

# 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.....	3
FILE DI LOG.....	7
FILE DI HASH.....	8

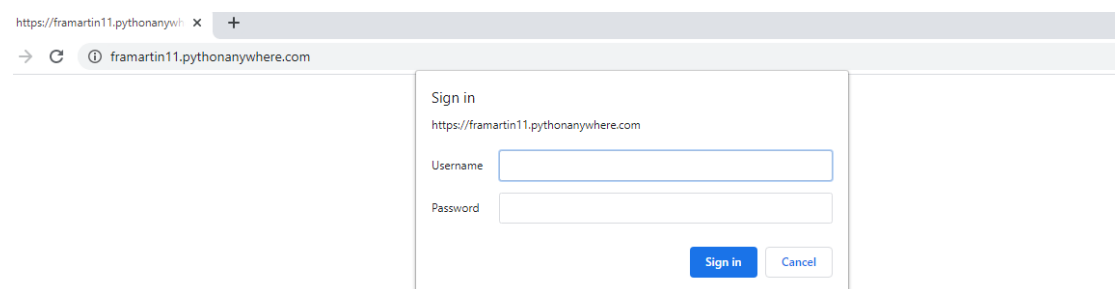
## INTRODUZIONE

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

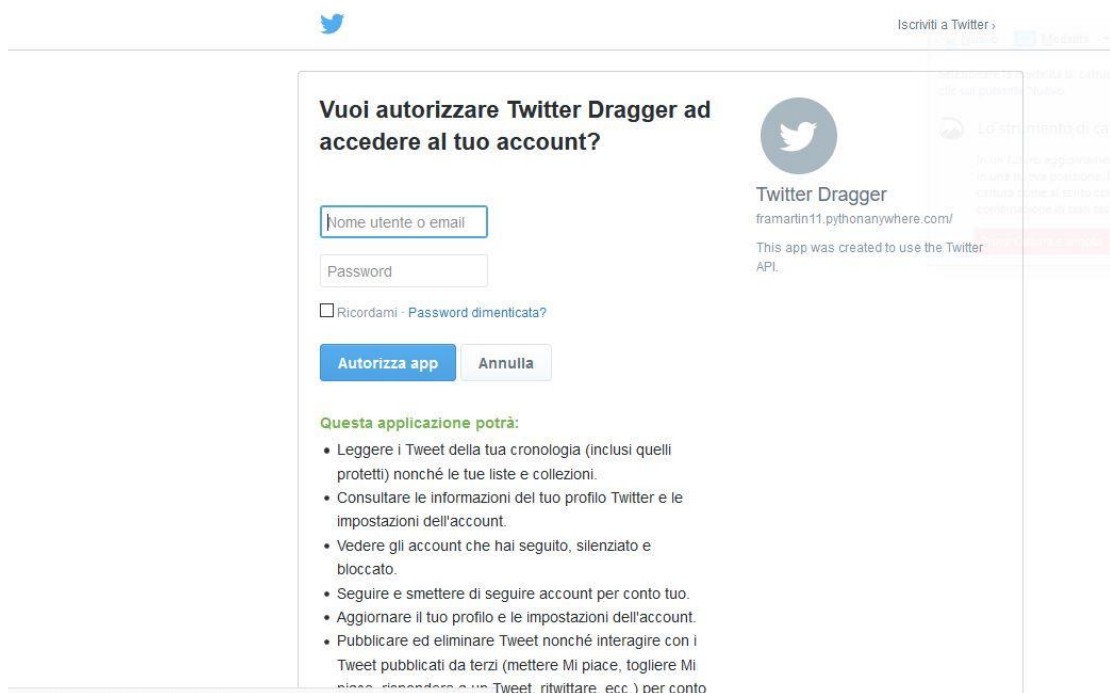
Una volta aperto il browser bisogna accedere al sito web della piattaforma:  
“ [framartin11.pythonanywhere.com](https://framartin11.pythonanywhere.com) ”

### 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

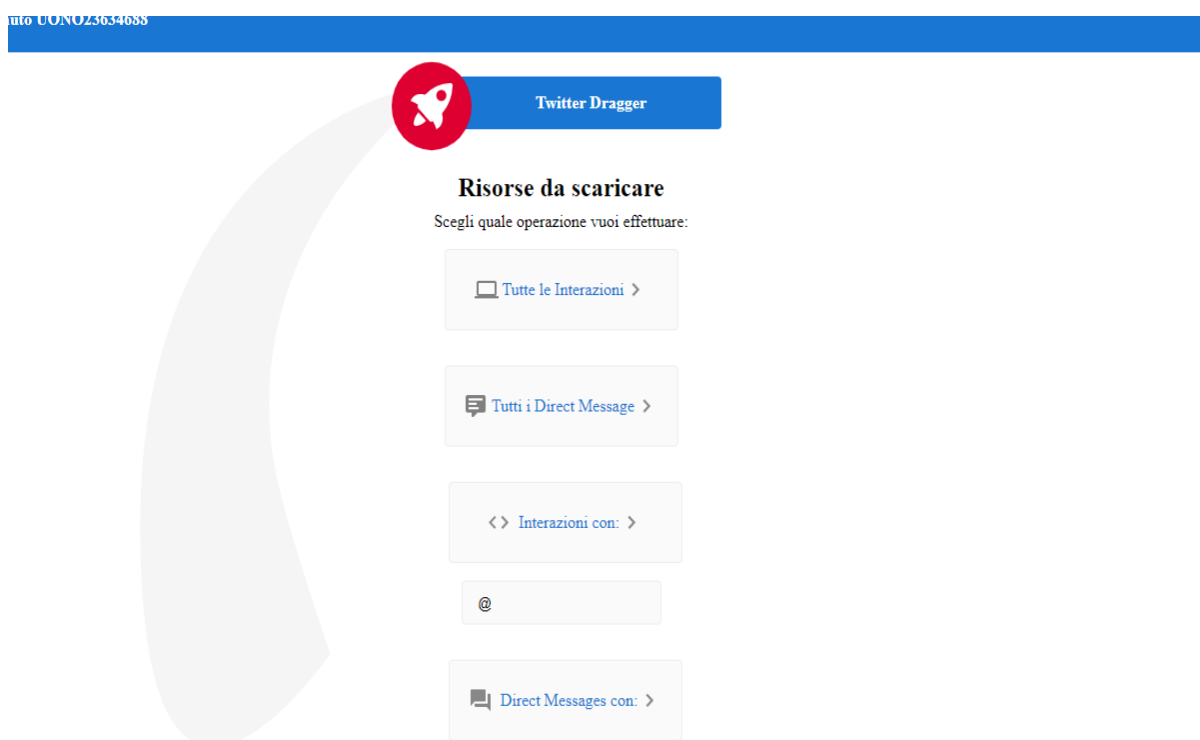


The screenshot shows the Twitter authorization interface. At the top, there's a Twitter logo and a link to 'Iscriviti a Twitter'. The main heading asks 'Vuoi autorizzare Twitter Dragger ad accedere al tuo account?'. Below this are input fields for 'Nome utente o email' and 'Password', a 'Ricordami' checkbox,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

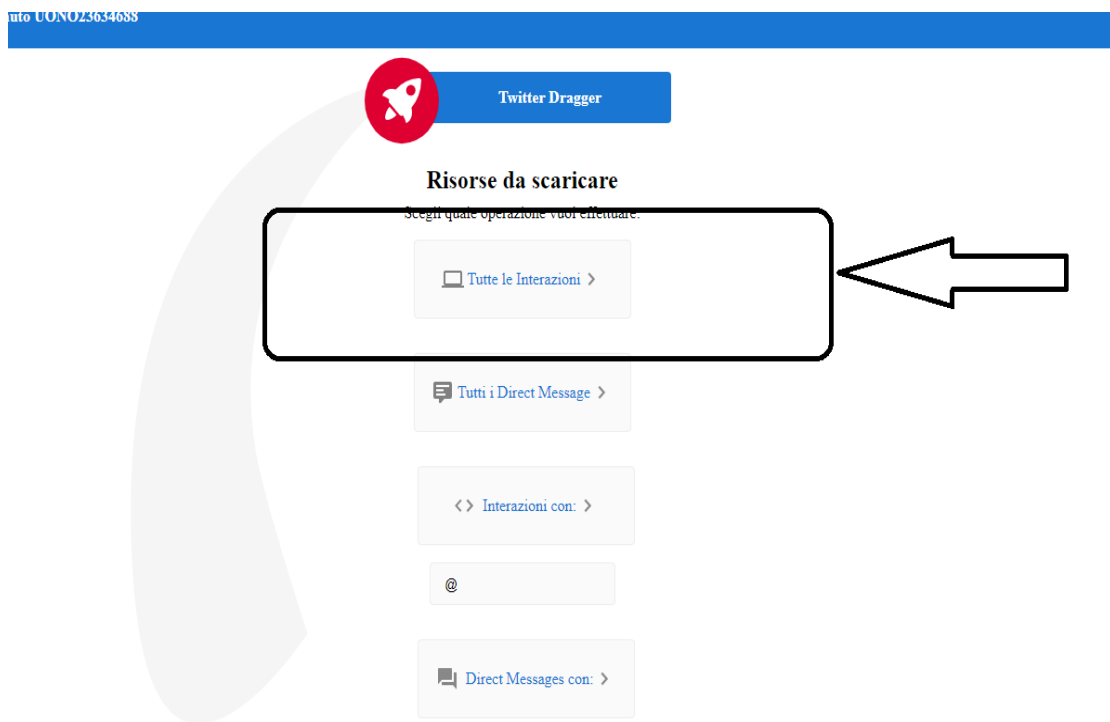
Si richiede all'utente l'inserimento di e-mail e password per autenticarsi come utente Twitter, per avere accesso alle Api.

## ISTRUZIONI

Schermata principale piattaforma:



## 1. TUTTE LE INTERAZIONI

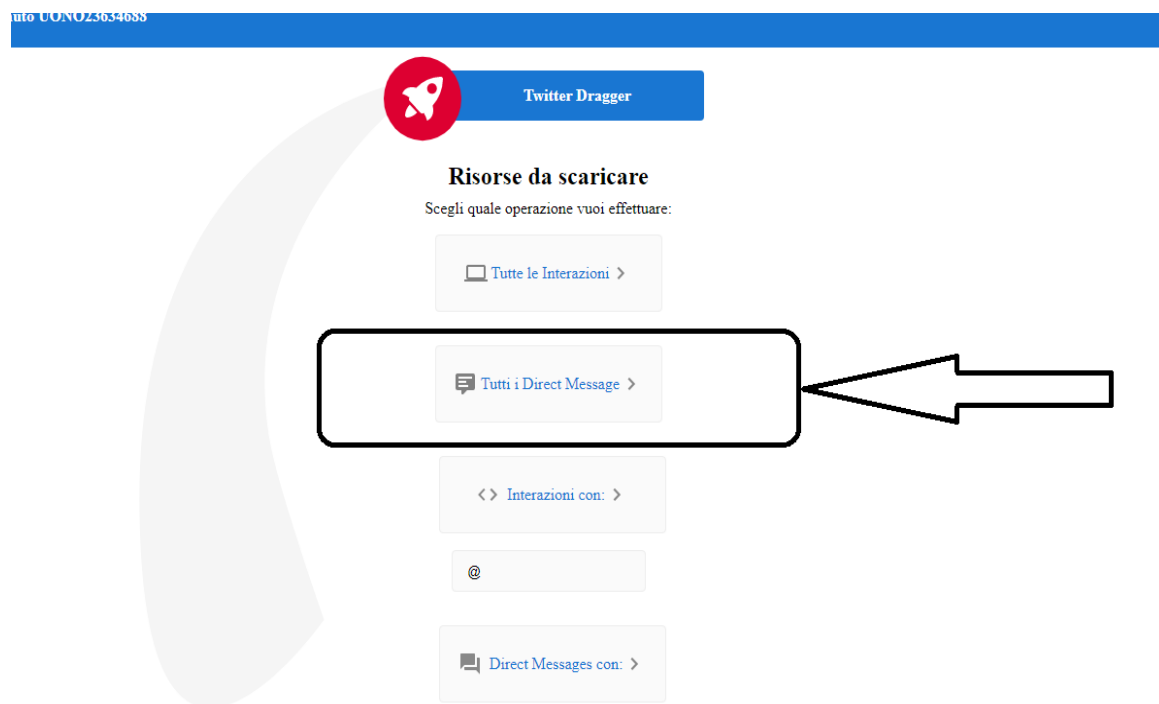


Nel primo caso d'uso sarà intuitivamente possibile digitare il pulsante in figura per accedere al file csv contenente le azioni dell'utente loggato. Il sistema genererà il file "allstatus.csv" contenente i campi:

- 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.



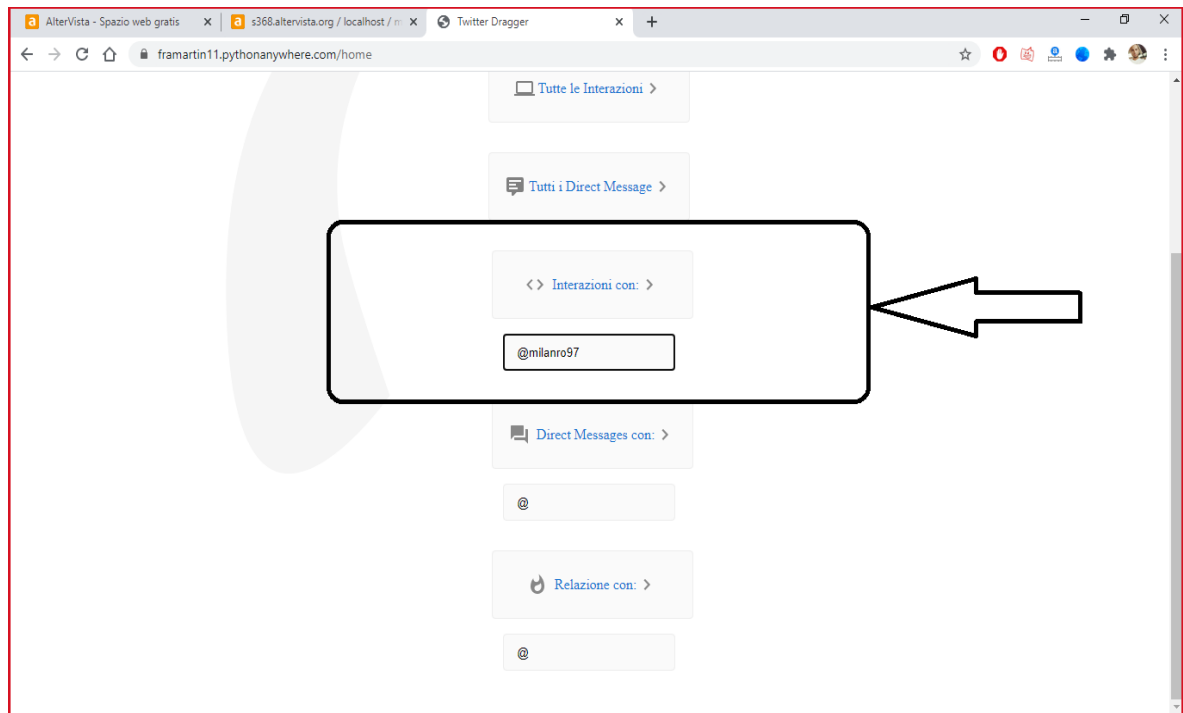
Inoltre, nelle opzioni successive il sistema consente anche di ricercare tramite una apposita checkbox una chat con uno specifico utente. Ovviamente il file risultante sarà un file csv che contiene le informazioni relative a quel determinato utente ricercato.

Ritornando ai direct message il software genera un file denominato “*dm\_NomiUtenti.csv*” che al suo interno conterrà le seguenti informazioni:

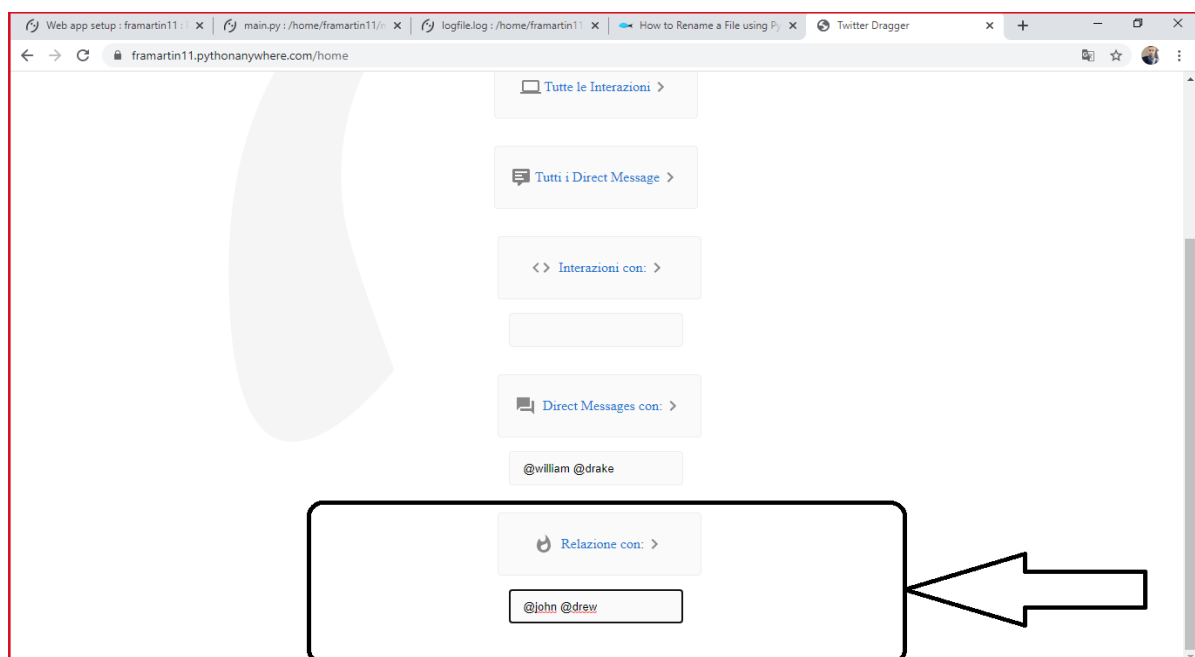
- Id: identificativo univoco del messaggio
- Timestamp: sequenza di caratteri che identificano l'evento di emissione del messaggio: data e ora
- SenderId: identificativo del mittente del messaggio
- Testo: testo del messaggio
- Username: username del mittente

### 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, dopo il salvataggio delle informazioni, crea un file di log che contiene tutte le operazioni compiute dall'utente autenticato

```
INFO: root:cb6dcdc7bdb9b577b478746279ffc912
INFO: root:8275d5ab2d5ada0d61c894a014448fac
INFO: root:589acae671f174436de3bb563725c283
INFO: root:SHA1:
INFO: root:1c6ca7ec340c79c7e6d25fcd8a55a5f97b256200
INFO: root:4087d355b81f8440d8c50c0a24d6020287e6738f
INFO: root:8bbfc641113ae950dc59ea80f5969e113ab44fc
INFO: root:b8c0b34dc46f443d6f52459057cb07c07e66f85
INFO: root:5ae353f22088fe7790ed3daf0c89fbdbb61d3a32
INFO: root:9a46a7c9d51815dea3b8915a86975bccff961364
INFO: root:23abb6c4db35f5a2577e8ee29ce8f5c52276e642
INFO: root:4a91905464b736db33fdeba078b376e8daf29bbe
INFO: root:f5ac2020b473f6a43c86baee963858ae7b431ca
INFO: root:a2ff74649ed6fc68cc8cb6545b594fd5fb0195d
INFO: root:20be63d3561eee13aa4c25346e5e3e9807c8a891
INFO: root:d1e4b737c5cf192b1da532c50c2c3f33e67b7f2e
INFO: root:954e0b09e8d971eb94aa901b0fc77938d8cad8c0
INFO: root:22026fee21de9c473e8cf0e300c587b0fd3ac709
INFO: root:/home/framartin11/mysite/dms@milanro97@_vcapo.csv ip: 10.0.0.124 Date: 24/02/2021 at 18:32:17
INFO: root:ID5:
INFO: root:8d3bead384fb8fab2f7c21d12d06e3e6
INFO: root:c200bf0da14cbfa9c4c615481dd53aff
INFO: root:c4a06ac355c1d5b94122653220491f9d
INFO: root:36fe158665e9c45cc3c53946dbcc86f8
INFO: root:SHA1:
INFO: root:8384fd36a49270739438877fad4b28dcf6b9351e
INFO: root:a085a3d8ddcc639bef9b17f979ff6904bc7f752
INFO: root:8c323171d73455c0c52eb16ea1579544cd953851
INFO: root:3ffb9916f7d73653c8501ff6f149b629cc76cb47
INFO: root:/home/framartin11/mysite/relationship@milanro97@_vcapo@dd.csv ip: 10.0.0.124 Date: 24/02/2021 at 18:32:42
INFO: root:ID5:
INFO: root:d19ce4bc92ef91e20a67e9b52d681d41
INFO: root:6fe179c0da0a016ec5b4c2816e1490b
INFO: root:SHA1:
INFO: root:8039df594581f6731bc0b4853d6af2948e74e06
INFO: root:99f86f3866d79036b4379ef704099169da6b222e
INFO: root:/home/framartin11/mysite/relationship@milanro97@_vcapo@dd@bafranco.csv ip: 10.0.0.124 Date: 24/02/2021 at 18:33:15
INFO: root:ID5:
INFO: root:6eb5c639ae29024fdabfd8091bf7c536
INFO: root:e9700500553d75a2dd0e34df7e9038e9
INFO: root:43b7379ae2b85df03303797ab0f4d021
INFO: root:SHA1:
INFO: root:21b0eabb405ed32a53837c9eba024d6b0a1a0f7c
INFO: root:8edae019b26742238d6f16a0d18c57ec95d173d
INFO: root:656084839dae56353d33cd3f4794dcbea9cb001
```



Il file in questione contiene informazioni circa l'IP dell'utente, data e ora delle operazioni, nome del file scaricato, e contiene anche il codice hash del file in questione (MD5 e SHA1).

## 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 non dovrà far altro che premere F5 dopo aver scaricato il file in questione. Dopo aver premuto F5 si aprirà sulla interfaccia un dialog contenente le informazioni relative a quel file (nome, ip, data, ora e message digest)

