

Описание работы парсера протоколов

Версия 1

(от 05.09.2020)

Каждый протокол делится на вкладки (или листы), соответствующие листам (sheet) в эксель-файле протокола. При этом название листа в эксель является названием вкладки протокола. Делить на вкладки можно по любым принципам, например:

- по дивизионам (например, WPC PL Equipped, WPC BP Unequipped, AWPC PL Raw и т.д.)
- по половому признаку (Men, Women)
- по возрастным категориям (Masters 1, Masters 2, Masters 3 и т.д.)
- или комбинация: например Women Masters 1, Men Masters 3 и т.д.

По названию вкладки парсер пытается определить вид спорта по вхождению строки.

Соответствие следующее:

- Пауэрлифтинг: 'пл', 'pl', 'пауэрлифтинг', 'powerlifting'
- Жим лежа: 'жим', 'bp', 'benchpress'
- Становая тяга: 'тяга', 'dl', 'deadlift'
- Приседание со штангой: 'присед', 'присяд', 'sq', 'squat'

Если нет ни одного вхождения, то вид соревнования определяется по найденным результатам (по колонкам). Если же и по колонкам не удалось определить вид, то вид спорта указывается как "пауэрлифтинг" (PL).

Если в вашем протоколе по логике получается всего одна вкладка, то можете задать ей произвольное имя.

Парсер протоколов пытается автоматически найти все нужные данные на каждом листе эксель файла, независимо от формата. На каждом листе парсер ищет строку с названиями колонок. Колонки могут идти в произвольном порядке в любом месте листа. При этом обязательными колонками являются **name** и **birth**. Остальные колонки не являются обязательными и если не будут найдены в форму загрузки протокола попадут пустые значения.

Основные колонки:

- **Name** - имя спортсмена (обязательно)
Ключевые слова для поиска колонки: 'name', 'фамилия', 'имя', 'фио'

Значение: Может содержать занятое спортсменом место вместе с именем спортсмена. В этом случае место будет извлечено отдельно (описано ниже)

Пример значений:

Name
- Polyakova Valeriya
1. Kukharchuk Vera
2. Sarycheva Yekaterina
3. Vasilenko Mariya
1. Safronova Svetlana

- **Birth** - дата рождения спортсмена (обязательно)

Ключевые слова для поиска колонки: 'год', 'рожд', 'bith', 'birth'

Значение: ячейка со значением может содержать дополнительные данные, главное чтобы в строке содержалась дата или год рождения в одном из следующих форматов:

- 'dd-mm-yyyy', 'dd.mm.yyyy' или 'dd/mm/yyyy' (например: '23-05-1982', '23.05.1982' или '23/05/1982')
- 'yyyy-mm-dd', 'yyyy.mm.dd' или 'yyyy/mm/dd' (например: '1982-05-23', '1982.05.23' или '1982/05/23')
- или просто год рождения 'yyyy' (например '1982')

Пример значений:

Age Class Bith date/Age
Juniors 20-23 (07.08.1997)/21
Open (28.12.1989)/29
Open (19.09.1992)/26
Open (28.07.1984)/34
Masters 45-49 (25.04.1972)/46

- **Place** - место, которое занял спортсмен (необязательно)

Ключевые слова для поиска колонки: '#', '№'

Значение: числовое значение. Если колонка с местом не найдена, парсер ищет место в ячейке с именем спортсмена (**name**). Например для значения: '1. Деточкин Роман' будет распознано первое число '1' как занятое место **place**

- **Gender** - пол спортсмена (необязательно)

Ключевые слова для поиска колонки: 'gender', 'sex', 'пол'

Значение: ячейка со значением может иметь одно из следующих значений:

- Мужской пол - 'муж', 'м', 'male', 'm'
- Женский пол - 'жен', 'ж', 'female', 'f'

Если значение в ячейке не соответствует ни одному значению из указанных выше, то пол остается не указанным для спортсмена.

- **Age** - возрастная категория спортсмена (необязательно)

Ключевые слова для поиска: 'возрастная группа', 'возрастная категория', 'age category', 'age class'

Значение: Допустимы следующие значения:

- С - "Дети (<12)"
 - С1 - "Дети (10-11)"
 - С2 - "Дети (12-13)"
 - С3 - "Дети (8-10)"
 - Т - "Юноши/Девушки"
 - Т0 - "Подростки (subteen)"
 - Т1 - "Юноши/Девушки 1"
 - Т2 - "Юноши/Девушки 2"
 - Т3 - "Юноши/Девушки 3"
 - J - "Юниоры"
 - О - "Открытый класс"
 - MS - "Суб-ветераны"
 - M30-34 - "Суб-ветераны 1"
 - M35-39 - "Суб-ветераны 2"
 - M - "Ветераны"
 - M1 - "Ветераны 1"
 - M2 - "Ветераны 2"
 - M3 - "Ветераны 3"
 - M4 - "Ветераны 4"
 - M5 - "Ветераны 5"
 - M6 - "Ветераны 6"
 - M7 - "Ветераны 7"
 - M8 - "Ветераны 8"
 - M9 - "Ветераны 9"
 - M10 - "Ветераны 10"
 - PFM - "Пожарные/Милиция/Военные"
 - S - "Студенты"
 - DS - "Инвалиды"
 -
- **Weight** - вес спортсмена (необязательно)
Ключевые слова для поиска колонки: 'вес', 'weight'
Значение: должно быть указано число.
 - **WeightCategory** - весовая категория спортсмена (необязательно)
Ключевые слова для поиска колонки: 'в/к', 'в.к.', 'weight category'
Значение: должно быть указано число. Если колонка не найдена, парсер пробует найти ячейки с содержимым: 'вес' + 'кат' или 'body' + 'weight' и извлекает весовую категорию оттуда. Все спортсмены указанные ниже строчки с указанием весовой категории, попадают в эту весовую категорию.
Пример:

№ п/п	Фамилия, имя	Год рожд.	Звание, разряд	Область город	Собств. вес	Прис.	Жим	Тяга	Сумма Разряд
Мужчины				Весовая категория до 52,0 кг					
Для всех спортсменов в этой области будет указана весовая категория до 52 кг.									
				Весовая категория до 56,0 кг					
Для всех спортсменов в этой области будет указана весовая категория до 56 кг.									
				Весовая категория до 60,0 кг					
Для всех спортсменов в этой области будет указана весовая категория до 60 кг.									

- Location** - местоположение спортсмена (необязательно)
Ключевые слова для поиска колонки: 'город', 'область', 'town', 'region', 'city'
Значение: любое строковое значение. Если значение ячейки полностью соответствует названию местоположения имеющегося в базе данных - то парсер использует его. Если же местоположение найдено не будет, вам придется выбрать значение из списка в форме протокола вручную.

№ п/п	Фамилия, имя	Год рожд.	Звание, разряд	Область город	Собств. вес	Прис.	Жим	Тяга	Сумма Разряд	Тренер очки
Мужчины				Весовая категория до 52,0 кг						
1	Шатаков Юрий	1976	МС	Ленинск-Кузнецкий	51,90	165,0	90,0	185,0	440,0	МС Кротов И.М. муж
2	Королев Денис	1979	КМС	Гурьевск	51,90	150,0	75,0	180,0	405,0	КМС Татаров Р.А. муж
3	Силаев Дмитрий	1979	КМС	Гурьевск	52,00	142,5	75,0	147,5	365,0	КМС Бокарев О.В. муж
4	Фельгрон Михаил	1979	1ю.	Л-Кузнецкий	50,70	125,0	57,5	150,0	332,5	б/р Докунаев А.Н. муж

#	Спортсмен	Пол	Д.Р.	Город	Вес	ВК	М
1	Shatakov Yuriy	♂	1976	Ленинск-Кузнецкий	51.9	-52	1
2	Korolev Denis	♂	1979	Гурьевск	51.9	-52	2
3	Silaev Dmitriy	♂	1979	Гурьевск	52	-52	3
4	Felgrom Mikhail	♂	1979	Л-Кузнецкий	50.7	-52	4

			Спортсмен				
-	#	Спортсмен	Пол	Д.Р.	Город		
x	1	Shatakov Yuriy	x	♂	1976	Ленинск-Кузнецкий	x 5
x	2	Korolev Denis	x	♂	1979	Гурьевск	x 5
x	3	Silaev Dmitriy	x	♂	1979	Гурьевск	x 5
x	4	Felgron Mikhail	x	♂	1979	Ленинск	x 5
x	5	Sedov Andrey	x	♂	1977	Россия > Кемеровская область > Ленинск-Кузнецкий (1405)	5
x	6	Abzalov Jurij	x	♂	1979	Казахстан > Ленинск (1541)	5
x	7	Proskurin Denis	x	♂	05.09.1979	Россия > Волгоградская область > Ленинск (2357)	5
x	8	Nezusev Denis	x	♂	1980	Россия > Иркутская область > Ленинск (6702)	5
x	9	Nabiev Vakhid	x	♂	1968	Антигуа и Барбуда > Ленинск-Кузнецкий (17778)	6
x	10	Krivushkin Dmitrij	x	♂	1977	Польсаево	x 5
x	11	Novgorodov Sergej	x	♂	1978	Ленинск-Кузнецкий	x 6
x	12	Dolgov Evgeniy	x	♂	1976	Кемеровская область	x 6
x	13	Nokhrin Sergej	x	♂	1972	Ленинск-Кузнецкий	x 6
x	14	Gusev Vasilij	x	♂	1977	Ленинск-Кузнецкий	x 6

- **Trainer** - тренер (необязательно)
Ключевые слова для поиска колонки: 'coach', 'тренер'
Значение: любое строковое значение
- **Points** - абсолютный зачет (необязательно)
Ключевые слова для поиска колонки: 'points', 'pts', 'очки'
Значение: должно быть указано число.
- **Team** - команда (необязательно)
Ключевые слова для поиска колонки: 'team', 'команда'
Значение: любое строковое значение
- **Comment** - комментарий (необязательно)
Ключевые слова для поиска колонки: 'коммент', 'comment'
Значение: любое строковое значение
- **DQ** - допинг контроль (необязательно)
Ключевые слова для поиска колонки: 'dq', 'дк'
Значение: любое строковое значение

Колонки с результатами выступлений

Колонки с результатами выступления не являются обязательными, и распознаются парсером только при наличии соответствующих ключевых слов. То есть могут быть распознаны как все три движения, так и отдельные: присед, жим или тяга.

- **Squat** - присед
Ключевые слова для поиска колонки: 'приседание', 'присед', 'прис', 'squat'
Пример названия колонки: 'Приседание со штангой', 'Прис.', 'Squat'

- **Benchpress** - жим лежа

Ключевые слова для поиска колонки: 'жим', 'benchpress'

Пример названия колонки: 'Жим лежа', 'Жим'

- **Deadlift** - становая тяга

Ключевые слова для поиска колонки: 'тяга', 'deadlift'

Пример названия колонки: 'Становая тяга', 'Тяга'

При этом парсер определяет номера подходов по колонкам следующим образом:

1. Номера подходов указаны под названием одной объединенной колонки

Squat			
1	2	3	Рек

2. Номера подходов указаны под названиями колонок

Тяга	Тяга	Тяга	Тяга
1	2	3	Рек

3. Номера подходов указаны вместе с названием колонок

Тяга 1	Тяга 2	Тяга 3	Тяга 4
--------	--------	--------	--------

Обратите внимание что 4-ый рекордный подход может быть указан как номером '4', так и значениями 'рес' или 'рек', или может быть не указан вовсе.

Если парсер распознал номера подходов, то итоговые результаты парсер не распознает, так как в форме загрузки протокола можно эти значения вычислить нажатием соответствующих кнопок:

Результаты																
Вес	ВК	М	AGE	П1	П2	П3	П	Ж1	Ж2	Ж3	Ж	Т1	Т2	Т3	Т	Ито

Если парсер не нашел номера подходов, а нашел только названия, то в форму загрузки протокола попадают только итоговый результат:

При распознавании результатов, зачеркнутое (75) значение или значение отмеченное красным цветом (75) распознается как неудачный подход и попадает в форму протокола со знаком минус в начале (-75).