NORTHWESTERN
UNIVERSITY

Character Strings for Search Tools

The following strings will be useful for finding data using many of these search tools. The tools search for data by matching character strings using regular expressions or regex. You can find a tutorial on using regex's at http://en.wikipedia.org/wiki/Regex.

These strings have been prepared for use with Cornell's Spider program. They should work for other search programs as well, but may require slight sytax modifications, depending on the version of regex your program supports.

You may combine any number of search strings into a single search by joining them with the "|" character. You may need to replace the "|" with "\|" for some versions of Unix, including Mac OS.

When conducting your search, you should also look at all files, including binary, as well as examine the contents of compressed files if possible. On a Unix system, grep -a will examine binary files as text.

Please examine the contents of any reported files carefully. Just because you get a particular hit does not automatically mean the data is of concern.

SSN w/ dashes: [0-7]\d{2}\-\d{2}\-\d{4}
SSN w/ dashes and breaks: \b[0-7]\d{2}\-\d{2}\-\d{4}\b

SSN – 9 consecutive digits: [0-7]\d{8}
SSN – 9 consecutive digits with breaks: \b[0-7]\d{8}\b

SSN w/ spaces: [0-7]\d{2}\s\d{2}\s\d{4}
SSN w/ spaces and breaks: \b[0-7]\d{2}\s\d{2}\s\d{4}\b

All SSN search options with breaks
\b[0-7]\d{2}\-\d{2}\-\d{4}\b|\b[0-7]\d{8}\b|\b[0-7]\d{2}\s\d{2}\s\d{4}\b

All SSN search options with no breaks
[0-7]\d{2}\-\d{2}\-\d{4}|[0-7]\d{8}|[0-7]\d{2}\s\d{2}\s\d{4}

Visa/Mastercard/Discover
\d{4}\-\d{4}\-\d{4}\-\d{4}|\d{4}\s\d{4}\s\d{4}\s\d{4}

Visa/Mastercard/Discover with breaks
\b\d{4}\-\d{4}\-\d{4}\-\d{4}\b|\b\d{4}\s\d{4}\s\d{4}\s\d{4}\b

American Express
\d{4}\-\d{6}\-\d{5}|\d{4}\s\d{6}\s\d{5}

American Express with breaks
\b\d{4}\-\d{6}\-\d{5}\b|\b\d{4}\s\d{6}\s\d{5}\b