2. WRITE CODE: Id3v2parser

Description

Id3v2parser is a program to retrieve information from ID3v2.4 tag of MP3 files. Structure of ID3 tag is as follows: “ID3 tag header” and “ID3 tag body” which consists of “frames”. Each frame is formed by “frame header” and “frame body”. In order to get information, “frame body” must be read according to the type of the frame and the encoding information as specified in “frame header”.

Because ID3 tag can include very much information, I have implemented parsing of the following data which are stored in the output text file or picture:

* textual information (e.g. title, album, artist),
* lyrics (namely “unsychronized lyrics” as stored in the ID3 tag),
* pictures stored within the ID3 tag.

The parser decides if the MP3 file starts with ID3 header and if so, it checks the version. Program can continue parsing if and only if the latest specification v2.4 of ID3 tag is used (older versions, e.g. v2.3, are not supported). The parser goes through the whole ID3 tag body and reads all “frame headers”. For each frame program prints these header data (including “code”, “size” and “flags”) to the standard output. If frame’s “code” is defined in the internal structure, program will process “frame body” and store data in the internal structure. Otherwise, that “frame body” is skipped. After ID3 is parsed, output files are created – one for textual data and lyrics, and binary files with picture data.

Implementation details

Source code is consisting from 2 files: id3v2parser.h (465 lines including a lot of comments, in order to fully understand parts of ID3 tag to parse it) and id3v2parser.c (645 lines including roughly 100 lines of comments). Code has been written in a modular way in such way that a) every more complex operation is written in distinct method, and b) methods are used in order to not repeat the code.

In order to store parsed textual and picture data there are pre-filled static structures. For instance, when parsing “Title” textual frame, its text is stored in static array id3v2\_textinfo[] of structure id3v2\_frame\_textinfo\_s. After the parsing is over, program goes through these structures and prints data which are filled. From the very large list of supported frames, I have selected only those data which are textual (e.g. lyrics) or picture. This approach is common for ID3 parsers in MP3 players because they interpret what they can (what is implemented) and skip what they do not know (what is not implemented).

Parser is adopting latest specification of ID3 tag in version 2.4. I was hoping to read also 2.3 version (popular e.g. in my MP3 files), however, the latest specification made several changes when treating frames. Hence, I was not able to read “frame headers” of v2.3 properly.

Although parser looks for ID3 tag in front of the file (and nowhere else), specification says that there are two other options where it can be. However, because in most cases it is in the front (and I was not able to get a proper sample file), I have not implemented two other options.

Bug fix

During writing this report I found a bug on 2 lines of code which has had caused wrong filenames of pictures. Unfortunately, this has not been discovered during final testing:

sscanf(id3frame\_apic\_type[i].mime, "%\*[^/],%s", extension);

len = sizeof(orig\_name) + sizeof(id3frame\_apic\_type[i].text) + sizeof(extension) + 3;

This is fix for the bug:

sscanf(id3frame\_apic\_type[i].mime, "%\*[^/]/%s", extension);

len = strlen(orig\_name) + strlen(id3frame\_apic\_type[i].text) + strlen(extension) + 3;

Appendix – results

> ./id3v2parser kabat2.4.mp3

MP3 file length: 6489377

ID3 header:

Version: ID3v2.4.0

Flags: 128

Unsynchronisation: 1,

Extended header: 0,

Experimental indicator: 0,

Footer present: 0

Size: 328017

Frame ID: TALB

Size: 8

Flags: 0

...

Frame ID: PRIV

Size: 28

Flags: 0

Parsed ID3 tag textual frames written into file kabat2.4.mp3.tag.txt

> cat kabat2.4.mp3.tag.txt

Textual information parsed from file kabat2.4.mp3:

Title: Virtuóz

Album: Corrida

Track number: 7

Lead artist: Kabát

Band: Kabát

Composer: Špalek

Content type: Rock

Recording time: 2006

Picture:

image/jpeg

description: Front Cover

picture is stored in file kabat2.4.mp3.cover front.jpeg

Picture:

image/jpeg

description: Performance

picture is stored in file kabat2.4.mp3.during performance.jpeg