Web log regular expression

Start and end of quoted region

Match [and then
1 or more characters not]
and then]
Capture the content
between [and]

Referrer
Field
need to allow
for empty value so *
" then
0 or more characters not "
then "

Start of string

Match literal
-

Path for
file name

Version of
HTTP

`'^([0-9.]+) - ([^ ]+) \\[([^]]+)\\] "([A-Z]+) ([^ ]+) HTTP/(1.[01])" ([0-9]+) (-|[0-9]+) "([^"]*)" "([^"]+)"'`

Simple match of
IP address by
1 or more
digits and .

login string -
everything but a
space, then
a space

Match the GET,
HEAD, POST, …

Fixed &
not captured

HTTP
Status
Code, e.g.,
200, 302,
404

Number of bytes
or -

" 1 or more
non-" characters
followed by "

Could add $ to end the string.
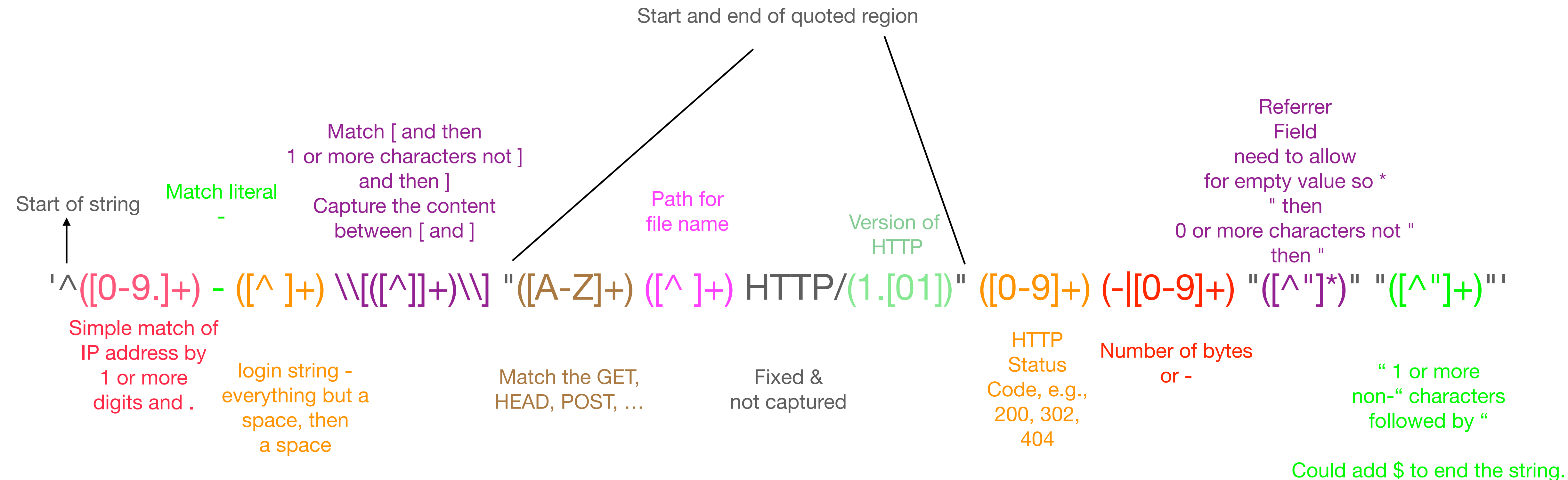
137.208.57.141 - - [01/Nov/2015:06:37:16 -0800] "HEAD /cgi-bin/testForm1.pl HTTP/1.1" 404 - "-" "-"

73.235.166.114 - - [01/Nov/2015:08:01:36 -0800] "GET /stat141/Data/digitsTrain.csv HTTP/1.1"
200 9135013 "-" "R (3.1.2 x86_64-w64-mingw32 x86_64 mingw32)"