Heureka Social Security Number Detection

August 2017

Heureka SOFTWARE®

CONTENTS

Summary

Heureka's Intelligence Platform includes auto detection of U.S. Social Security numbers. Social Security numbers are automatically displayed on the risk dashboard and may be used in conjunction with the search criteria page to narrow valid results in combination with keywords.

Social Security Number validation

The United States Social Security Administration began using social security number randomization on June 25, 2011. Before this date, the Social Security Administration published a "High Group List" which Heureka incorporates to determine whether a number is valid and assigned to a person. After June, 25, 2011 numbers became random and the High Group List is no longer used.

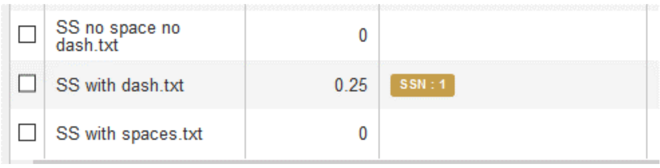
More information regarding the High Group List can be found here:
https://www.ssa.gov/employer/ssnvhighgroup.htm

More information on Social Security Number Randomization can be found here:
https://www.ssa.gov/employer/randomization.html

Pattern Recognition

Heureka uses the Social Security Administration's recognized pattern of 3 digits, hyphen, 2 digits, hyphen followed by four digits (XXX-XX-XXXX). Heureka does not automatically tag files outside of this structure. For example, numbers separated by periods (.) or 9-digit numbers will not be automatically tagged as Social Security Numbers. (XXX.XX.XXXX or XXXXXXXXX).

Tagging and pattern recognition can be demonstrated with three simple test files (see figure 1). Our test files are formatted in three separate ways; one has a valid number with no dashes or spaces, one is properly formatted and one replaces dashes with spaces. As you can see in figure 1, only the properly formatted file receives an automatic tag and risk score within Interrogate.



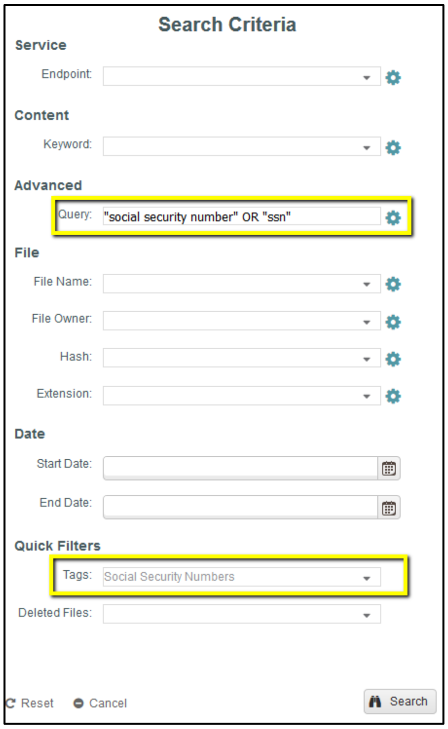
□	SS no space no dash.txt	0	
□	SS with dash.txt	0.25	SSN : 1
□	SS with spaces.txt	0	

Figure 1

Reducing False Positives

False positive tags can and will occur in the system. This is mainly due to the fact that social security numbers began to be randomized after June 25, 2011. Heureka helps to reduce false positives by limiting the pattern recognition to include dashes (-) instead of tagging 9-digit numbers or accepting random numbers with spaces or periods.

Additional reduction of false positives can be achieved by combining Heureka's "Quick Filters" tags with keywords or queries. This combination of searching allows the system to effectively search within a tagged file for additional information such as the query "social security number" OR "ssn". The system narrows the search to only the files that contain the automated Heureka social security number tag and your keyword/query.


Search Criteria

Query: "social security number" OR "ssn"

Tags: Social Security Numbers

Figure 2 Reducing False Positives

Heureka Software is a technical leader in endpoint search, identify and classification software. Our goal is to bring order to unstructured data by identifying risk while helping you realize the value of unstructured data across all endpoints.

Heureka Software, LLC

1382 W. 9th Street, Suite 410S

Cleveland, Ohio 44113