



NETWERK

KENNISCENTRUM GEFACILITEERD DOOR HET **NIOD**

**OORLOGS
BRONNEN**

OpenRefine

17 oktober 2017

Inge van Stokkom | Netwerk oorlogsbronnen

OpenRefine

- Open Refine is een gratis open-source tool waarmee je data kan:
 - Analyseren
 - Schonen & standaardiseren
 - Koppelen van eigen thesaurus aan externe bron
 - Verrijken: aanvullen van eigen data met data uit externe bron
 - “a powerful tool for working with messy data”
 - “Excel voor gestructureerde tekst”
 - stand alone desktop application, geen internet nodig
 - data schonen, normaliseren, transformeren naar verschillende formaten

Zoekresultaat

13 termen gevonden in 0.17

op ruiten gebracht

- op ruiten gebracht door insnijdingen
- op ruiten gebracht in grafiet
- op ruiten gebracht in grafiet of zwart krijt
- op ruiten gebracht in houtskool
- op ruiten gebracht in potlood
- op ruiten gebracht in rood en zwart krijt
- op ruiten gebracht in rood krijt
- op ruiten gebracht in roze krijt
- op ruiten gebracht in wit krijt
- op ruiten gebracht in zwart krijt
- op ruiten gebracht met pen en potlood
- op ruiten gebracht met pen in bruin

Alle sleutels

☐ Inclusief nauwere termen

Term uitklappen

thesaurusterm

beheergegevens

Term

Term

23D41

Engelse term

winter, 'Hyems'; 'Inverno' (Ripa)

Aziatische term

Relaties

Use

Used for

Broader Term

Narrower Term

Related Term

Equivalent Term

Bron en definitie

Bron

Iconclass

URI

<http://iconclass.org/23D41>

Scope Note

Bijzonderheden

Beschrij
Selecte
record:
weerge

Help

plaats:"Amsterdam"

plaats:"AMSTERDAM"

plaats:"A-dam"

plaats:"Adam"

plaats:"A'dam"

plaats:"Amsterdam"

plaats:"Amsterdam"

plaats:"Amsterdam."

plaats:"Amsterdam."

plaats:"AmsterdamM"

plaats:"Amsterdam-N"

plaats:"Amsterdam-oost"

plaats:"Amsterdam-Oost"

plaats:"Amsterdam-Sloten"

plaats:"Amsterdam-Sloten"

plaats:"Amsterdam-zuid"

plaats:"Amsterdam-Zuid"

plaats:"AMSTERDAN"

plaats:"Amsterdm"

plaats:"Amsterfoort"

plaats:"Amstgerdam"

plaats:"Amstredam"

plaats:"Amterdam"

plaats:"Amterdam"

plaats:"Amterdam"

plaats:"Amstelredam"

plaats:"Amstelledamme"

plaats:"Amst."

plaats:"Andenne"

penseel in bruin

penseel in bruin

penseel in bruinzwart

penseel in dekverf

penseel in dekverf en zilververf

penseel in dekverf in blauw en

penseel in dekverf in grijs en

penseel in dekverf in grijs en

penseel in dekverf in kleuren

penseel in dekverf in zwart

penseel in dekverf in zwart en

penseel in dekverf in zwart, rood

penseel in dekverf of olievert

penseel in donkerblauw

penseel in donkerbruin

penseel in donkerbruin en lichte

penseel in donkergrijs

penseel in donkerrode water

penseel in enig goud

penseel in enig rood

penseel in enig wit

penseel in enige kleuren

penseel in felgeel

penseel in geel

penseel in geelbruin

penseel in gewassen inkt

penseel in goud

penseel in goudverf

penseel in grijs

penseel in grijs en bister

penseel in grijs en blauw

NETWERK
OORLOGS
BRONNEN

OpenRefine

2014 11 21 penseel in grijs ... 2014 12 15 penseel in grijs ... +

127.0.0.1:3333/project?project=1730409142656 W - Wikipedia (nl)

Google refine 2014 12 15 penseel in grijs csv Permalink Open... Export Help

Facet / Filter Undo / Redo 14

1476 rows Show as: rows records Show: 5 10 25 50 rows Extensions: Named-entity recognition Freebase RDF

« first < previous 1 - 50 next > last »

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

All	recordnr	TK	MA	ME
☆	1. 352166	etsen, pen in bruin, pen in zwart, penseel in grijs	papier, inkt	ets, bewerkt met pen en penseel, proefdruk
☆	2. 154120	etsen, penseel in grijs	papier	ets
☆	3. 231448	pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	4. 145186	pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	5. 230784	pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	6. 205200	pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	7. 319405	potlood, pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	8. 174257	penseel in grijs, pen	papier, inkt	pen in grijs en zwart
☆	9. 247205	pen in grijs, penseel in grijs, doorgriffelen	papier, inkt	pen in grijs
☆	10. 225450	zwart krijt, pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	11. 236029	potlood, pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	12. 209917	pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	13. 145230	pen in grijs, penseel in grijs	papier, inkt, potlood	op ruiten gebracht in potlood
☆	14. 231542	pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	15. 26529	potlood, pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	16. 444552	loodstift, pen in bruin, penseel in grijs	papier	tekening
☆	17. 145242	pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	18. 323742	potlood, pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	19. 214538	pen in grijs, penseel in grijs, doorgriffelen	papier, inkt	pen in grijs
☆	20. 241551	pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	21. 226813	pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	22. 241679	pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	23. 236619	potlood, pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	24. 232042	penseel in grijs, pen in grijs	papier, inkt	pen in grijs
☆	25. 311902	potlood, pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	26. 230629	pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	27. 232390	pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	28. 54182	pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	29. 227066	pen in grijs, penseel in grijs	papier, inkt	pen in grijs
☆	30. 225835	zwart krijt, pen in grijs, penseel in grijs	papier, inkt	pen in grijs

Features

Screenshot of the SMM_objectName interface showing search results and facets.

Facets and filters

The interface displays 16 matching records (742 total) for the search term "stoe". The results are shown in a table with columns: record - domain, record - term, record - priref, record - date_in, and record - date_out.

Facet / Filter (record - term):

- 1 armstoel 1
- 1 armstoel met fronton 1
- 19 zetels en stoelen 1
- 2 armstoelen 1
- 2 stoelen 1
- 35 stoelen 1
- 6 stoelen 1
- achterstijl van stoel 1
- bureau-leunstoel 1
- ceremoniële stoelen 1
- draagstoelen 1
- gestoelte 1
- spaanse stoel 1
- stoelen 1
- stoelpoot 1
- stoelzitting 1
- (blank) 16

Table Data (Partial):

record - domain	record - term	record - priref	record - date_in	record - date_out
0	1 armstoel	1286	2005-04-01	domini
0	1 armstoel met fronton	1295	2005-04-01	domini
0	19 zetels en stoelen	1290	2005-04-01	domini
0	2 armstoelen	1294	2005-04-01	domini
0	2 stoelen	1306	2005-04-05	domini
0	35 stoelen	1297	2005-04-01	domini
0	6 stoelen	1285	2005-04-01	domini
0	achterstijl van stoel	1240	2005-03-18	domini
0	bureau-leunstoel	1235	2005-03-17	domini
0	ceremoniële stoelen	1220	2005-03-16	domini

FACET/FILTER

127.0.0.1:3333/project?project=1822407736442 — Adlibgebruikersgroep_workshop 3 csv — OpenRefine

Reader

Refine

Adlibgebruikersgroep_workshop 3 csv [Permalink](#)

Open... Export ▾ Help

Facet / Filter Undo / Redo 0

Refresh Reset All Remove All

Gemeente change

9 choices Sort by: name count Cluster

Antwerp 1

Antwerpen 1

Brussel 2

Frasnes-Lez_anvaing 2

Genk 1

Gent 4

Ghent 1

Mechelen 2

Sint-Denijs-Westrem 2

(blank) 8

Facet by choice counts

16 records

Extensions: undefined ▾

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

	Straat	Postcode	Postcode letters	Gemeente	Land	Datum ontwerp	Datum opleverin	Ontwerpers	Architect	Studie
	Meulesteedsesteenweg	9 000							noAarchitecten	
Sint-	Dieudonné Lefèvrestraat	1 020						Ambrosius Francken I	Architecten Tom Thys en Adinda Van Geystelen	Studiebureau Boydens nv
								Théodore Baron		
	Europalaan	3 600							PCP Architects	
oning	Rue Grandrieux	7 910		Frasnes-Lez_anvaing	België	2006	2008	Van Severen, Maarten	GWM architecten	PFC engine bvba
ilding al	Leuvensesteenweg	2 800		Mechelen	België	2006	2007	Ambrosius I Francken	DG architecten	Structor Engineering Mechelen
								Goossens, Willy		
	Oudeleeuwenrui	2 000		Antwerp	België	1997	2006	Brusselmans, Jean	Robbrecht en Daem architecten	Grontmij Vlaanderen
								Daem, Hilde		

UNDO/REDO

127.0.0.1:3333/project?project=1822407736442 — Adlibgebruikersgroep_workshop 3 csv - OpenRefine

Refine OPEN Adlibgebruikersgroep_workshop 3 csv Edit single cell on row 2, column Gemeente Undo

Facet / Filter Undo / Redo 2 Extensions: undefined

Extract... Apply...

Filter:

0. Create project
1. Edit single cell on row 1, column Gemeente
2. Edit single cell on row 2, column Gemeente

16 records

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

	▼ Straat	▼ Postcode	▼ Postcode letters	▼ Gemeente	▼ Land	▼ Datum ontwerp	▼ Datum opleverin	▼ Ontwerpers	▼ Architect	▼ Studie
	Meulesteedsesteenweg	9 000			België				noAarchitecten	
Sint-	Dieudonné Lefèvrestraat	1 020			België	2003 edit	2009	Ambrosius Francken I	Architecten Tom Thys en Adinda Van Geystelen	Studiebureau Boydens nv
								Théodore Baron		
	Europalaan	3 600		Genk	België				PCP Architects	
oning	Rue Grandrieux	7 910		Frasnes-Lez_anvaing	België	2006	2008	Van Severen, Maarten	GWM architecten	PFC engine bvba
ilding al	Leuvensesteenweg	2 800		Mechelen	België	2006	2007	Ambrosius I Francken	DG architecten	Structor Engineering Mechelen
								Goossens, Willy		
	Oudeleuwenrui	2 000		Antwerp	België	1997	2006	Brusselmans, Jean	Robbrecht en Daem architecten	Grontmij Vlaanderen
								Daem, Hilde		

Features

SMM_material - Google Chrome

127.0.0.1:3333/project?project=2217220435084

Cluster & Edit column "record - term"

This feature helps you find groups of different cell values that are very likely to refer to the same concept and merge them. For example, the two strings "New York" and "new York" probably refer to the same person. [Find out more ...](#)

Method: key collision

10 clusters found

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
3	3	<ul style="list-style-type: none">(roze) hout (1 rows)(roze?) hout (1 rows)roze hout (1 rows)	<input checked="" type="checkbox"/>	rozenhout
2	2	<ul style="list-style-type: none">ivoor (1 rows)ivoor? (1 rows)	<input type="checkbox"/>	ivoor
2	2	<ul style="list-style-type: none">marmer (1 rows)marmer? (1 rows)	<input type="checkbox"/>	marmer
2	2	<ul style="list-style-type: none">rode was (1 rows)was (rode) (1 rows)	<input type="checkbox"/>	rode was
2	2	<ul style="list-style-type: none">metaal (1 rows)metaal? (1 rows)	<input type="checkbox"/>	metaal
2	2	<ul style="list-style-type: none">keramiek (1 rows)keramiek? (1 rows)	<input type="checkbox"/>	keramiek
2	2	<ul style="list-style-type: none">was (1 rows)	<input type="checkbox"/>	was

Choices in Cluster

2 — 3

Rows in Cluster

2 — 3

Average Length of Choices

3.5 — 11

Length Variance of Choices

0.5 — 2

google-refine-2....zip

Show all downloads...

Cluster & Edit column "antwoord_bewerkt14"

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 metaphone3

516 clusters found

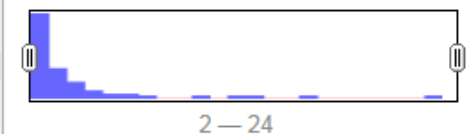
- 13 1987
- plaats:"Rotterdam " (1961 rows)
 - plaats:"Rotterdam Z" (5 rows)
 - plaats:"Rotterdam (N)" (4 rows)
 - plaats:"Rotterda (3 rows)
 - plaats:"Roterdam" (2 rows)
 - plaats:"Rotterdam (Hillgersberg)" (2 rows)
 - plaats:"Rotterdam (Schiebroek)" (2 rows)
 - plaats:"Rotterdam IJsselmonde" (2 rows)
 - plaats:"Rotterddam" (2 rows)
 - plaats:"Rotterdam " (1 rows)
 - plaats:"Rotterdam (Pernis)" (1 rows)
 - plaats:"Rotterdan" (1 rows)
 - plaats:"Rotterdm" (1 rows)

plaats:"Rotterdam "

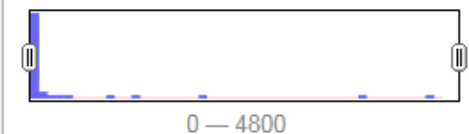
- 11 28
- plaats:"Villers le Temple" (11 rows)
 - plaats:"Flers en Escrebieux" (4 rows)
 - plaats:"Flers Borcelrieux" (2 rows)
 - plaats:"Flers Lille" (2 rows)
 - plaats:"Flers-en Escrelieux" (2 rows)
 - plaats:"Flers-en Escrelieux" (2 rows)

plaats:"Villers le Temple"

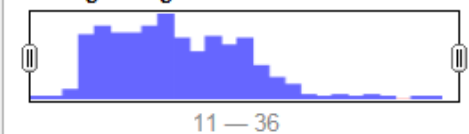
Choices in Cluster



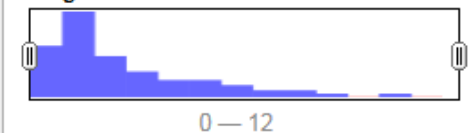
Rows in Cluster



Average Length of Choices



Length Variance of Choices



Select All Unselect All

Merge Selected & Re-Cluster

Merge Selected & Close

Close

Features

Google Refine Expression language

Add column based on column term_clean

New column name:

On error: ☒ set to blank ☐ store error

Expression: `cell.cross("AAT_ENG_NED", "skos:prefLabel@nl").cells["AAT:identifier"]1.value[0]`

Language: Google Refine Expression Language (GREL)

No syntax error.

Preview

23.	acrylverfschilderingen	300181918
24.	afdrukken	300178549
25.	affiches	300027221
26.	afgezaagde menselijke schedelkap	Error: Cannot retrieve field from null
27.	afgietsels	300047806

OK Cancel

waarden aanpassen (*Grel expressions*)

127.0.0.1:3333/project?project=2042098639427&ui=%7B%22facets%22%3A%5B%7B%22c%22%3A%7B%22type%22%3A%22list%22%2C%22name%22%3A%22Type%22%7D%7D%7D

Adlibgebruikersgroep_workshop - OpenRefine

Adlibgebruikersgroep_workshop - OpenRefine

Refine OPEN

Adlibgebruikersgroep_workshop Permalink

Open... Export ▾ Help

Extensions: undefined ▾

Facet / Filter Undo / Redo 9

16 records

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

« first ‹ previous 1 - 10 next › last »

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

Custom text transform on column Materiaal

Expression Language Google Refine Expression Language (GREL) ▾

`value.replace('potlood en inkt', 'potlood, inkt')` No syntax error.

Preview History Starred Help

row	value	value.replace('potlood en inkt', 'potlood, inkt')
1.	potlood en inkt	potlood, inkt
2.	potlood en inkt	potlood, inkt
3.	potlood en inkt	potlood, inkt
4.	potlood en inkt	potlood, inkt
5.	potlood en inkt	potlood, inkt
6.	potlood en inkt	potlood, inkt
7.	potlood en inkt	potlood, inkt

On error ☒ keep original ☐ Re-transform up to 10 times until no change
☐ set to blank
☐ store error

OK Cancel

Fotogalerij

://static.vai.be/sites/default/files/fotos/projecten/noaa_019_o
://static.vai.be/sites/default/files/fotos/projecten/noaa_019_o
://static.vai.be/sites/default/files/fotos/projecten/noaa_019_o
://static.vai.be/sites/default/files/fotos/projecten/noaa_019_o
://static.vai.be/sites/default/files/fotos/projecten/noaa_019_o
://static.vai.be/sites/default/files/fotos/projecten/20110306_J
://static.vai.be/sites/default/files/fotos/projecten/20110306_J
://static.vai.be/sites/default/files/fotos/projecten/20110306_J
://static.vai.be/sites/default/files/fotos/projecten/20110306_J
://static.vai.be/sites/default/files/fotos/projecten/20110306_J
://static.vai.be/sites/default/files/fotos/projecten/carbon_stad
://static.vai.be/sites/default/files/fotos/projecten/carbon_stad
://static.vai.be/sites/default/files/fotos/projecten/carbon_stad
://static.vai.be/sites/default/files/fotos/projecten/20110306_J
://static.vai.be/sites/default/files/fotos/projecten/20110306_J
://static.vai.be/sites/default/files/fotos/projecten/20110306_J
://static.vai.be/sites/default/files/fotos/projecten/20110306_J
://static.vai.be/sites/default/files/fotos/projecten/planckenda
://static.vai.be/sites/default/files/fotos/projecten/planckenda
://static.vai.be/sites/default/files/fotos/projecten/planckenda
://static.vai.be/sites/default/files/fotos/projecten/planckenda
://static.vai.be/sites/default/files/fotos/projecten/planckenda
://static.vai.be/sites/default/files/fotos/projecten/JAAR08_PF
://static.vai.be/sites/default/files/fotos/projecten/JAAR08_PF
://static.vai.be/sites/default/files/fotos/projecten/JAAR08_PF
://static.vai.be/sites/default/files/fotos/projecten/JAAR08_PF
://static.vai.be/sites/default/files/fotos/projecten/JAAR08_PF
://static.vai.be/sites/default/files/fotos/projecten/felix-stad_ar
://static.vai.be/sites/default/files/fotos/projecten/sint_felixpa

Features

SMM_material - Google Refine

127.0.0.1:3333/project?project=1923031089630

Google refine SMM_material Permalink

Facet / Filter Undo / Redo 1

Refresh Reset All Remove All

term_clean: judgment change

2 choices Sort by: name count

matched 30

none 1

Facet by choice counts

term_clean: best candidate's score change reset

92.00 — 100.02

31 rows

Show as: rows record

All term

☆	1.	beukenhout	2572	30
		Choose new match		
☆	2.	Congohout	802	30
		Choose new match		
☆	3.	dennehout	1445	30
		Choose new match		
☆	4.	dennenhout	2333	30
		Choose new match		
☆	5.	divuiti-hout	2088	30
		Choose new match		
☆	6.	ebbehout	1281	30
		Choose new match		
☆	7.	ebbenhout	1190	30
		Choose new match		
☆	8.	eikenhout	113	30
		Choose new match		
☆	9.	eikenhouten paneel	298	30
		Choose new match		
☆	10.	gezwart hout	2167	30
		Choose new match		

Reconciliation Services

Eigen thesaurus
Materiaal = olieverf

AAT
Term = oil paint (paint)
NL term = olieverf
<http://vocab.getty.edu/aat/300015050>

Eigen thesaurus
Materiaal = **olieverf**

AAT
Term = oil paint (paint)
olieverf
<http://vocab.getty.edu/aat/300015050>

Oil paint (paint)
Eigen thesaurus
Materiaal = **olieverf**

<http://vocab.getty.edu/aat/300015050>

AAT
Term = oil paint (paint)
olieverf
<http://vocab.getty.edu/aat/300015050>

Export maken

Open...

Export ▼

Export project

Tab-separated value

Comma-separated value

HTML table

Excel

ODF spreadsheet

Triple loader

MQLWrite

Custom tabular exporter...

Templating...

- Handleidingen:

- http://www.adlibug.nl/wp-content/uploads/2015/05/20150430_Handout_Open_Refine_workshop.pdf
- https://www.dropbox.com/s/4jcve4m7eczjmf2/Handleiding_OpenRefine.pdf?dl=0#
- Geo-info: <http://erfgeo.nl/wat-hoe/openrefine.html>

Gloednieuwe handleiding data opschonen & koppelen:

- https://www.projectcest.be/wiki/Bestand:MoMu_VerbindJeTermenOkt2017.pdf

✓ Met dank aan Alina Saenko van Packed vzw

Inge van Stokkom
l.van.stokkom@niod.knaw.nl



NETWERK

KENNISCENTRUM GEFACILITEERD DOOR HET NIOD

OORLOGS BRONNEN

www.oorlogsbronnen.nl
@Oorlogsbronnen
info@oorlogsbronnen.nl



Ministerie van Volksgezondheid,
Welzijn en Sport

VSBfonds,
iedereen doet mee