

```
#!/bin/gawk -f
#
# Some stuff found take from here:
# http://www.catonmat.net/blog/downloading-youtube-videos-with-gawk/
#
# Usage: Usage: gawk -f tubeload.awk VID1 VID2 VID3 ...
# You may use just the video Ids (like m-xlJRh5zSM) or
# the full youtube link (like http://www.youtube.com/watch?v=1i-L3YTeJJM)
```

```
BEGIN {
    numvideos = 0;

    for (i=1;i<ARGC;i++) {
        videoid = truncateurl(ARGV[i]);
        print "Input Video ID: "videoid;
        ID[numvideos]=videoid;
        numvideos++;
    }

    for (i = 0;i<numvideos;i++) {
        print "#####";
        print "Processing video " (i+1) " of "numvideos;
        print "#####";
        process(ID[i]);
    }
}
```

```
function truncateurl(url) {
    gsub(/http:\\\/\\\/, "", url);
    gsub(/www\\.\/, "", url);
    gsub(/youtube\\.com\\/watch\\?v=\/, "", url);

    if ((p = index(url, "&")) > 0) {
        url = substr(url, 1, p-1);
    }

    return url;
}
```

```
function process(thisid) {
    print "Processing video id: "thisid;

    #####
    # GET INFO
    #####
    title="";
    descr="___."thisid"";
    dlurl="";

    geturl = "http://www.youtube.com/watch?v="thisid;
    sys = "wget -O - -q " geturl;
    print "execute: "sys;

    found5 = 0;

    while ((sys|getline)>0) {
        # VIDEO URL MAP
        if (match($0, /"fmt_url_map": "([^\"]+)"/, matches)) {
            map = unescape(matches[1]);
            numentries = split(map,maparr,"|");
            for (entloop = 2;entloop <= numentries;entloop++) {
                entry = maparr[entloop];
                gsub(",[0-9]+$", "", entry);

                if (!found5) {
                    dlurl=entry;
                }
                # fmt number 5 is the best video
                if (entry ~ /itag=5/) {
                    dlurl=entry;
                    found5=1;
                }
            }
        }
    }
    ## Get Title of Video
    if (match($0, /<h1 *>([^\<]+)</, matches)) {
```

```

        title = matches[1];
        print "Title found: "title;
    }
}
close(sys);

#####

if (dlurl == "") {
    print "No appropriate VIDEO URL found. Skip.";
}
else {
    descr = "v_"title descr;
    gsub("[^0-9a-zA-Z]", "_", descr);

    outfile = "./"descr".flv";

    print "Video Info: "thisid;
    print "URL:         "dlurl;
    print "Title:        "title;
    print "Description: "descr;
    print "target file: "outfile;

    sys="wget -c -O \"\"outfile\" \"dlurl\"";
    print "Execute "sys;
    sysresult = system(sys);
    print "SYSRESULT:"sysresult;
}

}

function unescape(str, nmatches, entity, entities, seen, loop) {
    nmatches = find_all_matches(str, "[%0-9A-Fa-f][0-9A-Fa-f]", entities);
    for (loop = 1; loop <= nmatches; loop++) {
        entity = entities[loop];
        if (!seen[entity]) {
            if (entity == "%26") {
                gsub(entity, "\\&", str);
            }
            else {
                unentity = entity;
                sub("%", "", unentity);
                unentity = sprintf("%c", strtonum("0x" unentity));

                gsub(entity, unentity, str);
            }
            seen[entity] = 1;
        }
    }
    return str;
}

function find_all_matches(str, regex, resultarr, numresults, a, b) {
    numresults=0;
    a = RSTART;
    b = RLENGTH;

    while (match(str, regex) > 0) {
        numresults++;
        resultarr[numresults] = substr(str, RSTART, RLENGTH);
        str = substr(str, RSTART+RLENGTH);
    }

    RSTART = a;
    RLENGTH = b;
    return numresults;
}

```