

Aggregate record data

```
$ curl -s https://cdn.kernel.org/pub/linux/kernel/v4.x/linux-4.4.1.tar.xz | # Fetch Linux kernel
> xz -dc | # Uncompress
> tar tvf - >contents.txt # List archive contents
$ head contents.txt
drwxrwxr-x root/root          0 2016-01-31 21:29 linux-4.4.1/
-rw-rw-r-- root/root         31 2016-01-31 21:29 linux-4.4.1/.get_maintainer.ignore
-rw-rw-r-- root/root       1280 2016-01-31 21:29 linux-4.4.1/.gitignore
-rw-rw-r-- root/root       5537 2016-01-31 21:29 linux-4.4.1/.mailmap
-rw-rw-r-- root/root      18693 2016-01-31 21:29 linux-4.4.1/COPYING
-rw-rw-r-- root/root     97181 2016-01-31 21:29 linux-4.4.1/CREDITS
drwxrwxr-x root/root          0 2016-01-31 21:29 linux-4.4.1/Documentation/
-rw-rw-r-- root/root     17201 2016-01-31 21:29 linux-4.4.1/Documentation/00-INDEX
drwxrwxr-x root/root          0 2016-01-31 21:29 linux-4.4.1/Documentation/ABI/
-rw-rw-r-- root/root     3813 2016-01-31 21:29 linux-4.4.1/Documentation/ABI/README
$ awk '
> { size += $3; n++ } # Sum size and number of files
> END { # Print summary
>   print "Total size " size
>   print "Number of files " n
>   print "Average file size " size / n
>}
>' contents.txt
Total size 606232826
Number of files 55708
Average file size 10882.3
$ awk '
> {
>   sub("./", "", $6) # Remove path
>   if (!sub("./\\.", "", $6)) # Keep only extension
>     next # Skip files without extension
>   size[$6] += $3 # Tally size of extension
> }
> END {
>   for (i in size)
>     print i, size[i]
> }' contents.txt |
> sort -k 2nr |
> head
c 407336456
h 136403607
txt 14584930
S 10216790
dtsi 5508483
ihex 4685001
dts 4346879
xml 1951892
tmpl 1084413
HEX 900059
$
```

Extract attachments

```
$ head message.eml
Message-ID: <55A5EBE9.2040005@aueb.gr>
Date: Fri, 21 Feb 2020 14:00:16 +0200
From: Diomidis Spinellis <dds@aueb.gr>
MIME-Version: 1.0
To: reviews@publons.com
Subject: thank you
Content-Type: multipart/mixed;
  boundary="-----020706090003010408040400"

$ cat <<\EOF >split.awk
#!/bin/awk -f
$0 ~ "^" SEP "${ out++ }
out { print >>"file-" out ".att" }
EOF
```

```
$ ./split.awk SEP=-----020706090003010408040400 message.eml
$ ls
file-1.att  file-16.att  file-22.att  file-29.att  file-35.att  file-9.att
file-10.att file-17.att  file-23.att  file-3.att   file-36.att  message.eml
file-11.att file-18.att  file-24.att  file-30.att  file-4.att   split.awk
file-12.att file-19.att  file-25.att  file-31.att  file-5.att
file-13.att file-2.att   file-26.att  file-32.att  file-6.att
file-14.att file-20.att  file-27.att  file-33.att  file-7.att
file-15.att file-21.att  file-28.att  file-34.att  file-8.att
$
```