

Цифровой звук и MIDI

Звук является наиболее выразительным элементом мультимедиа. Для представления звуковых данных существует два принципиально разных подхода: цифровой звук и MIDI-Musical Instrument Digital Interface.

Данные в формате цифрового звука представляют собой коды уровня звукового сигнала, записанные с заданной частотой дискретизации, которая является основным параметром звукового файла. При использовании цифрового звука, как правило, осуществляется сжатие данных, поскольку данные в формате цифрового звука занимают большой объем памяти. Простейший способ сжатия PCM (Pulse Code Modulation – импульсно-кодовая модуляция), более лучшие результаты имеют усовершенствованные версии – DPCM, ADPCM, а также другие алгоритмы CCITT, GSM, DSP.

MIDI – цифровой интерфейс музыкальных инструментов. Стандарт MIDI разработан в начале 80-х годов для электронных музыкальных инструментов и компьютеров. Он определяет протокол обмена данными между звуковыми синтезаторами различных производителей. Данные в формате MIDI содержат ссылки на ноты, музыкальные инструменты, а также последовательности команд (сообщений), которыми записаны действия музыканта, например, нажатие или отпускание клавиши ли поворот регулятора громкости. Небольшое MIDI-сообщение может вызвать воспроизведение звука или последовательности звуков на музыкальном инструменте или синтезаторе, поэтому MIDI-файлы занимают существенно меньший объем, чем эквивалентные файлы оцифрованного звука.

Как уже указывалось в разделе 1, при оцифровке аналогового (непрерывного во времени и по уровню) звукового сигнала происходит его дискретизация по времени, квантование по уровню и кодирование величин уровней отсчетов. Эти процедуры производятся с помощью аналого – цифровых преобразователей (АЦП), входящих в состав звуковых плат. При упаковке звуковых данных в файловые форматы производится дополнительное кодирование величин уровней отсчетов, при котором возможно сжатие данных. Такие преобразования осуществляются с помощью программ, называемых аудиокодеками. Программы, ориентированные на обработку данных в виде цифрового звука носят название звуковых редакторов. На отечественном рынке наиболее широкое распространение получили звуковые редакторы семейств: Cool Edit (Syntrillium Software, <http://www.syntrillium.com>), Sound Forge (Sonic Foundry, <http://www.sfoundry.com>), Wave Lab (Steinberg SoftWare und HardWare <http://www.steinberg.net>).

В отличие от цифрового звука MIDI данные (сообщения) представляют не информацию о самом звуке, а управляющие команды, генерируемые и выполняемые специальными MIDI устройствами. К числу последних относятся: MIDI - клавиатуры, преобразующие действия музыканта в последовательность команд, а также синтезаторы звуковых карт и звуковые модули, интерпретирующие команды и преобразующие их в обычные

звуковые сигналы. Программы, предназначенные для работы с MIDI данными носят название секвенсоров. На отечественном рынке наиболее широкое распространение получили секвенсоры семейств : Cubase (Steinberg SoftWare und HardWare), Cakewalk (<http://www.cakewalk.com>).

Параметры цифрового звука

Основными параметрами данных в виде цифрового звука являются: частота дискретизации f_d , разрядность оцифровки C , количество каналов (моно/стерео/квадро) k , тип аудиокодека, время звучания t .

Частота дискретизации согласно теореме Котельникова должна быть вдвое выше верхней граничной частоты f_h спектра передаваемого звука и может изменяться в пределах 8 – 44,1 кГц. Формулировка теоремы Котельникова имеет вид: если непрерывная функция времени не содержит в составе своего спектра частот выше f_h , то она полностью определяется последовательностью своих значений, отсчеты которых взяты с частотой $f_d=2f_h$.

Разрядность оцифровки C представляет собой число двоичных разрядов, с помощью которых представляется амплитуда звукового сигнала. Число уровней n квантования связано с разрядностью оцифровки соотношением $n=2^m$. Минимальная разрядность оцифровки определяется типом используемого аудиокодека и составляет 1 – 8 bit, максимальная зависит от используемой звуковой карты и, как правило, не превышает 64 bit. С разрядностью оцифровки тесно связана величина динамического диапазона звукового сигнала D , который пропорционален логарифму отношения сигналов с наибольшей A_{max} и наименьшей A_{min} амплитудами, представимыми заданным числом двоичных разрядов

$$D = 20 \lg \frac{A_{max}}{A_{min}} = 20 \lg \frac{2^C}{2^0} = 20C \lg 2 \approx 6C, \text{ Децибел}$$

Объем памяти, занимаемый цифровым звуком без учета сжатия, которое возможно обеспечивает аудиокодек, может быть определен из соотношения

$$V = kC f_d t.$$

Например, при $k=2$, $C=16$, $f_d=41,1\text{КГц}$ всего лишь одноминутный стереозвук ($t=60\text{ с}$) требует значение $V=8,45\text{Мбит}$ памяти.

Основные типы аудиокодеков

Наибольшее распространение получили кодеки, реализующие следующие алгоритмы кодирования:

– РСМ (Pulse Code Modulation) – импульсно-кодовая модуляция (ИКМ). Наиболее широко распространенный алгоритм, реализующий линейное

кодирование уровня сигнала (величина кода прямо пропорциональна уровню) без сжатия. Различают:

- знаковый 8 bit PCM (коды уровней $-128 + 127$);
- беззнаковый 8 bit PCM (коды уровней $0 - 255$);
- знаковые 16 bit PCM формата LSB (для платформ DOS, OS/2, Windows на основе процессоров Intel, порядок следования байтов в многобайтных данных младший, за ним – старший);
- 16 bit PCM формата MSB (для платформ Macintosh, Amiga на основе процессоров Motorola, порядок следования байтов старший, за ним – младший).

– **СЦИТТ a-Law/u-Law(μ -Law)** – нелинейные PCM. Принцип нелинейного кодирования основан на том факте, что субъективное восприятие громкости звука человеком имеет логарифмическую зависимость от его уровня. Использование логарифмической компрессии позволяет сжимать данные за счет ухудшения качества кодирования отсчетов сигнала высоких уровней. При этом достигается коэффициент сжатия равный двум. Виды кодирования a-Law/u-Law(μ -Law) отличаются функциями отображения линейной шкалы уровней исходного сигнала на шкалы уровней сжатого сигнала (например, исходных 16 bit кодов на сжатые 8 bit) и распространены с одной стороны в Европе (a-Law) и с другой – в США (u-Law(μ -Law)). Функция вида u-Law(μ -Law), производящая логарифмическое сжатие и отображающая исходный уровень сигнала величиной $-1 \leq x \leq 1$ на сжатый величиной $-1 \leq y \leq 1$, имеет вид

$$y(x) = \text{sign}(x) \cdot \frac{\ln(1 + \mu|x|)}{\ln(1 + \mu)}$$

Функция, производящая обратное преобразование экспоненциального экспандирования, записывается в виде

$$x(y) = \frac{\text{sign}(x)}{\mu} [(1 + \mu)^{|y|} - 1]$$

Здесь $0 < \mu \leq 255$ параметр функций отображения, определяющий степень их нелинейности. При обработке уровней сигналов, амплитуда которых отлична от единицы, для шкал несимметричных относительно нуля (например, при использовании беззнаковых кодов), для получения целочисленных результатов, приведенные выше функции отображения должны быть модифицированы путем введения мультипликативных и аддитивных коэффициентов, производящих масштабирование и сдвиг функций, а также путем использования функции выделения целой части. Так, например, функция отображения исходных знаковых 16 bit кодов на сжатые беззнаковые 8 bit представляется в виде

$$y(x) = \text{trunc}\left\{2^7 \left[\text{sign}(x) \cdot \frac{\ln\left(1 + \frac{\mu|x|}{2^{15}}\right)}{\ln(1 + \mu)} + 1\right]\right\}$$

– **ADPCM** (Adaptive Differential PCM) – адаптивные дифференциальные PCM. Данные алгоритмы кодирования основываются на том факте, что последовательно идущие отсчеты сигнала, как правило, по уровням различаются незначительно. Для кодирования последующего отсчета используется значение не собственно его уровня, а его приращение по сравнению с предыдущим. Такой подход позволяет переходить от 8 bit кодов уровня к 1, 2, 4 или 6 bit кодам приращений. Если используемая частота дискретизации будет достаточно высока, то эти приращения, скорее всего, будут небольшими. В результате для хранения одного отсчета понадобится меньше битов.

В адаптивных (самоадаптирующихся) алгоритмах набор допустимых приращений задается не априорно, а определяется на основе кодируемых данных. К числу наиболее распространенных кодеков данного семейства относятся: Microsoft **ADPCM**, IMA **ADPCM**, DSP Group True Speech, GSM 6.10. В IMA **ADPCM**, например, для хранения каждого отсчета используется 4 бита. Кодирующий модуль берет разницу между двумя отсчетами и использует получившуюся величину в качестве очередного сжатого 4-битного выходного значения. Декомпрессор, наоборот, берет это 4-битное значение, умножает его на текущую величину шага и прибавляет получившийся результат к предыдущему отсчету, получая тем самым очередной отсчет. Изменяя размер шага, с помощью 4-битного кода можно записать большой диапазон значений.

– **MPEG** (Motion Picture Expert Group) – семейство алгоритмов, вошедших в стандарт MPEG–1 сжатия аудиовидеопотоков. Раздел, относящийся к аудио, определяет три уровня компрессии (Layer 1 – 3). Каждый последующий уровень позволяет получить большую степень сжатия за счет усложнения алгоритма и времени компрессии. Алгоритмы семейства основаны на спектральных преобразованиях. Являются алгоритмами сжатия с потерями. Ассиметричны, время сжатия существенно больше времени декомпрессии, которая, как правило, проходит в реальном масштабе времени. Примерное соотношение объемов несжатых звуковых данных и сжатых с помощью кодеков MPEG (Layer 1 – 3) составляет: 11: 2,25: 1,25: 1.

–

Классификация и внутренняя структура MIDI сообщений

На рис. 3.1 представлена классификация MIDI сообщений.

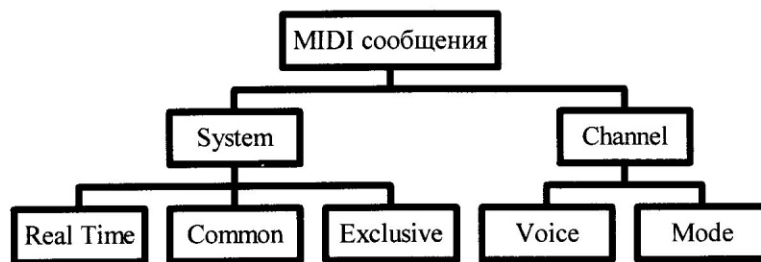


Рис. 3.1. Классификация MIDI сообщений

Различают системные System и каналные Channel сообщения.

Системные сообщения воздействуют на MIDI устройство в целом.

Среди системных сообщений выделяют:

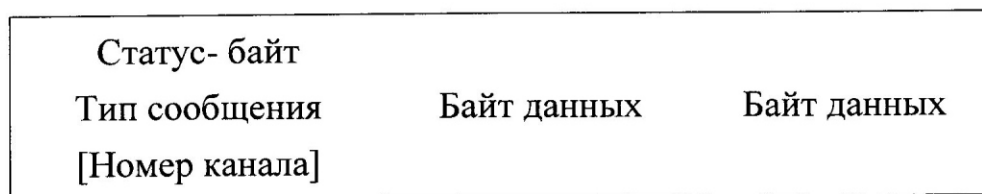
- Real Time – сообщения, синхронизирующие работу MIDI устройств (Start, Stop, Continue, MIDI Clock);
- Common – сообщения, одинаково воспринимаемые всеми устройствами (Song Position Pointer);
- Exclusive – сообщения, индивидуально определяемые производителями для MIDI устройств.

Канальные сообщения содержат номер канала и воздействуют только на соответствующий канал. Каждое MIDI устройство имеет 16 каналов (channels), по которым одновременно могут воспроизводиться голоса (voices), принадлежащие различным инструментам (instruments). Среди канальных сообщений выделяют:

- Mode – сообщения, определяющие режим работы канала (All Notes OFF);
- Voice – голосовые сообщения, несущие информацию о тембре, нотах, громкости и других характеристиках канала (Note ON/OFF).

MIDI сообщения имеют длину 1 – 3 байт, кроме Exclusive, которые могут быть и длиннее.

На рис. 3.2 представлена общая внутренняя структура MIDI сообщений а) и пример структуры голосового сообщения Note ON б).



а)



б)

Рис.3. 2. Общая внутренняя структура MIDI сообщений а) и пример структуры голосового сообщения Note ON б).

Стандарты MIDI

В настоящее время установлены следующие стандарты MIDI:

- **GM** (General MIDI) – стандарт фирмы Roland, регламентирующий набор инструментов в звуковом синтезаторе. В соответствии с этим стандартом синтезатор должен иметь 128 мелодических и 46 ударных инструментов;
- **GS** (General Standard) – расширенный стандарт фирмы Roland. По сравнению с GM имеет дополнительный набор инструментов, а также различные звуковые спецэффекты (скрип, крик, шум и т.п.);
- **GX** (Extended General) – альтернативный стандарту GS стандарт фирмы Yamaha.

Сравнительная характеристика цифрового звука и MIDI

Достоинства цифрового звука:

- позволяет передавать любые звуковые данные, в том числе и речь;
- при редактировании не требуют специальных музыкальных знаний;
- качество воспроизведения более аппаратно независимо;
- поддерживается более широким кругом программного обеспечения.

Достоинства MIDI:

- большая компактность (на 2 – 3 порядка);
- допускают изменение объема файлов при сохранении качества и громкости, но с изменением темпа;
- легко редактируются на уровне отдельных нот, что соответствует сегментам композиции длительностью до миллисекунд.

Аналогия в представлении графических и звуковых данных

В представлении графических и звуковых данных существует глубокая аналогия, которая иллюстрируется в табл.1. При этом растровым графическим данным соответствует цифровой звук, а векторным – MIDI..

Таблица 1. Аналогия в представлении графических и звуковых данных

Вид представления данных	Растровый	Цифровой звук	Векторный	MIDI
Примитив описания	Пиксел	Звуковой квант (дискретизованный и квантованный отсчет уровня сигнала)	Векторный графический примитив	Нота
Атрибуты	Цвет,	Громкость	Геометрически	Тембр,

примитива	прозрачно сть	(величина уровня)	е и негеометричес кие атрибуты	длительнос ть
-----------	------------------	----------------------	--------------------------------------	------------------

Основные файловые форматы звуковых данных

Форматы цифрового звука:

- **WAV (Microsoft Waveform Audio)** – звуковая версия общего файлового формата обмена ресурсами **riff** (Resource Interchange File Format). Организован как набор вложенных блоков данных, для идентификации содержимого которых используются дескрипторы. Поддерживает около сотни аудиокодеков. Наиболее распространенный формат цифрового звука на платформе Windows, по степени распространенности занимает такое же положение как **bmp** для представления графических данных и **avi** для видео. Поддерживается всеми звуковыми редакторами и плеерами;
- **IFF (Interchange File Format), AIFF (Audio Interchange File Format), AIFC (Audio Interchange File Format Extention for Compression)** – аналоги **riff** формата для платформ Amiga и Macintosh соответственно. В настоящее время доступны на платформе Windows;
- **MP3 (MPEG – 1, Layer 3)** – формат, использующий соответствующий кодек. Благодаря своей компактности получил весьма широкое распространение;
- **AU, SND (Audio Sun, Sound NeXT)** – форматы цифрового звука для компьютеров Sun и NeXT соответственно, получили широкое распространение в Internet;
- **RA (Real Audio)** - формат, разработанный фирмой RealNetworks для передачи данных в Internet в реальном масштабе времени;
- **VOC (Creative Labs Voice)** – формат, разработанный фирмой Creative Labs для демонстрации возможностей звуковых карт. Отличается относительной простотой внутренней структуры. Использует кодеки: 8, 16 bit PCM, a-Law/u-Law(μ -Law), 4 bit ADPCM.

Среди форматов, используемых для хранения звуковых данных, отличных от цифрового звука, помимо формата **MID** (MIDI), следует выделить:

- **MOD (Protracted Modules)** – формат, использующий представление, занимающее промежуточное положение между MIDI и цифровым звуком. Файлы данного формата начинаются с шаблонов (образцов цифрового звука инструментов) за которыми следует информация о нотах и их атрибутах. В отличие от MIDI полностью определяет звук. В отличие от несжатых форматов цифрового звука существенно компактнее (ориентировочно на 2 порядка);
- **SKD, SKP (SSEYO Koan Design, SSEYO Koan Play)** – форматы Коун-музыки, разработанные SSEYO Ltd. (<http://www.sseyo.com>). Отличаются исключительной компактностью – файлы объемом в несколько килобайт могут содержать данные с длительностью звучания в несколько часов. В

файле указываются основные параметры музыки, а она сама генерируется программой, причем процесс генерации носит элементы случайности, т.е. из одного и того же файла невозможно дважды воспроизвести идентичную музыку. Формат **SKP** представляет собой версию с защитой от редактирования;

– **CDA (Compact Disk Audio)** – формат представления звуковых данных, используемый в плеерах аудио компакт-дисков и поддерживаемый на компьютерных платформах.

Внутренняя структура файлов цифрового звука в формате VOC

Рассмотрим детально внутреннюю структуру файлов цифрового звука в формате **VOC**. Файлы данного формата включают заголовок длиной 26 байт и блоки данных. Порядок следования байтов в многобайтных данных младший, затем – старший, принятый для процессоров семейства Intel, используемых на платформе **Wintel**. Далее рассматриваются значения данных с использованием шестнадцатеричных (HEX) кодов. Для числовых значений данных, при необходимости, основание используемой системы счисления указывается в виде индекса.

Структура заголовка файла формата voc приведена в табл. 2.

Таблица 2. Структура заголовка файла формата voc

Длина данных, байт	Описание данных
19	Строковый идентификатор: Creative Voice File
1	Байт со значением $1A_{16}$ (26_{10}): Ctrl+Z – символ конца файла в операционной системе DOS
2	Байты со значением $1A_{16}$ и 00_{16} (26_{10}) – длина заголовка в байтах
2	Код версии формата. Как правило, значение равно 266_{10} (версия 1.10) или 176_{10} (версия 1.20)
2	Контрольное значение кода версии: 4659_{10} – код версии

В табл. 3 приведен пример шестнадцатеричных кодов первых 32 байтов файла монофонического цифрового звука в формате **VOC**, содержащие описание заголовка файла (байты 1 – 26) и заголовка первого блока данных (байты 27 – 32).

Таблица 3. Пример HEX кодов первых 32 байтов файла монофонического цифрового звука в формате **VOC**

43	72	65	61	74	69	76	65	20	56	6	69	6	65	20	46
										F		3			
69	6	65	1	1	00	0	01	29	11	01	C	3	00	D	00
	C		A	A		A					D	6		3	

Значение кода версии составляет: $010A_{16}=266_{10}$ (версия 1.10), контрольное значение кода версии: $4659_{10} - 1129_{16}=266_{10}$.

Блоки данных состоят из четырехбайтного заголовка, первый байт которого определяет тип блока, а оставшиеся три байта - длину данных и собственно данные. Исключение составляет блок признака окончания, который содержит единственный нулевой байт. Определены следующие типы блоков (значения типов указаны в скобках):

- звуковых данных (1);
- продолжения данных (2);
- паузы (3);
- маркерный (4);
- текстовый (5);
- повторения (6,7);
- расширения (8,9).

Блок звуковых данных (тип1) начинается с двух байт, определяющих код частоты дискретизации и используемый аудиокодек, после которых следуют собственно коды уровней отсчетов звукового сигнала. Частота дискретизации определяется из соотношения

$$f_d = \frac{10^6}{2^8 - \text{код частоты}} \quad (6)$$

Коды аудиокодеков, поддерживаемых форматом **VOC**, приведены в табл. 4.

Таблица 4. Коды аудиокодеков, поддерживаемых форматом **VOC**

Код	Аудиокодек
0	Беззнаковый 8 bit PCM
1	4 bit ADPCM
2	6 bit ADPCM
3	2 bit ADPCM
4	16 bit PCM
6	CCITT a-Law
7	CCITT u-Law(μ-Law)

В примере, приведенном в табл. 3, после заголовка файла непосредственно следует блок звуковых данных длиной $36CD_{16}=14$ кбайт с

частотой дискретизации $10^6/(2^8 - D3_{16})=22,2$ КГц, представленных с использованием беззнакового 8 bit PCM.

Файл формата **VOX** может иметь только один блок звуковых данных.

Блок продолжения данных (тип 2) используется в том случае, если возникает необходимость вставки в середину звуковых данных блока другого типа (например, блока паузы). Оставшаяся часть звуковых данных сохраняется в блоке продолжения. Блок продолжения в качестве данных содержит только коды уровней отсчетов звукового сигнала.

Блок паузы (тип 3) определяет период тишины. Данные такого блока занимают три байта, первые два из которых определяют количество отсчетов тишины, а третий – используемую в блоке частоту дискретизации.

Маркерный блок (тип 4) может использоваться мультимедийными приложениями для синхронизации событий. Например, событие входа в маркерный блок может вызывать смену изображения, запуск фильма или другого приложения. Данные маркерного блока представляют собой двухбайтное число, значения 0000 и FFFF зарезервированы и не должны использоваться.

Текстовый блок (тип 5) предназначен для сохранения внутри файла текстовой информации. Представляет собой строку в ASCII кодах с нулевым символом на конце.

Блоки повторения (типы 6, 7) используются для организации циклов внутри файла. Блок типа 6 помещается в начале области, которую необходимо повторить. Данные блока представляют собой двухбайтное число, показывающее количество повторений. Блок типа 7 помещается в конец области повторения и не содержит ассоциированных данных. Возможна организация вложенных циклов.

Блок расширения (тип 8) используется для поддержки файлов со стереозвуком. Если этот блок присутствует, он предшествует блоку звуковых данных (тип 1) и переопределяет информацию о частоте дискретизации и типе используемого аудиокодека. Данные блока содержат четыре байта, первые два из которых определяют код точного значения частоты дискретизации, а оставшиеся два – код аудиокодека и количество каналов (0 – моно, 1 – стерео) соответственно. Точное значение частоты дискретизации определяется из соотношения

$$f_d = \frac{2^8 \cdot 10^6}{2^{16} - \text{код частоты}}. \quad (6)$$

Для файлов стереозвука величину частоты дискретизации, получаемую из соотношения (6) следует уменьшить вдвое, поскольку коды уровней отсчетов звукового сигнала для левого и правого каналов в блоках типов 1 и 2 при этом чередуются и следуют с уменьшенной вдвое частотой.

В табл. 5 приведен пример HEX кодов первых 40 байтов файла стереофонического цифрового звука в формате vox, содержащие описание заголовка файла (байты 1 – 26), блока расширения (байты 27 – 34), и заголовка блока звуковых данных (байты 35 – 40).

Таблица 5. Пример HEX кодов первых 40 байтов файла стереофонического цифрового

звука в формате **VOC**

43	72	65	61	74	69	76	65	20	56	6	69	63	65	20	46
										F					
69	6	65	1	1	00	0	01	29	11	08	04	00	00	54	E
	C		A	A		A									9
00	01	01	02	40	00	E	00								
						9									

Точное значение частоты дискретизации, определенное в блоке расширения, согласно соотношению (7) составляет $2^8 \cdot 10^6 / (2^{16} - E954_{16}) = 44,1$ КГц. Для сравнения значение, определенное в блоке звуковых данных, согласно соотношению (6) составляет $10^6 / (2^8 - E9_{16}) = 43,5$ КГц. Данная величина является приближенной и полученной с использованием однобайтного кода $E9_{16}$ (байт 39), значение которого совпадает со значением старшего байта двухбайтного кода $E954_{16}$ (байты 31, 32), использованного при расчете точного значения величины частоты дискретизации. Поскольку данные описывают стереозвук, о чем свидетельствует единичное значение байта 34, реальная частота дискретизации будет вдвое меньше полученной из соотношения (7) и составит 22.05 КГц.

Блок расширения (тип 9) позволяет сохранять информацию о звуке в явном виде и вводить новые двухбайтные коды аудиокодеков. Если этот блок присутствует, он предшествует блокам типов 1 и 8 переопределяет содержащуюся в них информацию. Данные блока расширения типа 9 приведены в табл. 6.

Таблица 6. Данные блока расширения типа 9 файла формата **VOC**

Длина данных, байт	Описание данных
4	Частота дискретизации, Гц
1	Разрядность оцифровки, бит
1	Количество каналов (1 – моно, 2 – стерео)
2	<u>Код аудиокодека</u>
4	Зарезервированные байты

Значение частоты дискретизации не кодируется, а сохраняется непосредственно и не требует коррекции по числу каналов.

Приведенная выше спецификация формата **VOC** является полной и позволяет разрабатывать программы-разборщики произвольных файлов данного формата.