - 3 seconds random jitter could trigger up to 3 seconds earlier or up to 3 seconds later

Number of times the filter can be missing before sending a notification

After this many consecutive times that the CSS/xPath filter is missing, send a notification
Set to 0 to disable

Password

Password protection for your changedetection.io application.

Allow access to view diff page when password is enabled
Allow access to view watch diff page when password is enabled (Good for sharing the diff page)

Base URL

Base URL used for the `{{ base_url }}` token in notifications and RSS links.
Default value is the ENV var 'BASE_URL' (Currently ""), [read more here](#).

Extract <title> from document and use as watch title
Note: This will automatically apply to all existing watches.

Treat empty pages as a change?
When a page contains HTML, but no renderable text appears (empty page), is this considered a change?

Elaboración propia

Si estableces un chequeo demasiado constante de una página web puede que la web en cuestión te bloquee el acceso a la misma al considerar que estás atacando la web. Sufrirías un *baneo*.

Revision #5

Created 2023-02-04 10:07:35 CET by Pablo Ruiz

Updated 2023-07-20 18:09:27 CEST by Pablo Ruiz