

TWITTER SCRAPER

MANUALE

Francesco Martinelli mat.744041

Umberto Nocerino mat.747751

Magistrale di Sicurezza Informatica

Sommario

INTRODUZIONE.....	2
AVVIO APPLICAZIONE.....	2
1. AUTENTICAZIONE TWITTER SCRAPER.....	2
2. Autenticazione twitter.....	3
ISTRUZIONI.....	4
FILE DI LOG.....	8
FILE DI HASH.....	9

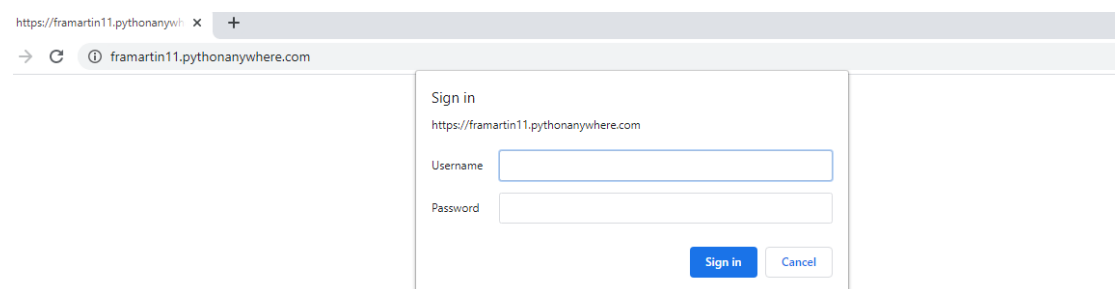
INTRODUZIONE

Il sistema realizzato ha lo scopo di effettuare lo scraping all'interno del Social Network Twitter. Dopo aver effettuato il doppio accesso, che può essere fatto dal browser, (è presente un *callback* che permette la reindirizzazione alla pagina di autenticazione di Twitter e successivamente di tornare sulla Web App) l'utente può richiedere lo scraping delle attività effettuate (tweet, retweet, commenti, etc.) oppure dei direct message. Il software contiene anche altre funzionalità aggiuntive, ovvero permette di mostrare le interazioni con un determinato utente (tramite l'username specifico) oppure mostra la relazione con quell'utente (follow, blocked, etc.). Il sistema produce un file di log che contiene le azioni effettuate dall'utente autenticato e produce i file csv crittografati tramite le funzioni di hash MD5 e SHA1.

AVVIO APPLICAZIONE

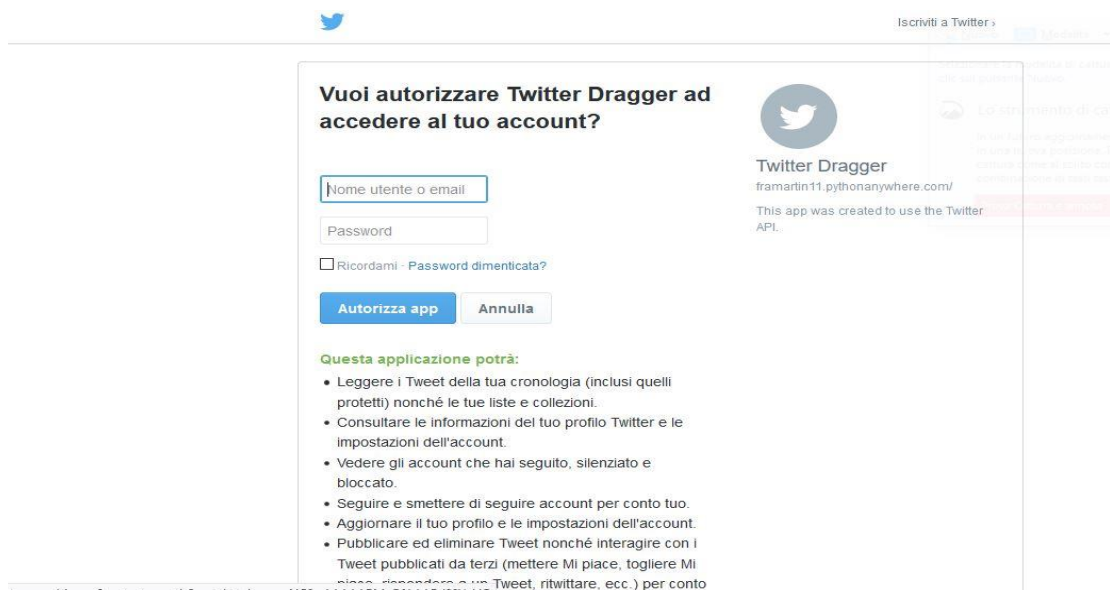
Una volta aperto il browser bisogna accedere al sito web della piattaforma:
“ framartin11.pythonanywhere.com ”

1. AUTENTICAZIONE TWITTER SCRAPER



In questa prima autenticazione bisognerà immettere username e password (forniti dall'amministratore) per avere accesso alla piattaforma. Il sistema quindi, prevede una doppia autenticazione per rendere il tutto più sicuro e soprattutto per poter essere ad esempio "affittato" ma molteplici utenti.

2. AUTENTICAZIONE TWITTER



The screenshot shows the Twitter authorization interface. At the top, there's a Twitter logo and links for 'Iscriviti a Twitter' and 'Modifica'. The main heading asks 'Vuoi autorizzare Twitter Dragger ad accedere al tuo account?'. Below this are input fields for 'Nome utente o email' and 'Password', followed by a checkbox for 'Ricordami - Password dimenticata?'. Two buttons, 'Autorizza app' and 'Annulla', are present. To the right, the app's profile is shown with the name 'Twitter Dragger', the URL 'framartin11.pythonanywhere.com/', and a note: 'This app was created to use the Twitter API.' Below the buttons, a section titled 'Questa applicazione potrà:' lists permissions: reading tweets, consulting profile info, seeing followed/blocked accounts, following/unfollowing, updating profile/settings, and publishing/deleting tweets.

Vuoi autorizzare Twitter Dragger ad accedere al tuo account?

Nome utente o email

Password

☐ Ricordami - [Password dimenticata?](#)

Autorizza app **Annulla**

Questa applicazione potrà:

- Leggere i Tweet della tua cronologia (inclusi quelli protetti) nonché le tue liste e collezioni.
- Consultare le informazioni del tuo profilo Twitter e le impostazioni dell'account.
- Vedere gli account che hai seguito, silenziato e bloccato.
- Seguire e smettere di seguire account per conto tuo.
- Aggiornare il tuo profilo e le impostazioni dell'account.
- Pubblicare ed eliminare Tweet nonché interagire con i Tweet pubblicati da terzi (mettere Mi piace, togliere Mi piace, rispondere a un Tweet, ritwittare, ecc.) per conto tuo.

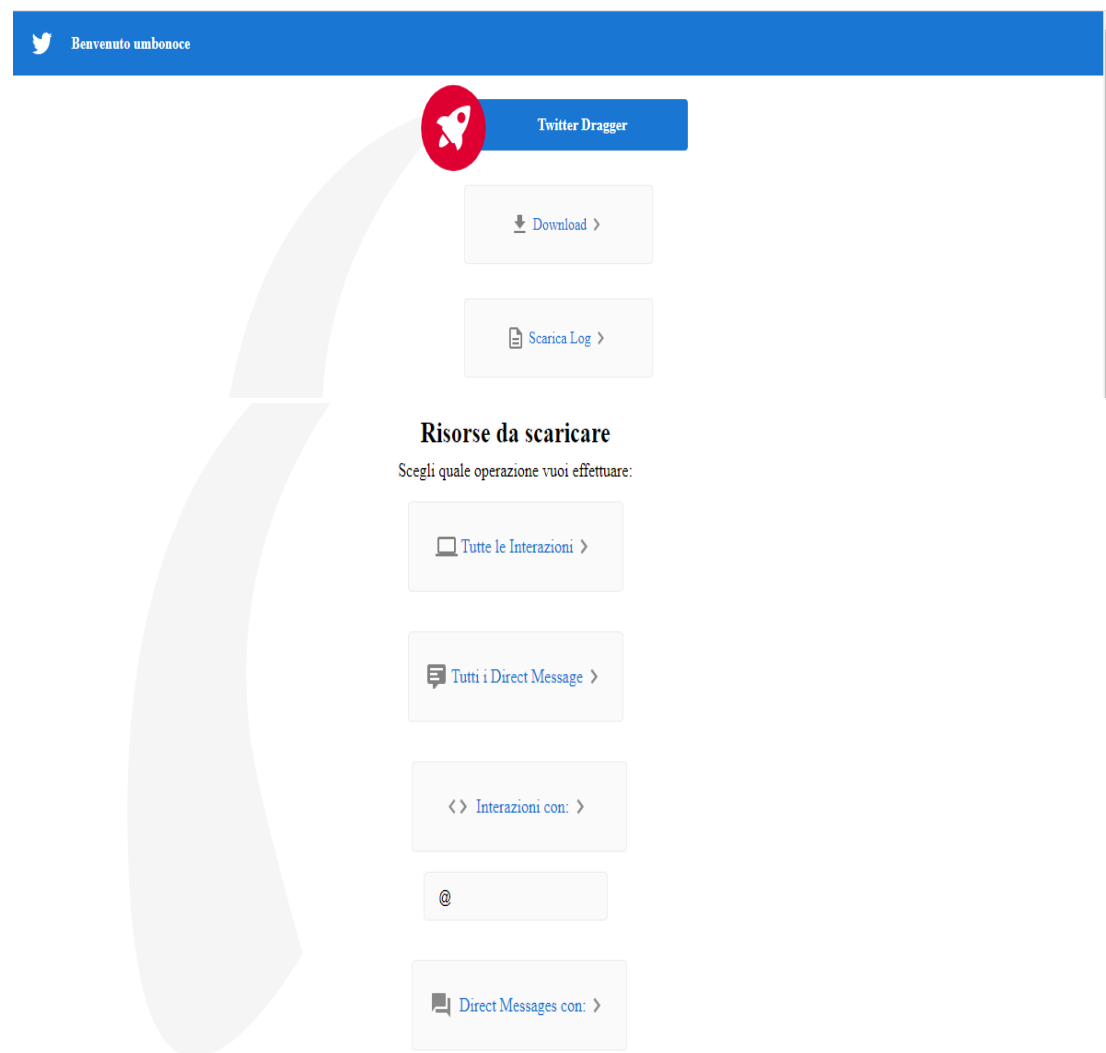
Twitter Dragger
framartin11.pythonanywhere.com/
This app was created to use the Twitter API.

Si richiede all'utente l'inserimento di e-mail e password per autenticarsi come utente Twitter, per avere accesso alle Api.

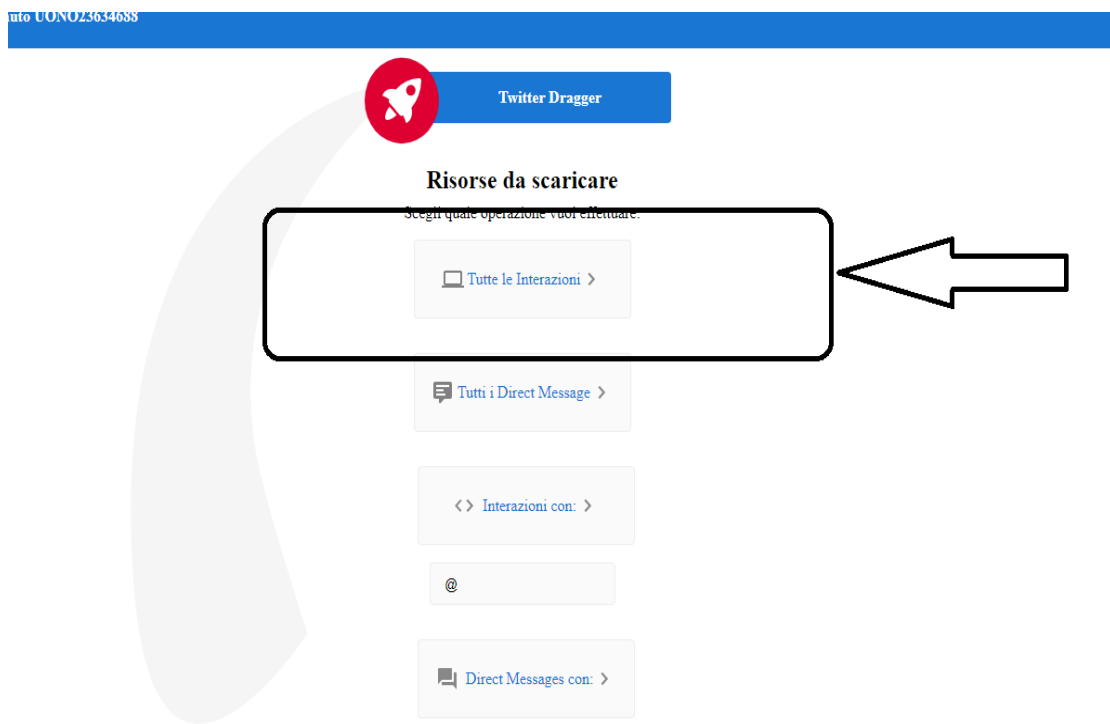
ISTRUZIONI

Tutti i file selezionati verranno accodati e successivamente scaricati in un file zip, tramite l'apposito pulsante “Download”.

Schermata principale piattaforma:



1. TUTTE LE INTERAZIONI



Nel primo caso d'uso sarà intuitivamente possibile digitare il pulsante in figura per accedere al file csv contenente le azioni dell'utente loggato. Il sistema genererà il file "allstatus.csv" contenente i campi:

- Id: identificativo univoco del tweet
- Data di creazione
- Testo: testo del tweet
- Lingua: lingua di scrittura del tweet autodeterminato da Twitter
- È un Retweet?
- Numero di retweet
- Numero di like
- Posizione: luogo di un post con geolocalizzazione
- Coordinate: coordinate di un post con geolocalizzazione

2. DIRECT MESSAGE

Nella seconda scelta è possibile ottenere tutti i direct message.



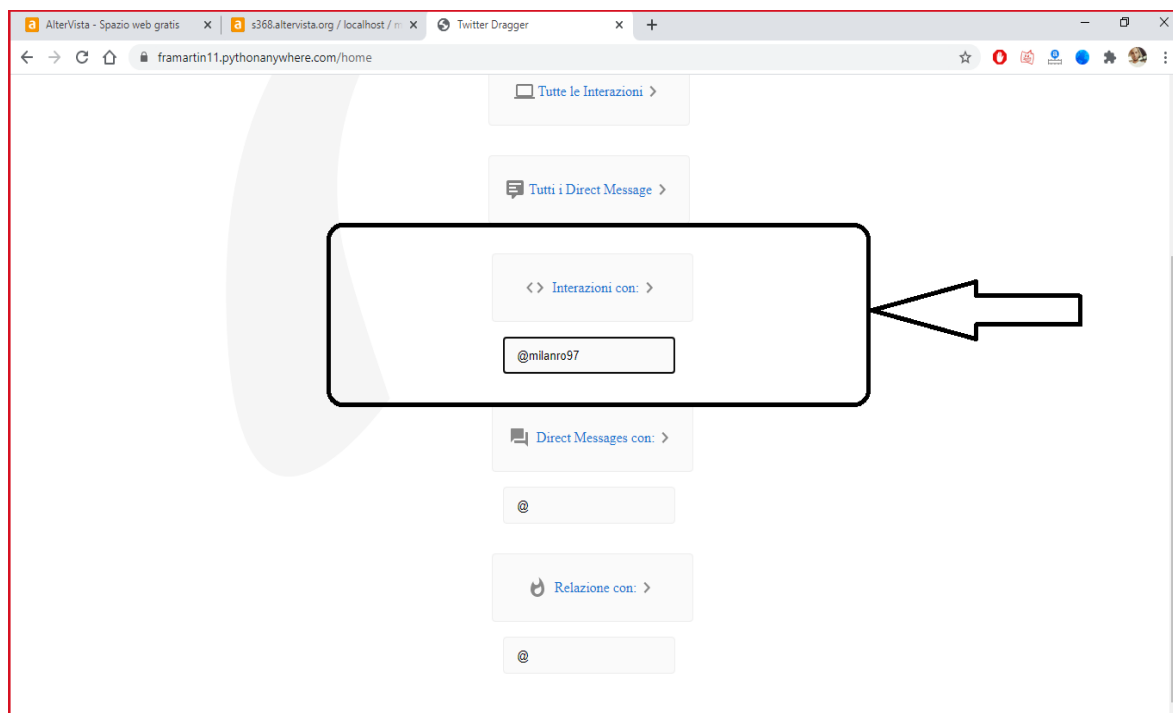
Inoltre, nelle opzioni successive il sistema consente anche di ricercare tramite una apposita checkbox una chat con uno specifico utente. Ovviamente il file risultante sarà un file csv che contiene le informazioni relative a quel determinato utente ricercato.

Ritornando ai direct message il software genera un file denominato "dm_NomiUtenti.csv" che al suo interno conterrà le seguenti informazioni:

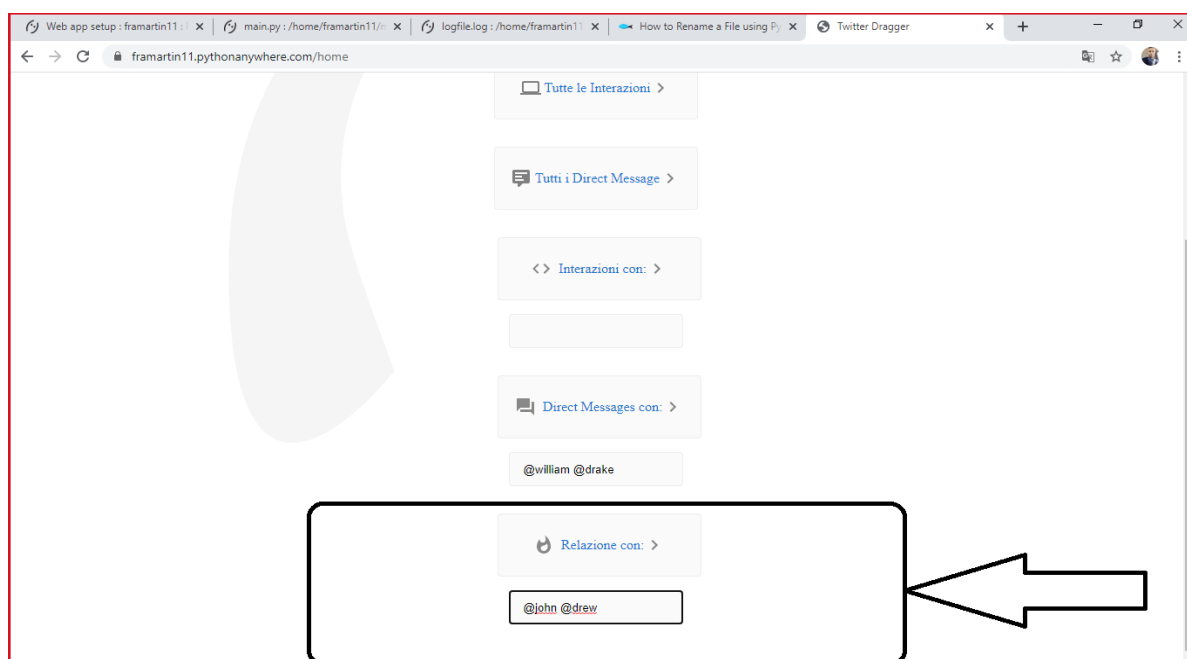
- Id: identificativo univoco del messaggio
- Timestamp: sequenza di caratteri che identificano l'evento di emissione del messaggio: data e ora
- SenderId: identificativo del mittente del messaggio
- Testo: testo del messaggio
- Username: username del mittente

3. INTERAZIONI

Nella terza scelta si prevede la possibilità di cercare tutte le interazioni con uno o più contatti. I dati verranno salvati tutti nel file *"interactions_NomiUtenti.csv"*. Ovviamente il file csv conterrà le stesse informazioni del file *allstatus.csv* con l'unica differenza di filtrare per l'username inserito.



4. RELAZIONE

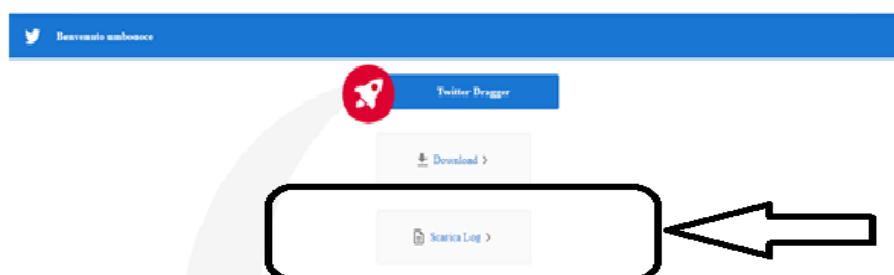


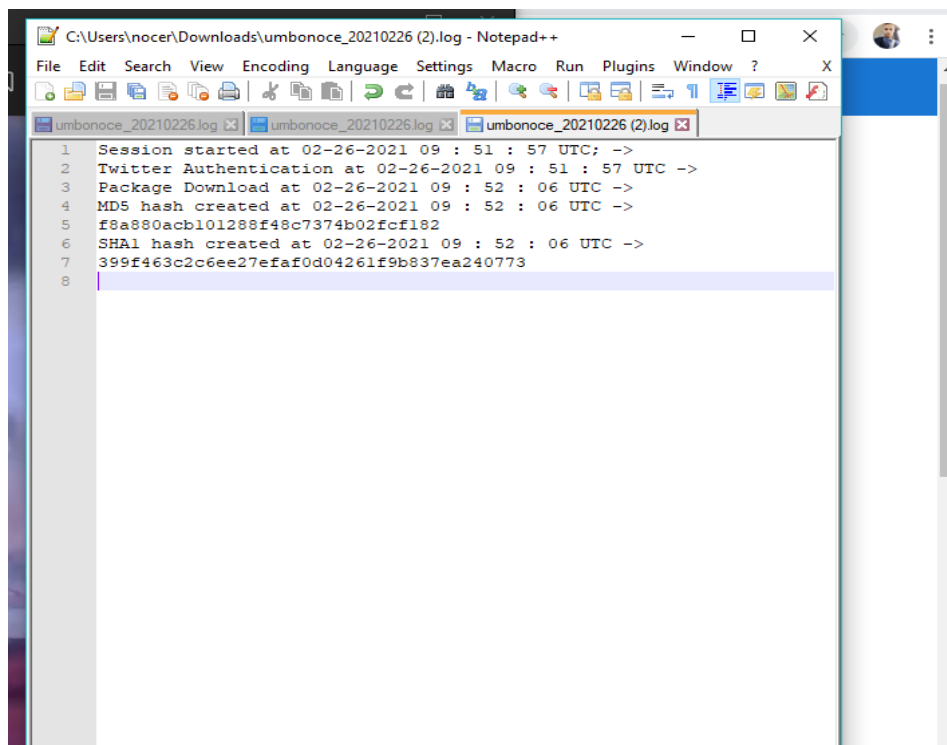
La quinta opzione permette di acquisire informazioni relative alla relazione che un determinato utente possiede con uno o più contatti. Nello specifico si valuta se l'account è seguito da quell'utente, se è stato bloccato, etc. I risultati vengono mostrati in un file csv chiamato: *"relationship_NomiUtenti.csv"* e i campi che possiede sono:

- Nome utente: nome dell'utente
- Screen name: username usato dall'utente
- Id: identificativo univoco dell'utente
- Is following: se è true account segue quell'utente
- Is followed: se è true l'account è seguito dall'utente
- Is muted: se è true l'account è stato mutato
- Is blocked: se è true account è stato bloccato
- Following request: se c'è una richiesta di following in sospeso
- No relationship: se è true i due account non hanno mai interagito

FILE DI LOG

Il sistema, durante il salvataggio delle informazioni, crea un file di log che contiene tutte le operazioni compiute dall'utente autenticato. Il file in questione contiene informazioni su data e ora delle operazioni (in formato UTC) e codice hash del file zip in questione (MD5 e SHA1).





The screenshot shows a Notepad++ window titled "C:\Users\nocer\Downloads\umbonoce_20210226 (2).log - Notepad++". The window contains a log file with the following content:

```
1 Session started at 02-26-2021 09 : 51 : 57 UTC; ->
2 Twitter Authentication at 02-26-2021 09 : 51 : 57 UTC ->
3 Package Download at 02-26-2021 09 : 52 : 06 UTC ->
4 MD5 hash created at 02-26-2021 09 : 52 : 06 UTC ->
5 f8a880acb101288f48c7374b02fcf182
6 SHA1 hash created at 02-26-2021 09 : 52 : 06 UTC ->
7 399f463c2c6ee27efaf0d04261f9b837ea240773
8
```

FILE DI HASH

Nello specifico il sistema effettua delle funzioni crittografiche di hashing ai file che vengono scaricati dall'utente. Per poter ottenere il message digest l'utente dovrà scaricare tramite l'apposito pulsante il file di log che contiene le informazioni relative allo zip.