



# OPEN REFINE WORKSHOP

Обработка данных в  
Open Refine для джедаев

Виталий Власов  
DH Week, Университет ИТМО, 2021

# Роль данных

Истории



Данные

# Какие данные бывают

GOOD DATA

VS

BAD DATA



Linked Data

Public Datasets

Open Government Data

PDF

Big Data

XML

CSV

Данные НПО

HTML

DOC

Open Spending

Статистика

XLS

Gov Budget Data

Неструктурированный данные

Dirty Data

# Открываем XLS

<input type="checkbox"/> All	<input type="checkbox"/> Качество поверх	<input type="checkbox"/> Column	<input type="checkbox"/> Column2	<input type="checkbox"/> Column3	<input type="checkbox"/> Column4	<input type="checkbox"/> Column5	<input type="checkbox"/> Column6	<input type="checkbox"/> Column7	<input type="checkbox"/> Column8	<input type="checkbox"/> Column9		
		1.		Название точки мониторинга р.Чу - Бурулдайский мост, 0,01 км выше моста; 0,2 м	р. Чу - г. Токмок, 1 км выше города; 0,2 м	р. Чу - г. Токмок, 0,5 км ниже города; 0,2 м	р. Чу - с. Милянфан, 2,5 км выше впадения дренажного коллектора N 17; 0,2 м	р. Чу - с. Милянфан, 1,5 км ниже впадения дренажного коллектора N 17; 0,2 м	р. Чу - с. Васильевка, 0,5 км выше села; 0,2 м	р. Чу - с. Васильевка, 0,3 км ниже села; 0,2 м		
		2.	Годы	Координаты	42.805462, 75.709088	42.856061, 75.282400	42.858074, 75.190389	42.975743, 74.885518	42.993731, 74.805681	43.114710, 74.573014	43.134715, 74.536480	
		3.	2011	Название веществ	Температура воды, град.  Взвешенные вещества, мг/л  Цветность, град.  рН  Кислород  Сульфаты  Хлориды  Сумма ионов (минерализация), мг/л  Общая жесткость, ммоль/л  Азот нитритный, мг/л  Азот нитратный, мг/л  ВПК 5, мг/л  Нефтепродукты, мг/л	11.9  112.3  5  8.1  9.47  35.5  9  198  2.89  0.006  0.79  0.62  0.01	14.3  104.3  5  8.05  9.03  39.3  10.1  217  3.12  0.003  0.89  0.51  0.01  16  74.9  5  8.15  8.87  44  12.3  243  3.41  0.023  1.25  1.53  0.02  3.64  0.012  1.62  1.49  0.01  4.2  0.015  1.68  0.84  0.01  4.3  0.012  1.87  0.84  0.01  4.32  0.079  1.86  1.6  0.01  15.5  15.9	17  39.4  5  8.21  8.86  55.8  12.6  270  3.64  0.012  1.62  1.49  0.01  4.2  0.015  1.68  0.84  0.01  4.3  0.012  1.87  0.84  0.01  4.32  0.079	16.3  46.7  5  8.2  8.82  79.7  14.6  317  4.2  0.015  1.68  0.84  0.01  4.3  0.012  1.87  0.84  0.01  4.32  0.079	14.8  70.9  10  8.08  8.92  100.7  20.1  345  4.2  0.015  1.68  0.84  0.01  4.3  0.012  1.87  0.84  0.01  4.32  0.079	15  61.7  10  8.09  8.89  104.2  22.2  352  4.3  0.012  1.87  0.84  0.01  4.32  0.079	
		4.	2012	Название веществ	Температура воды, град.  Взвешенные вещества, мг/л  Цветность, град.	12  17.7  4	15.4  27.4  4	16.4  26.9  4	17.9  14.4  5	18.7  49.3  5	15.5  153.4  6	15.9  98.4  8

<https://github.com/inxaoc/openrefine-workshop/blob/main/Data-on-Irrigation-water-quality.csv>

# Что не так с таблицей?

- Первая строка («шапка») - это название столбцов (обычно это параметры или ключевые слова «дата», «название», «категория», «геокоординаты»).
- В каждой колонке должны находиться значения одного типа (строка, число, дата, геокоординаты и пр.)
- Параметры которые находятся в Column2 должны быть в шапке таблицы
- Нельзя отсортировать данные например по «году»
- Много пустых ячеек
- Цель создания этой таблицы - сводная таблица статистики

# ПРОБЛЕМЫ ПРИ РАБОТЕ С ДАННЫМИ

- Не в машиночитаемом формате (html, doc, pdf)
- «Грязные» данные: не нормализованные данные (одни и те же данные записаны в разной форме) и пр.
- Отсутствие геокоординат (или разные системы) в датасетах с адресами
- Отсутствие технической поддержки
- Несвоевременное обновление данных
- Недостаточный уровень детализации данных
- Отсутствие актуальных наборов данных
- Использование латиницы вместо кириллицы

# ПРОБЛЕМЫ ПРИ РАБОТЕ С ДАННЫМИ

- Не в машиночитаемом формате (html, doc, pdf)
- «Грязные» данные: не нормализованные данные (одни и те же данные записаны в разной форме) и пр.
- Отсутствие геокоординат (или разные системы) в датасетах с адресами
- Отсутствие технической поддержки
- Несвоевременное обновление данных
- Недостаточный уровень детализации данных
- Отсутствие актуальных наборов данных
- Использование латиницы вместо кириллицы

# ФУНКЦИИ OPEN REFINE

- Преобразование данных к нужному формату
- Предварительная обработка. Нахождение ошибок и опечаток
- Получение представления о массиве данных

## **Установка:**

1. Скачать OpenRefine: <http://openrefine.org/>
2. Запустить установщик
3. Открыть браузер и перейти по ссылке: <http://127.0.0.1:3333>

<http://data.gov.spb.ru/datasets/69/>

# ЗАГРУЗКА МАССИВА ДАННЫХ

# Создаем новый проект

The screenshot shows the Google Refine interface for creating a new project. A red arrow points to the 'Create Project' button in the top-left corner of the sidebar. Another red arrow points to the 'Выбор файлов' (Select files) button, which is highlighted with a red border.

**Google refine** A power tool for working with messy data.

**Create Project** **Create Project** (highlighted by a red box)

Open Project

Import Project

Get data from

This Computer **Выбор файлов** (files not selected) (highlighted by a red box)

Web Addresses (URLs)

Clipboard

Google Data

Locate one or more files on your computer to upload:

Next »

Version 2.5 [r2407]

Help

About

# Загрузка исходного массива



A power tool for working with messy data.

Create Project

Open Project

Import Project

## Create a project by importing data. What kinds of data files can I import?

TSV, CSV, \*SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with Google Refine extensions.

Get data from

This Computer

Web Addresses (URLs)

Clipboard

Google Data

Locate one or more files on your computer to upload:

Выбор файлов



ГК.csv

Next »



Version 2.5 [r2407]

Help

About

# Настройка импорта файла. Выбор типа файла

Google refine A power tool for working with messy data.

Create Project    « Start Over    Configure Parsing Options    Project name csv    Create Project »

Open Project    Import Project

	_input	_num	_widgetName	_source	_resultNumber	_pageUrl
1.		1	opendatakg	opendatakg		1 http://zakupki.okmot.kg/pub/publicOrder.action;jsessionid=d8MH8u997k+M+iUrHNH9VonZ.u page=2

Parse data as    Character encoding    Update Preview

CSV / TSV / separator-based files    Line-based text files    Fixed-width field text files    PC-Axis text files    JSON files    RDF/N3 files    XML files    Open Document Format spreadsheets (.ods)    RDF/XML files

Character encoding:

Columns are separated by:

- commas (CSV)
- tabs (TSV)
- custom ,

Escape special characters with \

Ignore first 0 line(s) at beginning of file    Parse next 1 line(s) as column headers    Discard 0 initial row(s) of data    Load at 0 most row(s) of data    Parse cell text into blank

Version 2.5 [r2407]    Help    About

# Изменение имени проекта

Google refine

A power tool for working with messy data.

Create Project

« Start Over

Configure Parsing Options

Project name Contracts

Create Project »

Open Project

Import Project

b/orderDetail.action;jsessionid=d8MH8u997k+M+iUrHNH9VonZ.undefined?  
A%2F%2Fzakupki.okmot.kg%2Fpub%2FpublicOrder.action

column\_2\_link/\_source

/pub/orderDetail.action;jsessionid=d8MH8u997k+M+iUrHNH9VonZ.undefined?  
id=1876108&urlBack=http%3A%2F%2Fzakupki.okmot.kg%2Fpub%2FpublicOrder.action

## Настройка кодировки

Parse data as

Character encoding

Update Preview

CSV / TSV / separator-based files

Line-based text files

Fixed-width field text files

PC-Axis text files

JSON files

RDF/N3 files

XML files

Open Document Format  
spreadsheets (.ods)

RDF/XML files

Columns are separated by

commas (CSV)

tabs (TSV)

custom ,

Escape special characters with \

Ignore 0 line(s)  
first at beginning of file

Parse 1 line(s)  
next as column headers

Discard 0 row(s)  
initial of data

Load at 0 row(s)  
most of data

Parse cell text into  
blank numbers  Store  
blank rows



Version 2.5 [r2407]

Help

About

# Настройка кодировки

Google refine

A power tool for working with messy data.

Create Project

« Start Over

Configure Parsing Options

Project name Contracts

Create Project »

Open Project

Import Project

title	col_1_link/text	col_1_number	col_1_2	col_1_number_2	col_1_3	col_1_number_3	date	col_1_number_4	col_1_5	column_10
—Т—И—и—И—J—∞— ±—И—А— —Б—И— Г—ð—μ—Л— €—≤—И— И—l—∞— ≤—И—i— л—∞—±— €—П—л— μ—ї—μ—л— —E—и— —B—И— J—∞—A— A—≥—И— И—≥—И— €—И—л— ∞—j—∞—ї—	5; 0; 0	5 000		1	—ð—И—	5.06; 17; 13	05.06.2015	29.06; 10; 0	29.06.2015 10:00	—Ю—њ—Г— ±—ї—Е—Ї— И—≤—∞—љ—

## Select Encoding

Common Encodings

All Encodings

Encoding Aliases

ISO-8859-1 csISOLatin1, IBM-819, iso-ir-100, 8859\_1, ISO\_8859-1, I1, ISO8859-1, ISO\_8859\_1, cp819, ISO8859\_1, latin1, ISO\_8859-1:1987, 819, IBM819

US-ASCII cp367, ascii7, ISO646-US, 646, csASCII, us, iso\_646.irv:1983, ISO\_646.irv:1991, IBM367, ASCII, default, ANSI\_X3.4-1986, ANSI\_X3.4-1968, iso-ir-6

UTF-16 utf16, UTF\_16, UnicodeBig, unicode

UTF-16BE X-UTF-16BE, UTF\_16BE, ISO-10646-UCS-2, UnicodeBigUnmarked

UTF-16LE UnicodeLittleUnmarked, UTF\_16LE, X-UTF-16LE

UTF-8 UTF8, unicode-1-1-utf-8

Cancel

Update Preview

Ignore first 0 line(s) at beginning of file

Parse next 1 line(s) as column headers

Discard initial 0 row(s) of data

Load at most 0 row(s) of data

Parse cell text into numbers  Store blank rows



Version 2.5 [r2407]

Help  
About

# Настройка дополнительных параметров

**Google refine** A power tool for working with messy data.

Create Project    « Start Over    Configure Parsing Options    Project name: Contracts    Create Project »

Open Project    Import Project

b/orderDetail.action;jsessionid=d8MH8u997k+M+iUrHNH9VonZ.undefined?A%2F%2Fzakupki.okmot.kg%2Fpub%2FpublicOrder.action    column\_2\_link/\_source /pub/orderDetail.action;jsessionid=d8MH8u997k+M+iUrHNH9VonZ.undefined?id=1876108&urlBack=http%3A%2F%2Fzakupki.okmot.kg%2Fpub%2FpublicOrder.action

**Parse data as**

**CSV / TSV / separator-based files**

- Line-based text files
- Fixed-width field text files
- PC-Axis text files
- JSON files
- RDF/N3 files
- XML files
- Open Document Format spreadsheets (.ods)
- RDF/XML files

Character encoding:

Update Preview

Columns are separated by:

- commas (CSV)
- tabs (TSV)
- custom ,

Escape special characters with \

Ignore first 0 line(s) at beginning of file

Parse next 1 line(s) as column headers

Discard 0 initial row(s) of data

Load at 0 most row(s) of data

Parse cell text into numbers     Store blank rows

  
Version 2.5 [r2407]

Help    About

# ПРЕДВАРИТЕЛЬНАЯ ОЧИСТКА ДАННЫХ

# Основные элементы страницы

Google refine Contracts Permalink Open... Export Help

Facet / Filter Undo / Redo 4

Using facets and filters 

Количество строк в файле

Use facets and filters of your data to quickly find the right filter methods for each data column.

Not sure how to get started? Watch these screencasts

1480 rows

Show as: rows records Show: 5 10 25 50 rows

Extensions: Named-entity recognition ▾ Freebase ▾ RDF ▾

« first < previous 1 - 10 next > last »

All	_input	_num	_widgetName	_source	_resultNumber	_pageUrl
1.		1	opendatakg	opendatakg	1	<a href="http://zakupki.okmot.kg/pub/publicOrder.action;jsess">http://zakupki.okmot.kg/pub/publicOrder.action;jsess</a> page=2
2.		2	opendata	opendata	2	<a href="http://zakupki.okmot.kg/pub/publicOrder.action;jsess">http://zakupki.okmot.kg/pub/publicOrder.action;jsess</a> page=2
3.		3	opendatakg	opendatakg	3	<a href="http://zakupki.okmot.kg/pub/publicOrder.action;jsess">http://zakupki.okmot.kg/pub/publicOrder.action;jsess</a> page=2
4.		4	opendatakg	opendatakg	4	<a href="http://zakupki.okmot.kg/pub/publicOrder.action;jsess">http://zakupki.okmot.kg/pub/publicOrder.action;jsess</a> page=2
5.		5	opendatakg	opendatakg	5	<a href="http://zakupki.okmot.kg/pub/publicOrder.action;jsess">http://zakupki.okmot.kg/pub/publicOrder.action;jsess</a> page=2
6.		6	opendatakg	opendatakg	6	<a href="http://zakupki.okmot.kg/pub/publicOrder.action;jsess">http://zakupki.okmot.kg/pub/publicOrder.action;jsess</a> page=2
7.		7	opendatakg	opendatakg	7	<a href="http://zakupki.okmot.kg/pub/publicOrder.action;jsess">http://zakupki.okmot.kg/pub/publicOrder.action;jsess</a> page=2
8.		8	opendatakg	opendatakg	8	<a href="http://zakupki.okmot.kg/pub/publicOrder.action;jsess">http://zakupki.okmot.kg/pub/publicOrder.action;jsess</a> page=2

Количество отображаемых строк

название столбца

# Изменение наименований столбцов

473 rows						Extensions: Named-e
Show as: rows records			Show: 5 10 25 50 rows			
title	col_1_link/_tex	col_1_number	col_1_2	col_1_number	col_1_3	
Facet		.18	914 99818	2	Торги с неограниченным участием	
Text filter			2 300 00000	1	Торги с неограниченным участием	
Edit cells					Торги с неограниченным участием	
<b>Edit column</b>					Торги с ограниченным участием	
Transpose						
Sort...						
View						
Reconcile						
Extract named entities...						
адресу в г.Каракол ул. Иссык-Кульская № 8" Иссык- Кульской областной стоматологической поликлиники						
Приобретение продуктов питания	336; 700	336 70000				

Context menu for the 'Edit column' option:

- Split into several columns...
- Add column based on this column...
- Add column by fetching URLs...
- Add columns from Freebase ...
- Rename this column**
- Remove this column
- Move column to beginning
- Move column to end
- Move column left
- Move column right

Renaming dialog box:

Enter new column name: Статус

Buttons: Отменить (Cancel), OK

# Удаление столбцов

473 rows      Extensions: [!](#)

Show as: [rows](#) [records](#)      Show: [5](#) [10](#) [25](#) [50](#) rows

_link/_tex	col_1_number	col_1_2	col_1_number	col_1_3
ИМОНТ Х	Facet	▶ 18	2	Торги с неограниченным участием
ие ки для ных	Text filter		000	1 Торги с неограниченным
	Edit cells	▶		
	Edit column	▶	Split into several columns...	
	Transpose	▶	Add column based on this column...	
	Sort...		Add column by fetching URLs...	
	View	▶	Add columns from Freebase ...	
	Reconcile	▶	Rename this column	
	Extract named entities...		Remove this column	
			Move column to beginning	
			Move column to end	
			Move column left	
			Move column right	
и ческой и	336; 700	336 70000	2	Запрос котировок
и ческой и	658: 134	658 13400	1	Запрос
и ческой и				3.06

# Преобразование в числовой формат

Show: 5 10 25 50 rows « first < previous 1 - 10

link/_title	col_1_link/_tex	col_1_2	col_1_number_	col_1_3	col_1_nu
текущий ремонт социальных объектов		Facet	2 Торги с неограниченным участием	4.06; 17; 47	
приобретение спец.техники для коммунальных служб		Text filter			
приобретение продуктов питания		Edit cells	Transform...		
На капремонт кровли и отдельных помещений зданий (детской отделении) расположенного по адресу в г.Каракол ул. Иссык-Кульская № 8" Иссык-Кульской областной стоматологической		Edit column	Common transforms		
		Transpose	Fill down		
		Sort...	Blank down		
		View	Split multi-valued cells...		
		Reconcile	Join multi-valued cells...		
		Extract named entities...	Cluster and edit...		

473 rows Show as: rows records Show: 5 10 25 50 rows « first < previous

Custom text transform on column col\_1\_2

Expression: value.replace(" ", "") Language: Google Refine Expression Language (GREL)

No syntax error.

row	value	value.replace(" ", "")
1.	914 99818	91499818
2.	2 300 00000	230000000
3.	1 228 21600	122821600
4.	802 00000	80200000
5.	336 70000	33670000
6.	658 13400	65813400
7.	4 400 00000	440000000

On error:  keep original  set to blank  store error

Re-transform up to 10 times until no change

OK Cancel

Удаление пробелов: value.replace(" ", "")

# Разделение колонок на две

Screenshot of a data processing interface showing the process of splitting a column.

The main interface shows a table with columns: \_1\_number, col\_1\_4, col\_1\_number\_, col\_1\_5, and Статус. The col\_1\_5 column contains values like "25.06; 10; 0" and "24.06; 14; 0". A context menu is open over the col\_1\_5 column, with the "Edit column" option highlighted.

A modal dialog titled "Split column col\_1\_5 into several columns" is displayed. It contains two main sections:

- How to Split Column:**
  - by separator
  - by field lengths
- After Splitting:**
  - Guess cell type
  - Remove this column

The "Separator" field is set to ",". The "Split into" field is empty. Below these fields is a text input for "List of integers separated by commas, e.g., 5, 7, 15".

Buttons at the bottom of the dialog are "OK" and "Cancel".

Below the dialog, the table shows the results of the split operation. The col\_1\_5 column has been split into two new columns: col\_1\_5\_1 and col\_1\_5\_2. The first row now contains "25.06.2015" in col\_1\_5\_1 and "10:00" in col\_1\_5\_2. Subsequent rows show "24.06.2015" and "14:00", "24.06.2015" and "09:00", and "17.06.2015" and "11:00".

_1_number	col_1_4	col_1_number_	col_1_5	Статус
47	04.06.2015 17:47	25.06; 10; 0		
50	03.06.2015 18:50	24.06; 14; 0		
		<b>Split into several columns...</b>		
		Add column based on this column...		
		Add column by fetching URLs...		
		Add columns from Freebase ...		
		Rename this column		
		Remove this column		
		Move column to beginning		
		Move column to end		
		Move column left		
		Move column right		
18	03.06.2015 15:18	15.06; 9; 0	15.06.2015 09:00	Опубликован

	col_1_5_1	col_1_5_2
1	25.06.2015	10:00
	24.06.2015	14:00
	24.06.2015	09:00
	17.06.2015	11:00

# Преобразование в формат дат

« first < previous 1 - 10 next > last »

col_1_4	col_1_number_	col_1_51	col_1_52
4.06.2015 7:47	25.06; 10; 0		
3.06.2015 8:	Transform...	Facet Text filter	
	Common transforms	Edit cells	
3.06.2015 8:	Fill down	Edit column	
	Blank down	Transpose	
		Sort...	

8: Transform...

Common transforms

Fill down

Blank down

col\_1\_51

2015-06-25T00:00:00Z

2015-06-24T00:00:00Z

2015-06

Custom text transform on column col\_1\_51

Expression: value.replace(".", "/")

Language: Google Refine Expression Language (GREL)

No syntax error.

row	value	value.replace(".", "/")
1.	25.06.2015	25/06/2015
2.	24.06.2015	24/06/2015
3.	24.06.2015	24/06/2015
4.	17.06.2015	17/06/2015
5.	15.06.2015	15/06/2015
6.	10.06.2015	10/06/2015
7.	09.07.2015	09/07/2015

On error:  keep original  set to blank  store error

Re-transform up to 10 times until no change

OK Cancel

# Фасеты

col\_1\_3 col\_1\_number\_ col\_1\_4 col\_1\_5

Facet ► Text facet

Text filter

Numeric facet

Timeline facet

Scatterplot facet

Custom text facet...

Custom numeric facet...

Customized facets

Text filter

Edit cells

Edit column

Transpose

Sort...

View

Reconcile

Extract named entities...

Запрос котировок	3.06; 15; 18	03.06.2015 15:18	15.06; 9
------------------	--------------	---------------------	----------

col\_1\_3 change

5 choices Sort by: name count Cluster

Запрос котировок 110

Консультационные услуги 5

Торги с неограниченным  
участием 265

Торги с ограниченным участием 10

Электронные государственные  
закупки 3

(blank) 80

Facet by choice counts

80 matching rows (473 total)

Show as: rows records Show: 5 10 25 50 rows

All	_input	_num	_widgetName	_sour
Facet		213	GoszakupkiKG	Goszakup
► Edit rows				
► Edit columns				
► View				
Star rows				
Unstar rows				
Flag rows				
Unflag rows				
Remove all matching rows				

ИЗУЧЕНИЕ ДАННЫХ

# Использование фасетов для изучения встречающихся значений

## Текстовые фасеты

393 rows

Show as: rows records Show: 5 10 25 50 rows

\_number\_ Метод закупки col\_1\_number\_ col\_1\_4

2 Facet ▶ Text facet

Text filter Numeric facet

1 Edit cells Timeline facet

Edit column Scatterplot facet

1 Transpose Custom text facet...

Sort... Custom numeric facet...

View Customized facets ▶

Reconcile

Extract named entities...

**Статус** change

5 choices Sort by: name count Cluster

Изменен 36  
Опубликован 206  
Отменён 28  
Предложения приняты 41  
Создан протокол вскрытия 82  
Facet by choice counts

**Метод закупки** change

5 choices Sort by: name count Cluster

Запрос котировок 110  
Консультационные услуги 5  
Торги с неограниченным  
участием 265  
Торги с ограниченным  
участием 10  
Электронные государственные  
закупки 3  
Facet by choice counts

**Статус** change

5 choices Sort by: name count Cluster

Опубликован 206  
Создан протокол вскрытия 82  
Предложения приняты 41  
Изменен 36  
Отменён 28  
Facet by choice counts

**Метод закупки** change

5 choices Sort by: name count Cluster

Торги с неограниченным  
участием 265  
Запрос котировок 110  
Торги с ограниченным  
участием 10  
Консультационные услуги 5  
Электронные государственные  
закупки 3  
Facet by choice counts

Статус	Количество
Изменен	36
Опубликован	206
Отменён	28
Предложения приняты	41
Создан протокол вскрытия	82

Метод закупки	Количество
Запрос котировок	110
Консультационные услуги	5
Торги с неограниченным участием	265
Торги с ограниченным участием	10
Электронные государственные закупки	3

Метод закупки	Количество
Торги с неограниченным участием	265
Запрос котировок	110
Торги с ограниченным участием	10
Консультационные услуги	5
Электронные государственные закупки	3

# Использование фасетов для изучения встречающихся значений

## Временные фасеты

The screenshot shows a data analysis interface with a context menu open over a date facet. The menu items include:

- Text facet
- Numeric facet
- Timeline facet
- Scatterplot facet
- Custom text facet...
- Custom numeric facet...
- Customized facets
- Facet (highlighted with a red box)
- Text filter
- Edit cells
- Edit column
- Transpose
- Sort...
- View
- Reconcile
- Extract named entities...

To the right, there is a facet/filter panel for the "date" facet, which includes a histogram, a date selector (2013-01-19), and a time range (11:37:00 — 23:22:00). Buttons for Refresh, Reset All, and Remove All are also present.

The screenshot shows a data grid with the following details:

- Header: Facet / Filter, Undo / Redo 12, 259 matching rows (393 total), Extensions: Named-entity recognition ▾, Freebase ▾, RDF ▾.
- Filter: date (2014-11-20, 02:17:12 — 23:22:00).
- Show as: rows records, Show: 5 10 25 50 rows.
- Table Headers: \_tex, col\_1\_2, col\_1\_number\_, Метод закупки, col\_1\_number\_, col\_1\_4, date, col\_1\_number\_, Статус.
- Data Rows (partial):

_tex	col_1_2	col_1_number_	Метод закупки	col_1_number_	col_1_4	date	col_1_number_	Статус
91499818		2	Торги с неограниченным участием	4.06; 17; 47	04.06.2015 17:47	2015-04-06T17:47:00Z	25.06; 10; 0	Опубликован
230000000		1	Торги с неограниченным участием	3.06; 18; 50	03.06.2015 18:50	2015-03-06T18:50:00Z	24.06; 14; 0	Опубликован
122821600		1	Торги с неограниченным участием	3.06; 18; 8	03.06.2015 18:08	2015-03-06T18:08:00Z	24.06; 9; 0	Опубликован
80200000		1	Торги с ограниченным участием	3.06; 17; 2	03.06.2015 17:02	2015-03-06T17:02:00Z	17.06; 11; 0	Опубликован

# Использование фасетов для изучения встречающихся значений

## Числовые фасеты

305 matching rows (393 total) Extens

Show as: rows records Show: 5 10 25 50 rows

title	col_1_link/_tex	col_1_2	col_1_number_	Метод закупки
текущий ремонт социальных объектов	Facet	Text facet		
	Text filter	Numeric facet		
приобретение спец.техники для коммунальных служб	Edit cells	Timeline facet		
	Edit column	Scatterplot facet		
приобретение продуктов питания	Transpose	Custom text facet...		
	Sort...	Custom numeric facet...		
На капремонт кровли и отдельных помещений зданий (детской отделении) расположенного по адресу в г.Каракол ул. Иссык-Кульская № 8" Иссык-	View	Customized facets		
	Reconcile			
	Extract named entities...			

Facet / Filter Undo / Redo 12

Refresh Reset All Remove All

col\_1\_2 change reset

0.00 — 1,000,000,000.00

Numeric  Non-numeric  Blank  Error  
356 0 37 0

305 matching rows (393 total)

Show as: rows records Show: 5 10 25 50 rows

title	col_1_link/_tex	col_1_2	col_1_number_	Метод закупки
текущий ремонт социальных объектов		91499818	2	
приобретение спец.техники для коммунальных служб		230000000	1	
приобретение продуктов питания		122821600	1	
На капремонт кровли и отдельных помещений зданий (детской		80200000	1	

# Использование кластеризации

**Facet / Filter** Undo / Redo 12 **393 rows**

Refresh Reset All Remove All

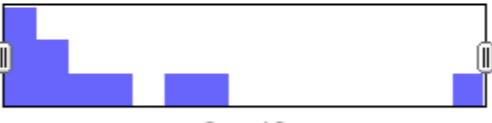
**col\_1\_link/\_text** change  
306 choices Sort by: name count Cluster Fk- /pub/orderDetail.action;jse: id=1874431&urlBack=http://er.action

products питания 9  
Приобретение продуктов питания 7  
Продукты питания 7  
Приобретение медикаментов и ИМН 5  
ГСМ 4  
медикаменты 4  
Приобретение ГСМ 4  
Завершение Строительства  
Неполной средней школы "Кызыл Окт с.Котур-Суу(Сары Камыш) Кабак а/о Джумгальского района" 3

**Cluster & Edit column "col\_1\_link/\_text"**

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find out more ...](#)

Method key collision Keying Function fingerprint 10 clusters found

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value	# Rows in Cluster
2	3	<ul style="list-style-type: none"> <li>приобретение медикаментов (2 rows)</li> <li>Приобретение медикаментов (1 rows)</li> </ul>	<input type="checkbox"/>	приобретение медикаментов	 2 — 16
2	2	<ul style="list-style-type: none"> <li>Приобретение медицинского оборудования (1 rows)</li> <li>приобретение медицинского оборудования (1 rows)</li> </ul>	<input type="checkbox"/>	Приобретение медицинского	Average Length of Choices
2	5	<ul style="list-style-type: none"> <li>медикаменты и ИМН (3 rows)</li> <li>Медикаменты и ИМН (2 rows)</li> </ul>	<input type="checkbox"/>	Медикаменты и ИМН	Facet / Filter Undo / Redo 13 393 rows
2	2	<ul style="list-style-type: none"> <li>Приобретение каменного угля (1 rows)</li> <li>приобретение каменного угля (1 rows)</li> </ul>	<input type="checkbox"/>	Приобретение каменного угля	
2	16	<ul style="list-style-type: none"> <li>продукты питания (9 rows)</li> <li>Продукты питания (7 rows)</li> </ul>	<input type="checkbox"/>	Продукты питания	
2	7	<ul style="list-style-type: none"> <li>медикаменты (4 rows)</li> <li>Медикаменты (3 rows)</li> </ul>	<input type="checkbox"/>	Медикаменты	
2	8	<ul style="list-style-type: none"> <li>Приобретение продуктов питания (7 rows)</li> <li>приобретение продуктов питания (1 rows)</li> </ul>	<input type="checkbox"/>	Приобретение продуктов питания	

Select All Deselect All

Refresh Reset All Remove All

**col\_1\_link/\_text** change  
296 choices Sort by: name count Cluster

products питания 16  
Приобретение продуктов питания 8  
медикаменты 7  
медикаменты и ИМН 5  
Приобретение медикаментов и ИМН 5  
ГСМ 4  
Приобретение ГСМ 4  
Продукты питания 4  
Завершение Строительства  
Неполной средней школы "Кызыл Октябрь" в с.Котур-Суу(Сары Камыш) Кабак а/о

Fk- /pub/orderDetail.action;jse: id=1874431&urlBack=http://er.action  
Fk- /pub/orderDetail.action;jse: id=1874431&urlBack=http://er.action  
Fk- /pub/orderDetail.action;jse: id=1874431&urlBack=http://er.action  
Fk- /pub/orderDetail.action;jse: id=1874431&urlBack=http://er.action  
Fk- /pub/orderDetail.action;jse: id=1874431&urlBack=http://er.action

Пример с паспортами МКД и такси

# Разделение колонок

- Используйте функцию Edit column - Split into several columns, в качестве разделителя «запятую»
- У вас появится новая колонка рядом. Переименуйте редактируемую колонку Edit column - Rename this column (как на скриншоте далее)
- Экспортируйте выборку в формат Comm Separated Value (CSV)

	Координаты	Год	Что измеряют
Facet	2015	Температура воды, град.	
Text filter	2015	Взвешенные вещества, мг/л	
Edit cells	2015	Цветность, град.	
Edit column	2015	pH	
Transpose		Split into several columns...	
Sort...		Add column based on this column...	
View		Add column by fetching URLs...	
Reconcile		Rename this column	
42.805462, 75.709088		Remove this column	
42.805462, 75.709088		Move column to beginning	
42.805462, 75.709088		Move column to end	
42.856061, 75.282400		Move column left	
42.856061, 75.282400		Move column right	
42.856061, 75.282400	2015	цветность, град.	
42.856061, 75.282400	2015	pH	

**Split column Координаты into several columns**

**How to Split Column**

by separator

Separator   regular expression

Split into  columns at most (leave blank for no limit)

by field lengths

List of integers separated by commas, e.g., 5, 7, 15

**After Splitting**

Guess cell type

Remove this column

**OK** **Cancel**

	Координаты 1	Координаты 2	Год	Что измеряют
Facet	2015	75.709088	Температура воды, г	
Text filter	2015	75.709088	Взвешенные веществ	
Edit cells	2015	75.709088	Цветность, град.	
Edit column	2015	75.709088	pH	
Transpose			Split into several columns...	
Sort...			Add column based on this column...	
View			Add column by fetching URLs...	
Reconcile			Rename this column	
42.805462			Remove this column	
42.805462			Move column to beginning	
42.805462			Move column to end	
42.856061			Move column left	
42.856061			Move column right	
42.856061	2015	75.282400	цветность, град.	

**Facet / Filter** Undo / Redo ↻

Refresh Reset All Remove All

**Год** change invert reset

5 choices Sort by: name count Cluster

2011	21
2012	21
2013	21
2014	21
<b>2015</b>	21

Facet by choice count =

**Что измеряют** change invert reset

13 choices Sort by: name count Cluster

Азот нитратный, мг/л	21
Азот нитритный, мг/л	21
Взвешенные вещества, мг/л	21
ВПК 5, мг/л	21
Кислород	21
Нефтепродукты, мг/л	21
Общая жесткость, ммоль/л	21
<b>Сульфаты</b>	21
Сумма ионов (минерализация), мг/л	21
Температура воды, град.	21
Хлориды	21

=

**21 matching rows (1365 total)**

Extensions: undefined ▾

Show as: **rows** records Show: 5 10 25 50 rows« first < previous **1 - 21** next > last »

			All	Место	LAT	LONG	Год	Что измеряют	Значение
★	58.	p.Чу - Бурулдайский мост, 0,01 км выше моста, 0,2 м	42.805462	75.709088	2015	Сульфаты	38.2		
★	123.	p. Чу - г. Токмок, 1 км выше города, 0,2 м	42.856061	75.2824	2015	Сульфаты	37.2		
★	188.	p. Чу - г. Токмок, 0,5 км ниже города, 0,2 м	42.858074	75.190389	2015	Сульфаты	38		
★	253.	p. Чу - с. Милянфан, 2,5 км выше владения дренажного коллектора N 17, 0,2 м	42.975743	74.885518	2015	Сульфаты	60.2		
★	318.	p. Чу - с. Милянфан, 1,5 км ниже владения дренажного коллектора N 17, 0,2 м	42.993731	74.805681	2015	Сульфаты	87.2		
★	383.	p. Чу - с. Васильевка, 0,5 км выше села, 0,2 м	43.11471	74.573014	2015	Сульфаты	82.1		
★	448.	p. Чу - с. Васильевка, 0,3 км ниже села, 0,2 м	43.134715	74.53648	2015	Сульфаты	82.3		
★	513.	p.Чу - пос. Нижне-Чуйский, 0,5 км выше поселка, 0,2 м	43.188808	74.370312	2015	Сульфаты	94.3		
★	578.	p.Чу - пос. Нижне-Чуйский, 0,7 км ниже поселка, 0,2 м	43.212833	74.312634	2015	Сульфаты	95.7		
★	643.	p.Чон - Кемин - устье, 0,3 км выше устья, 0,2 м	42.718415	75.886425	2015	Сульфаты	21.7		
★	708.	p. Кичи-Кемин - пос. Актюз, 3 км выше поселка, 0,2 м	42.886682	76.099286	2015	Сульфаты	17.7		
★	773.	p. Кичи-Кемин - пос. Актюз, 8 км ниже поселка, 0,2 м	42.846232	76.04528	2015	Сульфаты	20.1		
★	838.	p. Красная - г. Токмок, 11 км ниже города, 0,2 м	42.886948	75.130039	2015	Сульфаты	40.3		
★	903.	p. Ноуруз - с. Новопокровка, в черте села, 0,2 м	42.885323	74.704267	2015	Сульфаты	92.6		
★	968.	p. Аламедин - г. Бишкек, 1 км выше города, 0,2 м	42.796615	74.649026	2015	Сульфаты	25.6		
★	1033.	p. Аламедин - г. Бишкек, 2 км ниже города, 0,2 м	42.94701	74.673194	2015	Сульфаты	58		
★	1098.	p. Ала Арча - г. Бишкек, 4 км выше города, 0,2 м	42.75572	74.556464	2015	Сульфаты	16		
★	1163.	p. Ала Арча - г. Бишкек, 1 км ниже города, 0,2 м	42.957061	74.570197	2015	Сульфаты	70.8		
★	1228.	Канал ЗБЧК - с.Сокулук, 0,8 км ниже села	42.866423	74.234979	2015	Сульфаты	43.7		
★	1293.	p. Аксу - с. Тюлек, 1 км выше села, 0,2 м	43.066228	74.118287	2015	Сульфаты	151		
★	1358.	p. Аксу - с. Тюлек, 2,8 км ниже села, 0,2 м	43.140758	74.076106	2015	Сульфаты	228.7		

# Экстра практика

Объединение нескольких таблиц

Очистка повторяющихся значений

Расчет сумм, среднего в цикле

# Задача №1: Есть несколько таблиц с id, значения из которых надо объединить

Таблица с результатами экзаменов и school\_id

<input type="checkbox"/> id	<input type="checkbox"/> full_name	<input type="checkbox"/> math	<input type="checkbox"/> physics	<input type="checkbox"/> chemistry	<input type="checkbox"/> geometry	<input type="checkbox"/> biology	<input type="checkbox"/> geography	<input type="checkbox"/> history	<input type="checkbox"/> english_lang	<input type="checkbox"/> german_lang	<input type="checkbox"/>
1	АБДУЖАПАРОВА АЙДАЙ	30									
2	АЛИШЕРОВА ЭЛЬНУРА	19		11		9	<a href="#">edit</a>				
3	БАХАПОВА АЙСУЛУУ АЙБЕКОВНА	9		8				11			
4	ЖУНУСБЕКОВ ЕРМЕК	10	5								
5	КАДЫРОВ БААТЫР	14					20				
6	КОЖАЛИ КЫЗЫ БУРУЛКАН				8			7			
7	МАСЕЙИТОВ АЮКОЛ		4								
8	НУРМАМАТ УУЛУ ДАНИЯР					8					

Таблица с регионами

<input type="checkbox"/> id	<input type="checkbox"/> title
1	г. Бишкек
2	г. Ош
3	Баткенская область
4	Джалал-Абадская область
5	Иссык-Кульская область
6	Нарынская область
7	Ошская область
8	Таласская область
9	Чуйская область

# Таблица с описанием школ, их id и district\_id

<input type="checkbox"/> id	<input type="checkbox"/> title	<input type="checkbox"/> address	<input type="checkbox"/> phone	<input type="checkbox"/> district_id
1	Учебно-воспитательный комплекс с гимназическим образованием № 9	Молодая Гвардия, 68	64-59-17/20 64-32-07 (0517)23-04-75	1
2	Школа-комплекс лицей № 13	Московская, 174	+ (996 312)318801/38, +(996 312)318754, +(996 555)518858	1
3	Специальная (вспомогательная школа № 22)	Интергельпо, 10	+ (996 312)653716, +(996 772)169149	1
4	Средняя общеобразовательная школа № 25	Гагарина, 2	+ (996 312)541420, +(996 555)249821	1
5	Общеобразовательная школа – гимназия № 31	Павлова, 14	+ (996 312)643038, +(996 312)642807, +(996 555)904641	1
6	Средняя общеобразовательная школа № 40	Каховская, 7	+ (996 312)417506	<a href="#">edit</a> 1
7	Средняя общеобразовательная школа № 42	Алыкула, 1	+ (996 312)451029, + (996 555)070253	1
8	Средняя общеобразовательная школа № 47	Целинная, 16	+ (996 312)544988, +(996 312)544989, +(996 772)412103	1
9	Средняя общеобразовательная школа № 50	Мессароша, 3	+ (996 312)655799, +(996 772)341199	1
10	Средняя общеобразовательная школа № 44	Кривоносова, 11	+ (996 312)35-12-37, +(996 555)34-21-94,	1
11	Средняя общеобразовательная школа № 55	Гагарина 166	+ (996 312)492277, +(996 773)964673	1
12	Средняя общеобразовательная школа № 57	Интергельпо 2,720054	+ (996 312)653736, +(996 312) 653706, +(996 545)951415	1
13	Средняя общеобразовательная школа № 59	Луцихина 35	+ (996 312)647899, +(996 312)648331, +(996 772)248787	1
14	Учебно-воспитательный комплекс гимназия № 68	Токтогула 220	+ (996 312)454463, +(996 312)454529	1
15	Средняя общеобразовательная школа-лицей № 74	Ремесленная 94,720007	+ (996 312)350816, +(996 312)350171, +(996 515)775839	1
16	Средняя общеобразовательная школа № 77	Ашар 4.720077	+ (996 312)412266	1
17	Средняя школа Арча Бешик № 79	Садыбакасова-Эртайлак, 720047	+ (996 312)493026, +(996 543)971168	1
18	Средняя общеобразовательная школа № 80	мкрн.Джал 29 дом 8, 720038	+ (996 312)493026, +(996 543)971168	1
19	Средняя школа № 82	Ала Тоо, 51/1, 720043	+ (996 312)419051(48,49), +(996 543)987069	1
20	Чараплиева школа № 84	Дүй-Оюн, Ашар 6/и 720077	+ (996 212)10127/261 4/008 7791456681	1

# Таблица с id всех районов

<input type="checkbox"/> id	<input type="checkbox"/> region_id	<input type="checkbox"/> title
1	1	Ленинский район
2	1	Октябрьский район
3	1	Первомайский район
4	1	Свердловский район
5	2	г. Ош
6	3	Баткенский район
7	3	Кадамжайский район
8	3	Лейлекский район
9	3	г. Баткен
10	3	г. Кызыл-Кыя
11	3	г. Сулукта
12	4	Аксыйский район
13	4	Ала-Букинский район
14	4	Базар-Коргонский район
15	4	Ноокенский район
16	4	Сузакский район
17	4	Тогуз-Тороуский район
18	4	Токтогульский район
19	4	Чаткальский район
20	4	г. Джалаал-Абад
21	4	г. Кара-Куль
22	4	г. Кок-Джангак
23	4	г. Майлуу-Суу
24	4	г. Таш-Кумыр
25	5	Аксуйский район

# Требования к объединению таблиц

1. Имена датасетов латиницей в одно слово
2. Имена колонок в датасетах латиницей в одно слово

11355 rows Extensions

Show as: [rows](#) [records](#)    Show: [5](#) [10](#) [25](#) [50](#) rows    [« first](#) [< previous](#) [1 - 25](#) [next](#) [» last](#)

<input type="checkbox"/> All	<input type="checkbox"/> id	<input type="checkbox"/> full_name	<input type="checkbox"/> math	<input type="checkbox"/> physics	<input type="checkbox"/> chemistry	<input type="checkbox"/> geometry	<input type="checkbox"/> biology	<input type="checkbox"/> geography	<input type="checkbox"/> history	<input type="checkbox"/> english_lang	<input type="checkbox"/>
1.	1	АБДУЖАПАРОВА АЙДАЙ	30								
2.	2	АЛИШЕРОВА ЭЛЬНУРА	19		11			9			

**Refine** OPEN A power tool for working with messy data.

Create Project  
**Open Project**  
Import Project  
Language Settings

Last modified	Name
today 2:13 PM	<a href="#">school</a>
today 2:13 PM	<a href="#">region</a>
today 2:12 PM	<a href="#">district</a>
today 2:12 PM	<a href="#">ntc_result</a>
yesterday 7:32 PM	<a href="#">VINestate</a>
a week ago	<a href="#">Для серафима</a>
a week ago	<a href="#">для серафима</a>
a week ago	<a href="#">schools_devided</a>

# Добавление названий школ в файл с результатами

1. Edit column - Add column based on this column
2. Введите команду `cell.cross("school", "id")[0].cells["title"].value`
3. Введите имя колонки

The screenshot shows the ntc\_result interface with a modal dialog titled 'Add column based on column school\_id'. In the 'New column name' field, 'school\_title' is entered. The 'set to blank' radio button is selected. The 'Expression' field contains the code `cell.cross("school", "id")[0].cells["title"].value`, which is highlighted with a red arrow. The 'Language' dropdown is set to 'General Refine Expression Language (GREL)'. Below the dialog, a preview table shows the first seven rows of data, all containing the value 'Средняя общеобразовательная школа имени Владимира Ильича Ленина'. At the bottom of the dialog are 'OK' and 'Cancel' buttons.

Значение какой колонки будет импортировано (school.title)

Значение текущей ячейки будет сопоставляться с значением этой колонки (school.id)

Имя таблицы, откуда будут взяты значения

## **Добавление остальных колонок**

1. Добавление id района из таблицы district  
cell.cross("school", "id")[0].cells["district\_id"].value
2. Добавление title района из таблицы  
cell.cross("district", "id")[0].cells["title"].value
3. Добавление id региона из таблицы  
cell.cross("district", "id")[0].cells["region\_id"].value
4. Добавление title района из таблицы  
cell.cross("region", "id")[0].cells["title"].value

## Задача №2: Удаление в колонке повторяющихся строк

Edit cells - Blank down

The screenshot shows a data grid with several columns. The first column has a dropdown arrow icon. The second column is labeled 'school\_id' with a dropdown arrow icon. The third column is labeled 'school\_title' with a dropdown arrow icon. The fourth column is labeled 'district\_id' with a dropdown arrow icon. The fifth column is labeled 'd' (likely 'district\_name') with a dropdown arrow icon.

A context menu is open over the first row of the grid. The menu items are:

- Facet
- Text filter
- Edit cells** (highlighted)
- Edit column
- Transpose
- Sort...
- View
- Reconcile

Under the 'Edit cells' option, a submenu is displayed with the following items:

- Transform...
- Common transforms
- Fill down
- Blank down** (highlighted)
- Split multi-valued cells...
- Join multi-valued cells...
- Cluster and edit...

The data in the grid rows is as follows:

	school_id	school_title	district_id	d
1	394	Средняя общеобразовательная школа имени Владимира Ильича	9	г. Батайск
2		Средняя общеобразовательная школа имени Владимира Ильича		г. Батайск

# Задача №3: Подсчет количества учеников в школах

Строки (rows) vs записи (records)

1. Перенести колонку в начало

	school_id	school_title	district_id	district_name	
Facet		я разовательная	9	г. Баткен	
Text filter		имени			
Edit cells		ири Ильи			
Edit column		▶ Split into several columns...			
Transpose		▶ Add column based on this column...			
Sort...		▶ Add column by fetching URLs...			
View		▶ Rename this column			
Reconcile		▶ Remove this column			
	Средняя общеобразовательная школа им. Владимира Ленина	▶ Move column to beginning			
		▶ Move column to end			
		▶ Move column left			
	Средняя общеобразовательная школа имени Владимира Ильи	▶ Move column right			
	Ленина				

2. Выбрать “записи”

862 records								
Show as: rows records Show: 5 10 25 50 records								
All	school_id	id	full_name	math	physics	chemistry	geometry	biology
1.	394	1	АБДУЖАПАРОВА АЙДАЙ	30				
2			АЛИШЕРОВА ЭЛЬНУРА	19		11		9
3			БАХАЛОВА АЙСУЛУУ АЙБЕКОВНА	9		8		
4			ЖУНУСБЕКОВ ЕРМЕК	10	5			
5			КАДЫРОВ БААТЫР	14				
6			КОЖАЛИ КЫЗЫ БУРУЛКАН					8
7			МАСЕЙТОВ АЮКОЛ		4			
8			НУРМАМАТ УУЛУ ДАНИЯР					

## 2. Добавить колонку

Add column based on column school\_id

New column name: count  
 set to blank  store error  copy value from original column

Expression: 1  
Language: General Refine Expression Language (GREL)  
No syntax error.

Preview:

row	value
1.	394
2.	null
3.	null
4.	null
5.	null
6.	null
...	

## 3. Объединить значения в строках для 1 записи

Facet

Text filter

Edit cells ► Transform...

Edit column ► Common transforms ►

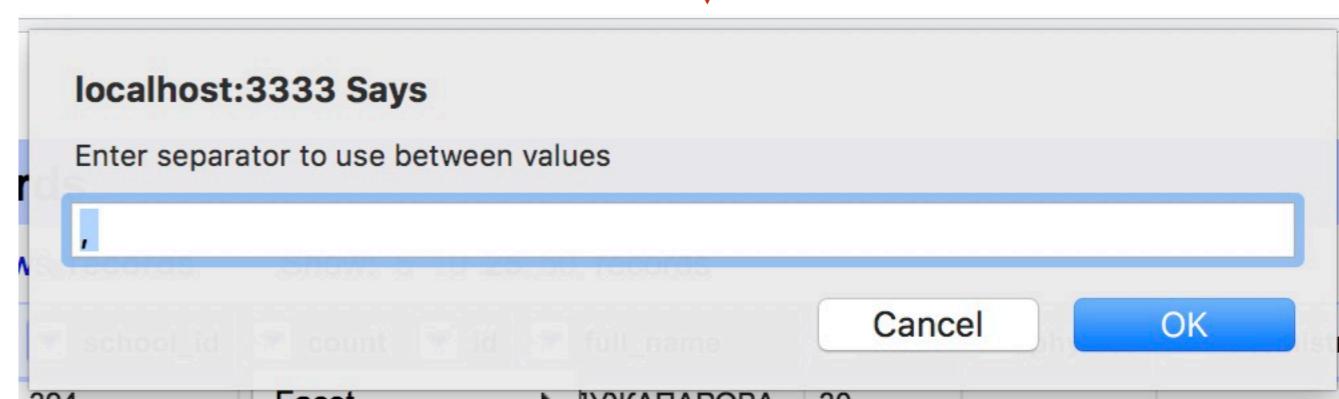
Transpose ► Fill down

Sort... ► Blank down

View ► Split multi-valued cells...

Reconcile ► Join multi-valued cells... ► Cluster and edit...

	school_id	count	id	full_name	math	phy
2.	393	1, 1, 1, 1	12	АКМАТОВА ЗАРИНА	12	
			13	АЛИЕВА ГУЛЗАР	9	
			14	ИДИРИСОВА НУРАЙЫМ	21	
			15	ТАЙИРОВА ЖАЗГУЛ	11	
3.	212	1	16	ЖЭЭНАЛИ К АЙЗИРЕК	3	
4.	222	1	17	АБДИРАНШТ	24	



4. Сложить значения, разделенные запятой  
forEach(value.split(','),v,v.toNumber()).sum()

Custom text transform on column count

Expression      Language General Refine Expression Language (GREL)

```
forEach(value.split(','),v,v.toNumber()).sum()
```

No syntax error.

Preview   History   Starred   Help

row	value	forEach(value.split(','),v,v.toNumber()).sum()
1.	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1	11
2.	null	Error: split expected followed by an
3.	null	Error: split expected followed by an
4.	null	Error: split expected followed by an
5.	null	Error: split expected followed by an

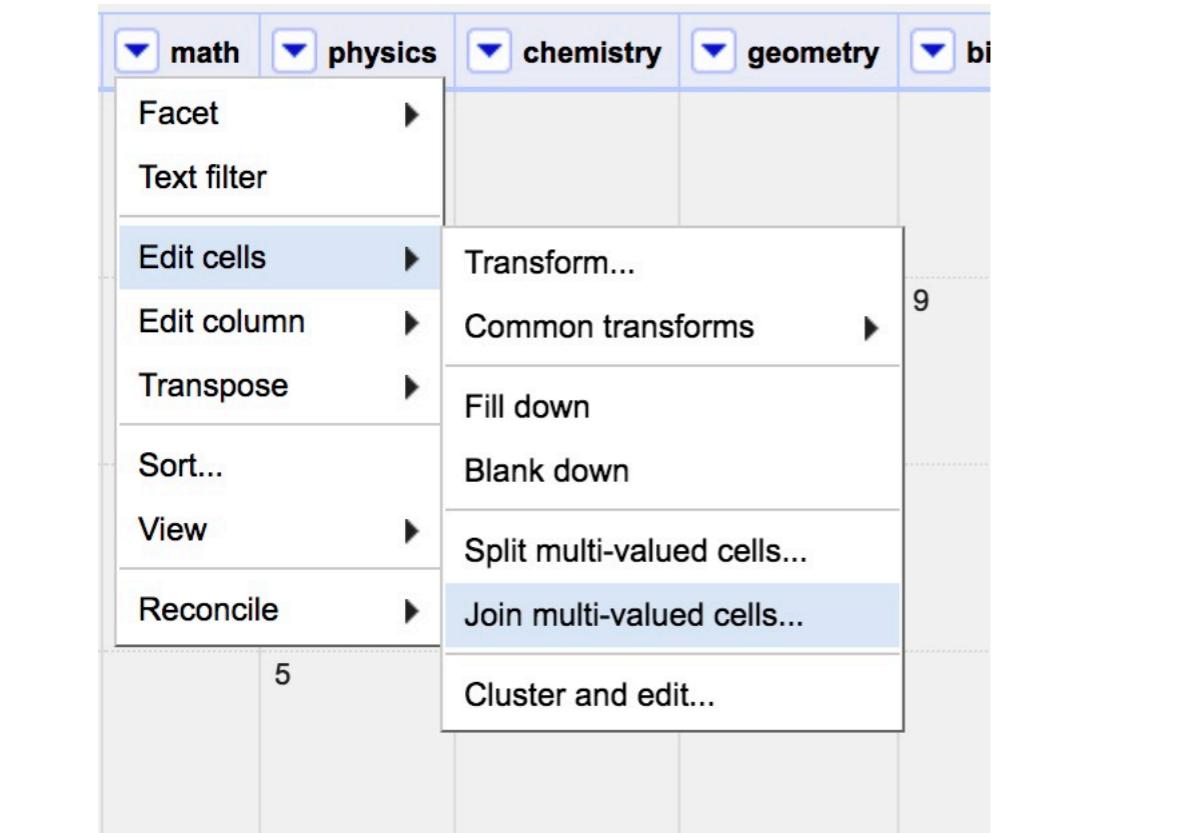
862 records

Show as: rows records   Show: 5 10 25 50 records

All	school_id	count	id	full_name	math	physics	chemistry	g
1.	394	11	1	АБДУЖАПАРОВА АЙДАЙ	30			
2.		2		АЛИШЕРОВА ЭЛЬНУРА	19			11
3.		3		БАХАПОВА АЙСУЛУУ АЙБЕКОВНА	9			8
4.		4		ЖУНУСБЕКОВ ЕРМЕК	10	5		
5.		5		КАДЫРОВ БААТЫР	14			
6.		6		КОЖАЛИ КЫЗЫ БУРУЛКАН				

# Задача №4: Подсчет количества учеников в школах

1. Объединить значения  
в строках для 1 записи



Show as: rows records Show: 5 10 25 50 records

All	school_id	count	id	full_name	math
1.	394	11	1	АБДУЖАПАРОВА АЙДАЙ	30, 19, 9, 10, 14
		2		АЛИШЕРОВА ЭЛЬНУРА	
		3		БАХАЛОВА АЙСУЛУУ АЙГЕКОРДА	

2. Сумма / длина массива

Сумма

`forEach(value.split(','),v,v.toNumber()).sum()`

Длина массива:

`length(value.split(','))`

Custom text transform on column math

Expression

`forEach(value.split(','),v,v.toNumber()).sum()/length(value.split(','))]`

Language General Refine Expression

Preview

History

Starred

Help

row value

`forEach(value.split(','),v,v.toNumber()).sum()/length(value.split(','))`

1. 30, 19, 9, 10, 14

16.4

# Как работает forEach

ЦИКЛ  
(возвращает массив)

создается переменная v, в которую записывается элемент массива (поочередно)

```
forEach(value.split(','), v, v.toNumber()).sum()
```

формируется массив  
элементов для перебора

элементы  
массива  
суммируются

каждый элемент приводится к  
числовому типу, чтобы можно  
было суммировать

# Задача №5: Удаляем лишние строки

1. Переключиться на row
  2. Сделать фасет по пустым
  3. Удалить выбранные строки

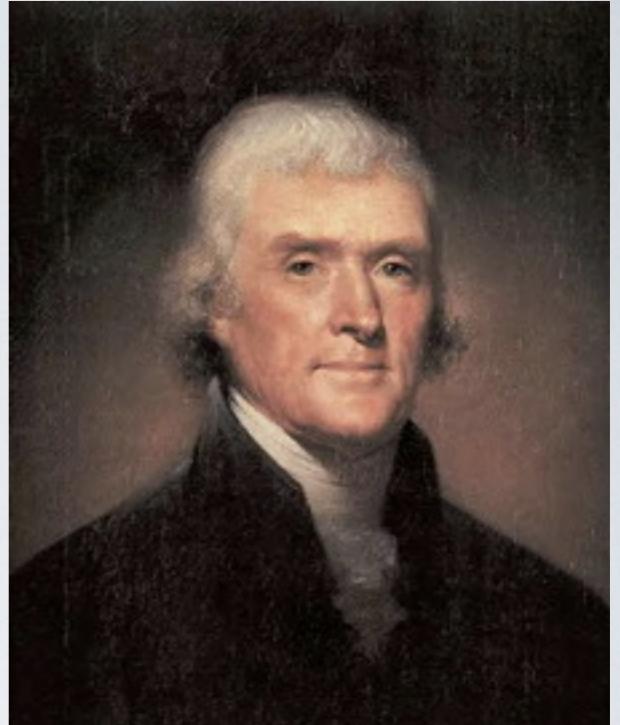
school_id	count	id	full_name	math	physics	chemistry
Facet			Text facet	16.4		
Text filter			Numeric facet			
Edit cells			Timeline facet			
Edit column			Scatterplot facet			11
Transpose			Custom text facet...			
Sort...			Custom Numeric Facet...			
View			Customized facets	Word facet		
Reconcile				Duplicates facet		
		4	ЖУНУСБЕКОВ ЕРМЕК	Numeric log facet		
		5	КАДЫРОВ БААТЫР	1-bounded numeric log facet		
		6	КОЖАЛИ КЫЗЫ БУРУЛКАН	Text length facet		
				Log of text length facet		
				Unicode char-code facet		
				Facet by error		
				Facet by blank		

10493 matching rows (11355 total)				
Show as: <a href="#">rows</a> <a href="#">records</a>		Show: <a href="#">5</a> <a href="#">10</a> <a href="#">25</a> <a href="#">50</a> rows		
All	school_id	count	id	full_name
Transform			2	АЛИШЕРОВА ЭЛЬНУРА
Facet				
Edit rows		Star rows		A
Edit columns		Unstar rows		HA
View		Flag rows		
		Unflag rows		KOB
		4.		
				Remove all matching rows

Все исходные материалы доступны тут:  
<https://github.com/inxaoc/openrefine-workshop>

*«Информация - это валюта демократии»*

*Томас Джефферсон*



***«Данные - это валюта демократии»***