-->

RAND Data Sets

Description

This dataset contains 5,000 examples of file created in the Pure Random Numbers RAND format. This data set forms part of a much larger mixed file data set, called NapierOne, created by the School of Computing at Edinburgh Napier University in 2021. When using this dataset attribution should be given to Edinburgh Napier University as well as following any specific attribution requirements. Details of which appear in the License section below. All files in the dataset have been validated to be of the correct format, having the correct extension and magic numbers where appropriate.

The files have also been scanned for viruses using Windows Defender© at the following level:

Scan Component	Component Version
Antimalware Client Version	4.18.2101.9
Engine Version	1.1.17800.5
Antivirus Version	1.331.1067.0
Anti-spyware Version	1.331.1067.0

The files have also been submitted to www.virustotal.com to confirm that they do not contain any malicious code.

Creation

To populate this dataset, the files were created by downloading random numbers from https://www.random.org/, which is a provider of pure random numbers https://en.wikipedia.org/wiki/Randomness

To automate the file download, the random.org API interface was used, utilizing the tool developed by Koen Dercksen https://github.com/KDercksen/randompy

To provide file name consistency across the entire dataset, the retrieved files were renamed using the following syntax:

> **\<sequence-number\>-\<original-extension\>\<-extra-information\>.\<extension\>**

Where:

File name fragment	Description
\<sequence-number\>	is unique number within this dataset. Files derived from this RAND dataset, for example the RAND-NOMAGIC and RAND-PASSWORD datasets, will retain the same sequence number across all datasets. This facilitates the easy identification of related files between related datasets.
\<original-extension\>	is the original extension that the file had, indicating the format of the file. Some datasets contain files, where the actual extension does not reflect the actual format of the original file. This especially relates to files that have been encrypted by malware.
\<-extra-information\>	is a placeholder in the file name that can be used to provided optional additional information relating to the file.
\<extension\>	is the extension of the file name. Normally the characters after the '.' in a file name. The extension is normally

File name fragment	Description

used to identify the content and format of the file.

So for this sub dataset an example of a file name could be: 0001-rand-pure.rand

A mapping of the original file name to the new dataset file name is held by the researchers and is available on request.

Information

If you require further information, then please feel free to contact the researchers below:

Researcher	Contact
Simon Davies	s.davies@napier.ac.uk ✉
Rich Macfarlane	r.macfarlane@napier.ac.uk ✉
Bill Buchanan	b.buchanan@napier.ac.uk ✉

Related Data Sets

- RANDOM/PURE
- RANDOM/PSEUDO

Provenance

The data used to populate the files in this dataset was provided by https://www.random.org/

License

Dataset Contents

Documentation created at 16:47 on May 24, 2021