

По результатам обсуждения 8 апреля в ин-те Склифасовского под председ. Сумского ЛИ., и последовавшим за этим обмена мнениями, формулирую в окончательном виде формат UDF. Отличия и нововведения выделены красным.

Формат данных UDF (Universal Data Format).

А.В. Пирожено, МБН, Москва.

При разработке этого формата я придерживался следующих требований:

1. Совместимость с EDF (European Data Format).
2. Возможность хранения ЭЭГ, ВП и полиграфических данных.
3. Хранение дополнительной информации о пациенте и регистрации.

Вся информация хранится в одном файле с расширением **edf**. Файл данных состоит из блока параметров EDF, блока дополнительных параметров и последовательности блоков данных.

1. Блок параметров EDF.

Блок имеет следующий формат:

8 ascii : версия EDF (0)
80 ascii : краткая информация о пациенте
80 ascii : краткая информация о регистрации
8 ascii : дата начала регистрации (dd.mm.yy)
8 ascii : время начала регистрации (hh.mm.ss)
8 ascii : общий размер заголовка, включая блок EDF и дополнительных параметров
44 ascii : зарезервировано
8 ascii : число блоков данных (-1 если не известно)
8 ascii : длина одного блока данных, в секундах
4 ascii : число сигналов (ns) в блоке данных
Итого: 256 байт
ns * 16 ascii: ns * типы и названия отведений (например *EEG Fpz-A1* или *EOG REOG*)
ns * 80 ascii: ns * типы электродов (например *AgAgCl electrode*)
ns * 8 ascii: ns * физическая размерность (например *uV* или *%*)
ns * 8 ascii: ns * физический минимум (например *-500* или *34*)
ns * 8 ascii: ns * физический максимум (например *500* или *40*)
ns * 8 ascii: ns * цифровой минимум (например *-2048*)
ns * 8 ascii: ns * цифровой максимум (например *2047*)
ns * 80 ascii: ns * параметры фильтров
ns * 8 ascii: ns * пг отсчетов в каждом блоке данных
ns * 32 ascii: ns * зарезервировано
Итого: ns * 256 байт

Комментарии.

Информация в ASCII строках должна начинаться с левой позиции и дополняться справа пробелами. Для совместимости с имеющимися западными программами все символы должны быть латинскими. Русские буквы транслитерируются в латинские в соответствии с Приложением 1.

Рекомендуется, чтобы длительность блока данных была целым значением. Имеются два исключения:

1. По рекомендации EDF размер одного блока данных не должен превышать 61440 байт.
2. В случае ВП длительность эпохи может быть менее 1 секунды.

В этих случаях нужно задавать меньшую длину блока (например *0.01*).

В качестве версии EDF следует задавать 0.

Значение 'общий размер заголовка' указывает отступ первого блока данных от начала файла. Наличие этого параметра обеспечивает совместимость формата UDF с EDF.

Значения цифровых минимума и максимума (d0 и d1) каждого сигнала должны соответствовать предельным значениям, которые могут встречаться в блоках данных. Физические максимум и минимум (a0 и a1) должны соответствовать этим цифровым значениям.

Физическое значение a, соответствующее цифровому значению d может быть рассчитано по формуле

$$a = a0 + (a1-a0) * (d-d0) / (d1-d0). \quad (1)$$

Дополнительные рекомендации UDF:

1. Везде, где возможно, задавать длину блока данных 1 секунда. Большинство имеющихся программ понимают только это значение.
2. В информации о пациенте указывать ФИО и возраст, например *Ivanov Ivan Ivanovich, 32 years old*.
3. В информации о регистрации указывать систему, на которой были записаны данные, и тип усилителя, например *Neurokartograph 4.32, MBN-20, (C) MBN, Moscow*.
4. Нет причин указывать в качестве количества блоков данных -1. Однако некоторые имеющиеся программы это делают и при чтении файла это нужно учитывать.
5. Типы и названия отведений. Тип и название канала разделяются пробелом. Стандартизированы следующие типы:
EEG – электроэнцефалограмма,
EOG – электроокулограмма,
EMG – электромиограмма,
ECG – электрокардиограмма,
Respir – дыхание,
CPAP – давление CPAP,
Pres – давление,
Micro – микрофон,
HF – частота сердечных сокращений,
SaO2 – насыщение кислородом,

Pos – положение тела,
GSR – кожно-гальваническая реакция,
PPG – фотоплетизмограмма без SaO₂,
EXT – неопределенный.

Возможно наличие служебных каналов ЭЭГ (например в усилителе МБН-20 все каналы записываются относительно A1, дополнительно регистрируется A2-A1; в усилителях Neurofax и Schwarzet все каналы записываются относительно виртуальной земли Z и дополнительно записываются A1-Z, A2-Z, E-Z). Для таких каналов также указывается тип EEG.

Для ЭЭГ каналов указывается название канала и референт через дефис. Хотя в стандарте EDF это прямо не указано, это соглашение, похоже, стало негласной традицией для программ, работающих с EDF. Рекомендуется записывать данные в файл так, как они считаны с энцефалографа, без перекоммутации, а регистрировать рекомендуется с общим референтом. Т.е. у всех ЭЭГ и служебных каналов после дефиса должен быть один и тот же референт.

Список должен начинаться с активных каналов ЭЭГ. Служебные каналы ЭЭГ должны следовать после активных ЭЭГ каналов. Затем следуют полиграфические каналы. У полиграфических каналов референт может не указываться, в этом случае дефис тоже не пишется.

6. Параметры фильтров рекомендуется указывать в следующем формате:
HP:0.03Hz LP:75Hz NF:off
HP:0.1Hz LP:150Hz NF:on

2. Блок дополнительных параметров UDF.

Блок имеет следующий формат:

```
4 ascii : идентификатор формата: UDF
4 ascii : версия: 1.1
64 ascii : имя базы данных, к которой отнесена эта регистрация (напр. Эпилептики)
32 ascii : фамилия пациента
32 ascii : имя отчество пациента
16 ascii : дата рождения (dd.mm.yyyy)
2 ascii : пол (M – мужской, F – женский; буквы латинские)
2 ascii : тип лаборатории (A - амбулатория, S – стационар, H – на дому и т.д.)
16 ascii : номер медицинской карты (идентификатор пациента)
256 ascii: диагноз (причина обследования)
16 ascii : номер (идентификатор) регистрации, например 97-00129
8 ascii : тип обследования

Далее следуют дополнительные технические параметры.
8 ascii : индифферентный электрод
8 ascii : земляной электрод (допускается заполнить это поле пробелами)
ns * int : координаты электродов по оси X (см. комментарии ниже)
ns * int : координаты электродов по оси Y
ns * int : координаты электродов по оси Z
ns * float: импеданс в кОм
ns * float: ФВЧ (HP) фильтр в Гц
ns * float: ФНЧ (LP) фильтр в Гц
ns * int : режкторный (notch) фильтр в Гц: 50, 60, 0 – выключен
int : базовая частота опроса bFreq (обычно 256 или 200). Частоты опроса каналов должны быть целыми делителями этой базовой частоты. Положение маркеров указывать в единицах количества отсчетов с частотой bFreq от начала записи до маркера
int : количество маркеров функциональных проб (ФП) и комментариев nMark
nMark * dword: положение маркеров
nMark * int: тип маркера (см. комментарии ниже)
nMark * 64 ascii: текст маркера
int : количество отметок стимулятора nStim
nStim * dword: массив положений отметок стимулятора

Далее следуют параметры экранного отображения сигналов:
float : масштаб по горизонтали в мм/сек
int : количество отведений на экране nRefs
nRefs * int: nRefs * индекс активного электрода в списке отведений
nRefs * int: nRefs * индекс пассивного электрода в списке отведений
nRefs * int: nRefs * цвет графика
nRefs * int: nRefs * полярность графика: 0 – позитивность вверх, 1 – позитивность вниз
nRefs * int: nRefs * тип масштабных коэффициентов val0 и val1, записанных далее. См. комментарии
nRefs * float: nRefs * val0 (см. комментарии)
nRefs * float: nRefs * val1 (см. комментарии)

Далее следует текстовое заключение (описание)
4 ascii : формат заключения: TXT – DOS текст, RTF – rich text format, DOC – MSWord
dword : размер блока с заключением в байтах cLeng (при отсутствии заключения задается 0)
cLeng * byte: заключение
```

Далее следует блок, специфический для программы.

```
16 ascii : идентификатор программы (для ЭЭГ программы МБН - "NKWIN")
? : блок параметров произвольного формата
```

Комментарии.

Следует иметь в виду, что европейские фирмы могут использовать свой формат блока дополнительных параметров. Наличие блока в формате UDF определяется по его идентификатору в начале блока. Блок неизвестного формата нужно игнорировать и использовать параметры по умолчанию.

Блок параметров UDF имеет неопределенный размер. Конец блока параметров и начало данных можно определить по значению 'размер заголовка' в начале блока параметров EDF.

Информация в ASCII строках имеет тот же формат, что в EDF: начинается с левой позиции и дополняется справа пробелами. Используется DOS кодировка русских букв, приведенная в Приложении 2, чтобы максимально упростить работу DOS программ.

Формат бинарных данных соответствует архитектуре Intel процессоров серии 86/87:
int – двухбайтное знаковое целое (Pascal – integer, C – short int, Assembler - word);
dword – четырехбайтное целое (Pascal – longint, C – long, Assembler - dword);
float – четырехбайтное с плавающей точкой (Pascal – single, C – float, Assembler - dword)

Дата рождения указывается в формате dd.mm.yyyy, например 25.07.1962. Если программа знает только возраст пациента, следует указывать в качестве даты 1 января. Если нет никакой информации, поле заполняется пробелами.

Некоторые программы нуждаются в информации о типе обследования, чтобы предложить пользователю соответствующие функции обработки. Предлагается стандартизировать следующие типы:

EEG – ЭЭГ обследование;
EMG – ЭМГ обследование;
REG – реографическое обследование;
POLY – полиграфическое обследование;
EP – вызванные потенциалы;
P300 – когнитивные ВП

Данный формат предназначен для записи ЭЭГ и полиграфических обследований, поэтому тип обследования должен быть EEG или POLY. Остальные типы резервируются на будущее; вероятно, для этих типов формат будет отличаться, поэтому для совместимости с последующими версиями желательно проверять значение этого поля при чтении данных.

Координаты электродов указываются для ЭЭГ электродов. Для других электродов могут указываться произвольные значения (0). Координатная система описана в Приложении 3.

Определены следующие типы маркеров:

0 – текстовый комментарий,
1 – разрыв записи,
2 – изменение параметров усилителя,
3/4 – включение/выключение калибровочного сигнала,
5 – измерение импеданса,
10/11 – начало/конец фоновой записи,
12/13 – начало/конец ФП 'открытие глаз',
14/15 – начало/конец ФП 'гипервентиляция',
16/17 – начало/конец ФП 'фотостимуляция',
18/19 – начало/конец ФП 'фоностимуляция',
20/21 – начало/конец сна (time in bed).
99/100 – начало/конец неопределенной ФП

Классификация маркеров может потребоваться для автоматизированного анализа. Программы, не имеющие информации о типе данного маркера, указывают тип 0.

Поля “активный электрод” и “пассивный электрод” предназначены для описания монтажа отведений при выводе на экран.

Указывается номер электрода в списке электродов или одно из следующих значений:

252 – A1 (левое ухо),
253 – A2 (правое ухо),
254 – AA (объединенные уши),
255 – Av (усредненный электрод),
256 – Av1 (усредненный по левому полушарию),
257 – Av2 (усредненный по правому полушарию),
258 – Sd (source deviation – отведение от источника),
259 – M1 (левый мостоид),
260 – M2 (правый мостоид),
261 – MM (объединенные мостоиды),
262 – E (земля),
263 – неопределенный (для пассивного электрода: не вычитать ничего из значений активного)

Масштабные коэффициенты могут быть двух типов:

тип 0 – канал постоянного тока (dc), например Pres, HF, SaO2, Pos. В этом случае val0 задает минимальное, val1 – максимальное значение в физических единицах в той полосе экрана, в которой отображается график. Например, для SaO2 пользователь может задать val0 = 70 %, val1 = 100 %.

тип 1 – канал переменного тока (ac), например EEG, EOG, ECK, EMG и т.д. В этом случае val0 задает сдвиг относительно нулевой линии в физических единицах, val1 – масштаб в физических единицах на 1 мм. Например для ЭЭГ обычно задают val0=0 uB, val1=7 uB/мм.

Если заключение (описание) имеет текстовый вид, русские буквы должны быть преобразованы в кодировку DOS (Приложение 2). Другие форматы не преобразуются. При считывании заключения в формате RTF или DOC программа может записать его в отдельный файл с соответствующим расширением.

Блок, специфический для программы, имеет произвольный формат, и предназначен для того, чтобы каждая программа могла хранить в файле дополнительные параметры, которые могут ей пригодиться при реимпорте данных. Каждый производитель может выбрать произвольный идентификатор, но не короче 4 символов. Желательно сообщить этот идентификатор мне.

3. Данные UDF.

Последовательные блоки содержат двоичные 16-битные значения в формате, соответствующем “Модифицированному EDF” Nielsen 1988:

signal1 : nrsamples[1] * integer;
signal2 : nrsamples[2] * integer;
signal3 : nrsamples[3] * integer;
...и т.д.

Цифровые значения преобразуются в физические в соответствии с формулой (1).

Приложение 1. Преобразование русских букв в латинские.

Для того чтобы обеспечить возможность обратного преобразования, которое может понадобиться в дальнейшем, предлагается следующая таблица преобразования:

а	a	и	i	р	r	ш	sh
б	b	й	y	с	s	щ	sh
в	v	к	k	т	t	ъ	'
г	g	л	l	у	u	ы	yi
д	d	м	m	ф	f	ь	'
е, ё	e	н	n	х	h	э	e
ж	j	о	o	ц	c	ю	yu
з	z	п	p	ч	ch	я	ya

Например: Шалыгин -> Shalyigin.

Приложение 2. Кодировка русских букв.

Для совместимости с DOS программами нужно использовать стандартную MS DOS кодировку русских букв:

А	128	И	136	Р	144	Ш	152	а	160	и	168	р	224	ш	232
Б	129	Й	137	С	145	Щ	153	б	161	й	169	с	225	щ	233
В	130	К	138	Т	146	Ъ	154	в	162	к	170	т	226	ъ	234
Г	131	Л	139	У	147	Ы	155	г	163	л	171	у	227	ы	235
Д	132	М	140	Ф	148	Ь	156	д	164	м	172	ф	228	ь	236
Е	133	Н	141	Х	149	Э	157	е	165	н	173	х	229	э	237
Ж	134	О	142	Ц	150	Ю	158	ж	166	о	174	ц	230	ю	238
З	135	П	143	Ч	151	Я	159	з	167	п	175	ч	231	я	239

Использование перекодировки в формат DOS обязательно для всех, т.к. даже Windows программа, ожидающая DOS кодировку, не сможет корректно прочитать строки в Windows кодировке.

Приложение 3. Координаты электродов.

Координаты электродов задаются в трехмерном виде. Ось X направлена от левого уха к правому. Ось Y направлена от затылочного бугра (inion) к переносице (nasion). Ось Z направлена от базиса к вертексу. Значения указываются в миллиметрах. Электроды, установленные по системе 10-20%, для некоего 'среднего' пациента (данные Ю.М.Коптелова) имеют в этой системе следующие координаты:

Название	X	Y	Z
Fp1	-27	94	34
Fp2	27	94	34
F3	-45	60	80
F4	45	60	80
C3	-55	0	96
C4	55	0	96
P3	-45	-60	80
P4	45	60	80
O1	-27	-95	34
O2	27	-95	34
F7	-68	54	34
F8	68	54	34
T3	-80	0	34
T4	80	0	34
T5	-67	-55	34
T6	67	-55	34
Fpz	0	99	34
Fz	0	66	97
Cz	0	0	120
Pz	0	-66	98
Oz	0	-100	34
A1	-70	0	-15
A2	70	0	-15

А.В. Пироженко
07 июня 1998.