

Answers Lab2

Alicia Ruiz Tovar

13/01/2019

Soluciones del laboratorio 2

Documento Markdown R con las soluciones del laboratorio 3 para la asignatura de Data Driven Security del Máster en CyberSecurity Management.

CPE

1. Obtención del fichero

Nos descargamos el fichero y lo transformamos a xml para poder tratarlo.

```
compressed_cpes_url <- "https://nvd.nist.gov/feeds/xml/cpe/dictionary/official-cpe-dictionary_v2.3.xml"
cpes_filename <- "cpes.zip"
download.file(compressed_cpes_url, cpes_filename)
unzip(zipfile = cpes_filename)
cpe.file <- "./official-cpe-dictionary_v2.3.xml"
cpes.xml <- xml2::read_xml(cpe.file)
```

2. Creación del data frame

Una vez transformado a xml, lo convertimos en un data frame.

```
cpe.title <- xml2::xml_find_all(cpes.xml, "//*[cpe-23:cpe23-item]/*[1]")
cpe.name <- xml2::xml_find_all(cpes.xml, "//*[cpe-23:cpe23-item]/*/@name")

cpe <- data.frame(title = xml2::xml_text(cpe.title),
                  cpe.23 = xml2::xml_text(cpe.name),
                  stringsAsFactors = F)
```

3. CPE limpio

Modificamos el data frame añadiendo el nombre de las columnas y almacenando la información de una forma más elegante.

```
col.names <- c("cpe",
               "v",
               "part",
               "vendor",
               "product",
               "version",
               "update",
               "edition",
               "language",
               "sw_edition",
               "target_sw",
```

```

      "target_hw",
      "other")

cpe$cpe.23 <- stringr::str_replace_all(cpe$cpe.23, "\\:\\:", ";")
cpe.clear <- tidyr::separate(data = cpe, col = cpe.23, into = col.names, sep = ":", remove = F)
cpe.clear$cpe <- as.factor(cpe.clear$cpe)
cpe.clear$v <- as.factor(cpe.clear$v)
cpe.clear$part <- as.factor(cpe.clear$part)
cpe.clear$vendor <- as.factor(cpe.clear$vendor)
cpe.clear$product <- as.factor(cpe.clear$product)
cpe.clear$language <- as.factor(cpe.clear$language)
cpe.clear$sw_edition <- as.factor(cpe.clear$sw_edition)
cpe.clear$target_sw <- as.factor(cpe.clear$target_sw)
cpe.clear$target_hw <- as.factor(cpe.clear$target_hw)
cpe.clear$other <- as.factor(cpe.clear$other)

```

Tiempo dedicado

El tiempo total de dedicación a la resolución de este laboratorio, en el que se incluyen el tiempo de comprensión del ejercicio, búsqueda de información, investigación de librerías, programación y ejecución de comandos y redacción del documento, ha sido de: **14 horas**.