

Heliaca Finder

Sommario

Heliaca Finder	1
Introduzione	2
Step 1 - Individuazione	2
Step 2 – Tracciamento.....	2
Step 3 – Quantificazione	2
Home.....	3
Data Discovery	4
Data Discovery Option	4
Esecuzione della Data Discovery	5

Introduzione

Heliaca Finder è un software per l'esecuzione della data discovery. In pratica viene avviato un processo finalizzato a "Individuare", "Tracciare" e "Quantificare" i dati sensibili presenti all'interno della sua azienda. Quindi i tre step fondamentali sono:

Step 1 - Individuazione

Attraverso un accurato processo di ricerca – la cui durata dipenderà dalla posizione e dal volume di dati da analizzare – vengono identificati i dati sensibili contenuti nei file nei formati più comuni, quali PDF, Word, Excel, ecc. Anche i formati immagine (ad es. JPG, PNG, ecc.), che solitamente non possono essere analizzati automaticamente, vengono processati grazie a tecnologie avanzate basate sull'intelligenza artificiale (OCR), permettendo così il riconoscimento e l'estrazione del testo anche da questi file.

Step 2 – Tracciamento

Una volta individuati tutti i file contenenti dati sensibili, il processo prosegue con il tracciamento di tali dati. In pratica, si ottiene un elenco dettagliato che indica il file sensibile trovato, il tipo di dato sensibile, il campione, il numero di occorrenze per ogni file e infine il livello di criticità. Successivamente sarà possibile esportare questi dati in formato Excel (.xlsx) e PDF.

Step 3 – Quantificazione

Con l'esportazione del file PDF si procede alla quantificazione dei dati e dei file sensibili. Esso consente di determinare non solo il numero di file sensibili, ma anche la quantità di dati sensibili presenti in ciascun file (in termini di occorrenze). Questo step permette, quindi, di valutare la mole di dati sensibili trovati, distinguendo anche il tipo di dato rilevato – ad esempio, un classico dato sensibile è il codice fiscale, ma ne esistono numerosi altri con differenti livelli di criticità.

È importante sottolineare che, secondo il regolamento europeo sul GDPR, per dato "personale" (ossia sensibile) si intende qualsiasi informazione riguardante una persona fisica identificata o identificabile. Tale definizione evidenzia l'importanza di una corretta quantificazione dei dati sensibili individuati.

Home

Per ora nella home non c'è tanto da vedere a parte il logo (che apprezzerete di più se laziali) e lo slogan, il logo se cliccato fornirà accesso al sito dell'azienda.

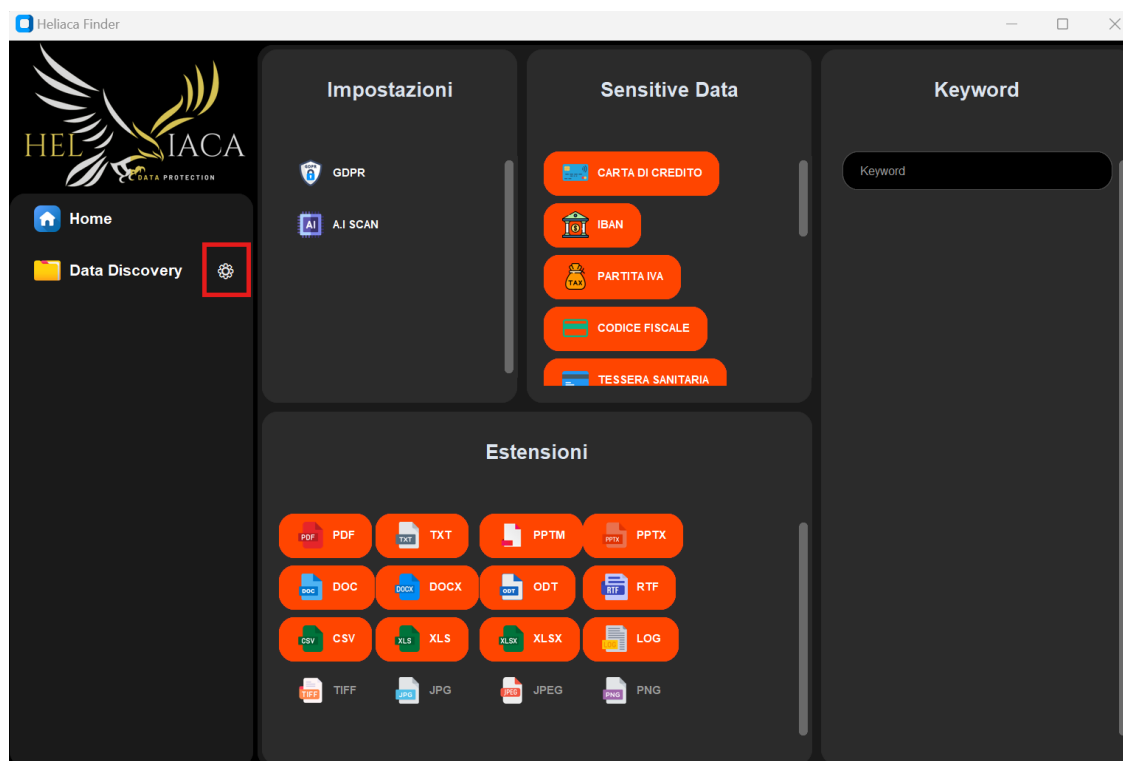


Data Discovery

Come già scritto il processo di data discovery porta ad individuare inizialmente i file sensibili presenti. Per avviarla dobbiamo procedere innanzitutto a controllare la configurazione della discovery, anche se sono impostati di default dei settaggi per cui potete iniziare a farla direttamente dalla sezione apposita.

Data Discovery Option

In questa sezione, la quale si accede premendo l'ingranaggio rappresentato nello screenshot, potrete selezionare vari pulsanti per modificare la ricerca.



Nella sezione “**Impostazioni**” premendo sul tasto GDPR, potrete applicare una selezione o deselezione rapida dei tipi di dati sensibili contenuti nella sezione “Sensitive Data”, premendo sul tasto A.I SCAN potrete abilitare l’ocr che permetterà l’estrazione del testo da scansioni in PDF e dai formati TIFF, JPG, JPEG e PNG i quali questi ultimi disabilitati di default.

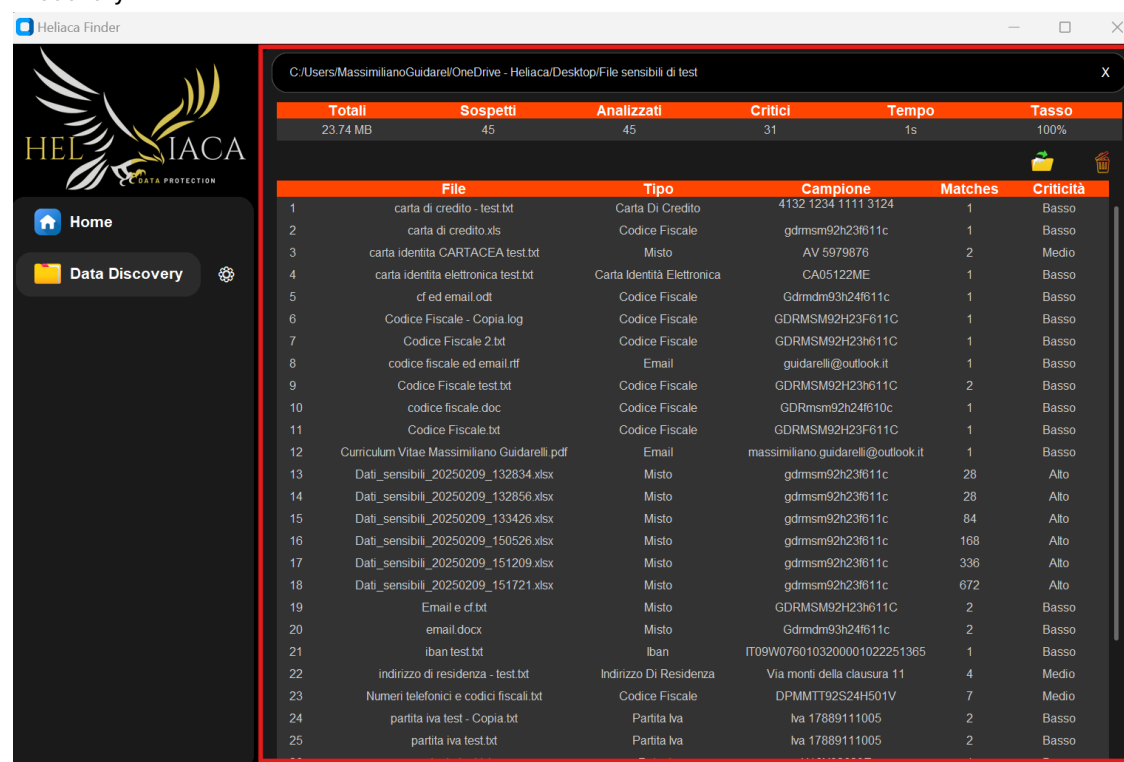
Nella sezione “**Sensitive Data**” potrete selezionare o deselezionare il tipo di dato sensibile che volete ricercare, essi corrispondono a dei pattern (regex) di testo che individuano appunto il tipo di dato relativo.

Nella sezione “**Estensioni**” potrete selezionare il tipo di estensione desiderata che volete far analizzare, tutti gli altri formati deselezionati saranno ignorati dalla ricerca. I formati TIFF, JPG, JPEG e PNG devono essere abilitati dal tasto A.I SCAN per poter essere selezionati.

Nella sezione “**Keyword**” potrete inserire delle parole chiave che volete usare nella ricerca. Se ad esempio volete cercare il vostro nome nei file compresi nella ricerca potete aggiungerlo grazie a questa sezione.

Esecuzione della Data Discovery

Una volta fatte le dovute impostazioni potete procedere con la Data Discovery.



The screenshot shows the Heliaca Finder application window. The left sidebar contains a logo for HELIACA and two buttons: 'Home' and 'Data Discovery'. The main area displays a table of search results for the path 'C:/Users/MassimilianoGuidarelli/OneDrive - Heliaca/Desktop/File sensibili di test'. The table has two main sections. The first section is a summary table with columns: Totali, Sospetti, Analizzati, Critici, Tempo, and Tasso. The second section is a detailed table with columns: File, Tipo, Campione, Matches, and Criticità.

Totali	Sospetti	Analizzati	Critici	Tempo	Tasso
23.74 MB	45	45	31	1s	100%

File	Tipo	Campione	Matches	Criticità
1 carta di credito - test.bt	Carta Di Credito	4132 1234 1111 3124	1	Basso
2 carta di credito.xls	Codice Fiscale	gdrmsm92h23f611c	1	Basso
3 carta identità CARTACEA test.bt	Misto	AV 5979876	2	Medio
4 carta identità elettronica test.bt	Carta Identità Elettronica	CA05122ME	1	Basso
5 cf ed email.odt	Codice Fiscale	Gdrmdm93h24f611c	1	Basso
6 Codice Fiscale - Copia.log	Codice Fiscale	GDRMSM92H23F611C	1	Basso
7 Codice Fiscale 2.bt	Codice Fiscale	GDRMSM92H23h611C	1	Basso
8 codice fiscale ed email.rtf	Email	guidarelli@outlook.it	1	Basso
9 Codice Fiscale test.bt	Codice Fiscale	GDRMSM92H23h611C	2	Basso
10 codice fiscale.doc	Codice Fiscale	GDRMSM92h24f610c	1	Basso
11 Codice Fiscale.bt	Codice Fiscale	GDRMSM92H23F611C	1	Basso
12 Curriculum Vitae Massimiliano Guidarelli.pdf	Email	massimiliano.guidarelli@outlook.it	1	Basso
13 Dati_sensibili_20250209_132834.xlsx	Misto	gdrmsm92h23f611c	28	Alto
14 Dati_sensibili_20250209_132856.xlsx	Misto	gdrmsm92h23f611c	28	Alto
15 Dati_sensibili_20250209_133426.xlsx	Misto	gdrmsm92h23f611c	84	Alto
16 Dati_sensibili_20250209_150526.xlsx	Misto	gdrmsm92h23f611c	168	Alto
17 Dati_sensibili_20250209_151209.xlsx	Misto	gdrmsm92h23f611c	336	Alto
18 Dati_sensibili_20250209_151721.xlsx	Misto	gdrmsm92h23f611c	672	Alto
19 Email e cf.bt	Misto	GDRMSM92H23h611C	2	Basso
20 email.docx	Misto	Gdrmdm93h24f611c	2	Basso
21 iban test.bt	Iban	IT09W0760103200001022251365	1	Basso
22 indirizzo di residenza - test.bt	Indirizzo Di Residenza	Via monti della clausura 11	4	Medio
23 Numeri telefonici e codici fiscali.bt	Codice Fiscale	DPMMTT92S24H501V	7	Medio
24 partita iva test - Copia.bt	Partita Iva	Iva 17889111005	2	Basso
25 partita iva test.bt	Partita Iva	Iva 17889111005	2	Basso

Per iniziatala basta fare doppio click nella entry in alto così da selezionare la cartella o percorso da analizzare, si può anche copiare un percorso all'interno e poi far partire il processo premendo Enter. Una volta partito saranno trovati tutti i file contenuti dentro le cartelle e sottocartelle specificate dal percorso immesso.

Nella prima tabella possiamo vedere descritti da sinistra verso destra:

- **Totali:** Il volume di dati analizzato.
- **Sospetti:** Il numero di file sospetti che sono stati rilevati in base ai criteri impostati.
- **Analizzati:** Il numero di file che vengono analizzati durante il processo che dovrà corrispondere al termine con quelli sospetti.
- **Critici:** Il numero di file critici trovati.
- **Tempo:** Il tempo trascorso dall'inizio del processo.
- **Tasso:** Percentuale di completamento.

Nella seconda tabella vengono conservati i risultati dell'analisi, sempre da sinistra verso destra, troviamo:

- **File:** Il nome del file trovato, con un doppio click possiamo aprire il file in questione sperando che il nostro sistema abbia l'editor giusto.
- **Tipo:** Il tipo di dato sensibile trovato.

- Campione: Il primo dato sensibile trovato nel file.
- Matches: Che rinominerò in “Occorrenze” indica quanti dati sensibili sono stati trovati all’interno del file. Cliccando sulla cella relativa si aprirà una finestra che conterrà tutti i dati sensibili trovati nel file.
- Criticità: Grazie ad un algoritmo in base al numero di occorrenze e tipo di dato trovato viene deciso il livello di criticità che sarà “Basso”, “Medio” e “Alto”.

Una volta terminato il processo di ricerca puoi estrapolare un report pdf e xlsx che conterranno il risultato della data discovery premendo il tasto “Folder” tra le due tabelle a destra.

Premendo sul tasto del “Cestino” puoi cancellare la ricerca effettuata.