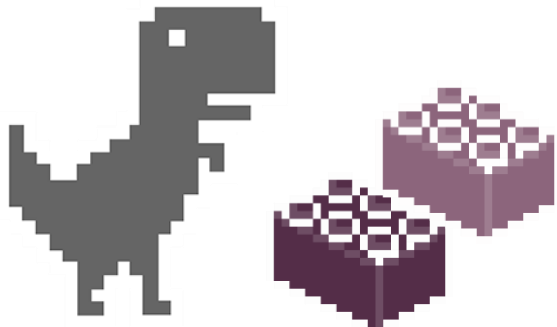
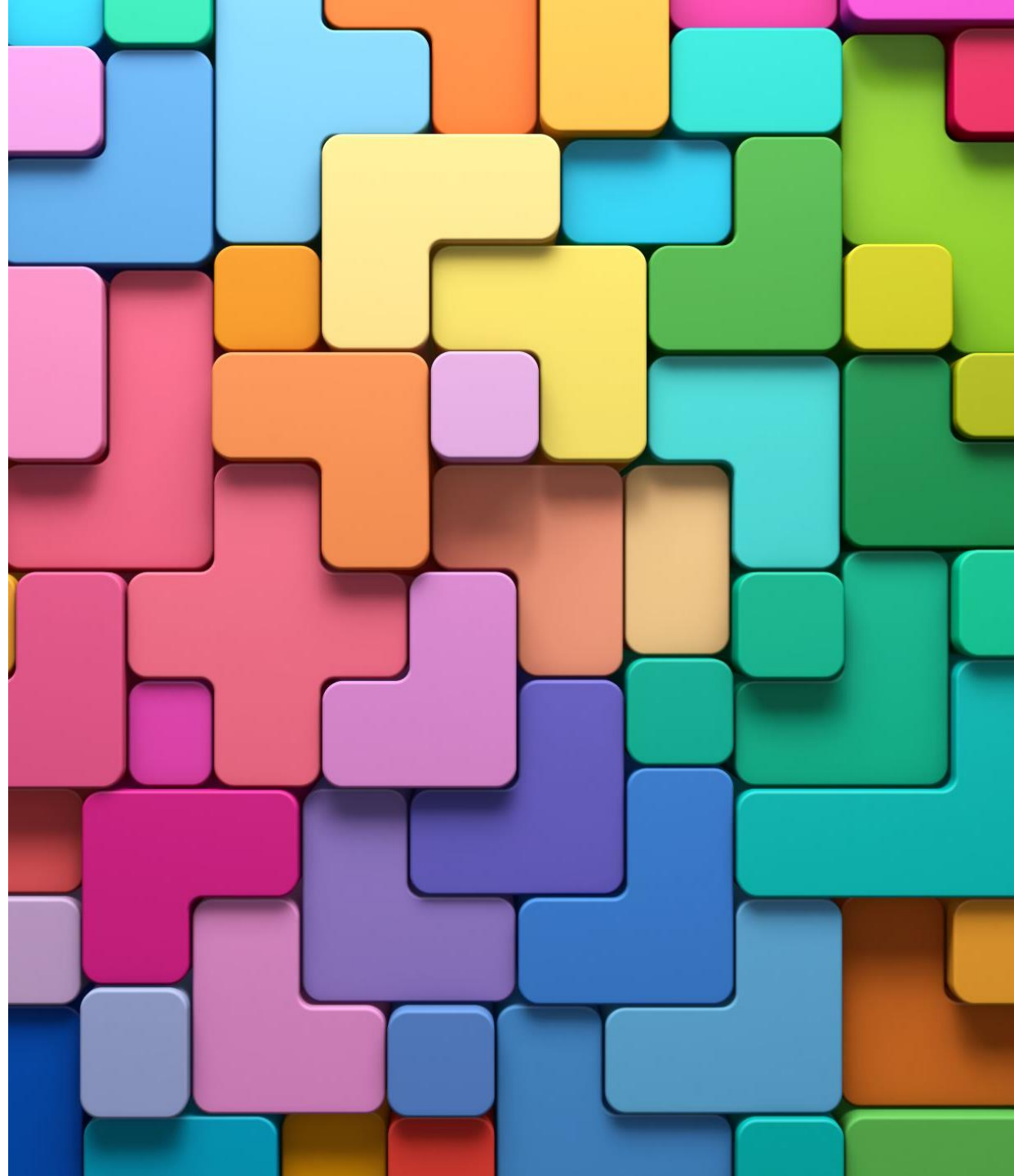

THE THESAURUS BUILDER



Shannon van Muijden
Informatiespecialist
Rijksmuseum



1.000.000

OBJECTS IN THE COLLECTION



500

MUSICAL INSTRUMENTS



2.000

GLASSWARE



2.500

SCULPTURES



3.500

COSTUMES & ACCESSORIES



5.000

PAINTINGS



6.000

FURNITURE



6.500

WEAPONS AND OBJECTS OF MILITARY HISTORY



10.000

TEXTILES



12.000

COINS & TOKENS



15.000

CERAMICS



49.000

HISTORICAL OBJECTS



60.000

DRAWINGS



150.000

PHOTOGRAPHS



500.000

PRINTS



VISIBLE OBJECTS IN MUSEUM



DIVIDED OVER CENTURIES



8.000 (1%)

<16TH

17TH

18TH

19TH

20TH

5%

18%

20%

38%

19%

THESAURUS

Een ontsluitingsmiddel waarbij unieke concepten door hiërarchische, equivalente en associatieve relaties verbonden worden, een semantisch netwerk voor het toegankelijk maken en koppelen van (collectie)gegevens. (bron: wikipedia)



RIJKSMUSEUM THESAURUS

- Termen met URI's naar de AAT / Wikidata
 - Termen met broader terms
 - Termen met broader terms die corresponderen met de AAT hiërarchie
 - Termen met broader terms waarbij de BT in een ander domein zit dan de term
 - Termen met een URI maar geen broader term
 - Termen zonder URI en zonder broader term
 - *Termen zonder enige relaties aan andere records*
-

FOCUS

1. Termen met een URI maar zonder broader term (wezen)
 2. Termen waarvan de broader term in een ander domein zit dan het domein van de term (verkeerde ouder)
-

WEZEN

Thesaurusterm

Term

Term ambtsketen

Soort term objectnaam

Status niet gedefinieerd

Relaties

Use

Used for

Broader Term

Narrower Term

Related Term

Equivalent Term

DOEL

Identificeer termen uit je eigen thesaurus die als broader term kunnen dienen voor de wezentermen op basis van de hiërarchie van de AAT (match op URI-niveau)



EXPORT

1. Recordnummer

2. Term

3. AAT URI

(zet ID uit URI in aparte kolom)

<input checked="" type="checkbox"/>	rec...	term	PID_overige.URI	term.s...	term.st...	broader_term	narrower_term
<input checked="" type="checkbox"/>	1	bord (vaatwerk)	http://vocab.getty.edu/aat/300042991	objectnaam	niet gedefini...	schaal	diep bord
<input checked="" type="checkbox"/>	6	diepte	http://vocab.getty.edu/aat/300072633	afmeting	niet gedefini...		
<input checked="" type="checkbox"/>	13	schenking	http://vocab.getty.edu/aat/300138913	verwervings...	niet gedefini...		
<input checked="" type="checkbox"/>	16	schilderij	http://vocab.getty.edu/aat/300177435	objectnaam	niet gedefini...		
<input checked="" type="checkbox"/>	18	olieverf	http://vocab.getty.edu/aat/300015050	materiaal	niet gedefini...		
<input checked="" type="checkbox"/>	19	doek	http://vocab.getty.edu/aat/300014078	materiaal	niet gedefini...		
<input checked="" type="checkbox"/>	21	hoogte	http://vocab.getty.edu/aat/300055644	afmeting	niet gedefini...		
<input checked="" type="checkbox"/>	22	breedte	http://vocab.getty.edu/aat/300055647	afmeting	niet gedefini...		
<input checked="" type="checkbox"/>	24	brons	http://vocab.getty.edu/aat/300010957	materiaal	niet gedefini...	metaal	
<input checked="" type="checkbox"/>	27	meubilair	http://vocab.getty.edu/aat/300037680	objectnaam	niet gedefini...		ameublement
<input checked="" type="checkbox"/>	30	metaal	http://vocab.getty.edu/aat/300010900	materiaal	niet gedefini...		brons
<input checked="" type="checkbox"/>	31	triplex	http://vocab.getty.edu/aat/300012849	materiaal	niet gedefini...	hout	
<input checked="" type="checkbox"/>	32	draaiend bewerken	http://vocab.getty.edu/aat/300053158	techniek	niet gedefini...		
<input checked="" type="checkbox"/>	36	papier	http://vocab.getty.edu/aat/300014109	materiaal	niet gedefini...		aluminiumpapier
<input checked="" type="checkbox"/>	37	hout	http://vocab.getty.edu/aat/300011914	materiaal	niet gedefini...	plantaardig mater...	abachi
<input checked="" type="checkbox"/>	39	plastic	http://vocab.getty.edu/aat/300014570	materiaal	niet gedefini...		
<input checked="" type="checkbox"/>	44	glas	http://vocab.getty.edu/aat/300043230	objectnaam	niet gedefini...	drinkgerei	bekerglas
<input checked="" type="checkbox"/>	45	lood (gewicht)	http://vocab.getty.edu/aat/300011022	objectnaam	niet gedefini...		plaatlood
<input checked="" type="checkbox"/>	46	verf	http://vocab.getty.edu/aat/300015029	materiaal	niet gedefini...		acrylverf
<input checked="" type="checkbox"/>	49	lengte	http://vocab.getty.edu/aat/300055645	afmeting	niet gedefini...		
<input checked="" type="checkbox"/>	52	gereedschap	http://vocab.getty.edu/aat/300024841	objectnaam	niet gedefini...		avegaar
<input checked="" type="checkbox"/>	54	zilver	http://vocab.getty.edu/aat/300011029	materiaal	niet gedefini...	metaal	bladzilver
<input checked="" type="checkbox"/>	67	draagteken	http://vocab.getty.edu/aat/300193994	objectnaam	niet gedefini...	penning	pelgrimsinsigne
<input checked="" type="checkbox"/>	69	persen	http://vocab.getty.edu/aat/300053136	techniek	niet gedefini...	vormen	kalanderen
<input checked="" type="checkbox"/>	81	been	http://vocab.getty.edu/aat/300011798	materiaal	niet gedefini...		gewei
<input checked="" type="checkbox"/>	86	textiel	http://vocab.getty.edu/aat/300231565	objectnaam	niet gedefini...		borduurwerk

EXCEL

	A	B	C	D
1	recordnummer	term	PID_overige.URI	AAT ID
2	1	bord (vaatwerk)	http://vocab.getty.edu/aat/300042991	300042991
3	6	diepte	http://vocab.getty.edu/aat/300072633	300072633
4	13	schenking	http://vocab.getty.edu/aat/300138913	300138913
5	16	schilderij	http://vocab.getty.edu/aat/300177435	300177435

CTRL+E

	A	B	C	D
1	recordnummer	term	PID_overige.URI	AAT ID
2	1	bord (vaatwerk)	http://vocab.getty.edu/aat/300042991	300042991
3	6	diepte	http://vocab.getty.edu/aat/300072633	
4	13	schenking	http://vocab.getty.edu/aat/300138913	
5	16	schilderij	http://vocab.getty.edu/aat/300177435	

OPENREFINE

GREL

```
'http://vocabsservices.getty.edu/AATService.asmx/AATGetParents?subjectID='+value
```

Levert je de gehele hiërarchie van de AAT; alle termen boven de term

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

« first < previous 1 next > last »

▼ preref	▼ term	▼ PID_other.URI	▼ AAT ID	▼ AAT parents	▼ Parent strings
16	schilderij	http://vocab.getty.edu/aat/300177435	300177435	<pre><?xml version="1.0" encoding="utf-8"?> <Vocabulary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://vocabsservices.getty.edu/Schemas/AAT/AATGetParents.xsd"> <Subject Subject_ID="300177435"> <Parent_Relationships> <Preferred_Parent> <Parent_Subject_ID>300033638</Parent_Subject_ID> <Relationship_Type>Parent/Child</Relationship_Type> <Historic_Flag>Current</Historic_Flag> <Parent_String>paintings by form [300033638], paintings (visual works) [300033618], visual works by material or technique [300191091], visual works (works) [300191086], Visual Works (hierarchy name)\$http://vocab.getty.edu/aat/300179869, Visual and Verbal Communication (hierarchy name)\$http://vocab.getty.edu/aat/300264552, Objects Facet\$http://vocab.getty.edu/aat/300264092</Parent_String> <Hier_Rel_Type>Genus/Species-BTG</Hier_Rel_Type> </Preferred_Parent> </Parent_Relationships> </Subject> </Vocabulary></pre>	<p>paintings by form\$http://vocab.getty.edu/aat/300033638, paintings (visual works)\$http://vocab.getty.edu/aat/300033618, visual works by material or technique\$http://vocab.getty.edu/aat/300191091, visual works (works)\$http://vocab.getty.edu/aat/300191086, Visual Works (hierarchy name)\$http://vocab.getty.edu/aat/300179869, Visual and Verbal Communication (hierarchy name)\$http://vocab.getty.edu/aat/300264552, Objects Facet\$http://vocab.getty.edu/aat/300264092</p>
117	serviesgoed	http://vocab.getty.edu/aat/300236054	300236054	<pre><?xml version="1.0" encoding="utf-8"?> <Vocabulary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://vocabsservices.getty.edu/Schemas/AAT/AATGetParents.xsd"> <Subject Subject_ID="300236054"> <Parent_Relationships> <Preferred_Parent> <Parent_Subject_ID>300133146</Parent_Subject_ID> <Relationship_Type>Parent/Child</Relationship_Type> <Historic_Flag>Current</Historic_Flag> <Parent_String>sets (groups) [300133146], object groupings by general context [300241508], object groupings [300241507], Object Groupings and Systems (hierarchy name) [300241489], Objects Facet [300264092]</Parent_String> <Hier_Rel_Type>Genus/Species-BTG</Hier_Rel_Type> </Preferred_Parent> </Parent_Relationships> </Subject> </Vocabulary></pre>	<p>sets (groups)\$http://vocab.getty.edu/aat/300133146, object groupings by general context\$http://vocab.getty.edu/aat/300241508, object groupings\$http://vocab.getty.edu/aat/300241507, Object Groupings and Systems (hierarchy name)\$http://vocab.getty.edu/aat/300241489, Objects Facet\$http://vocab.getty.edu/aat/300264092</p>
154	persoonlijk object	http://vocab.getty.edu/aat/300238982	300238982	<pre><?xml version="1.0" encoding="utf-8"?> <Vocabulary xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://vocabsservices.getty.edu/Schemas/AAT/AATGetParents.xsd"> <Subject Subject_ID="300238982"> <Parent_Relationships> <Preferred_Parent> <Parent_Subject_ID>300239170</Parent_Subject_ID> <Relationship_Type>Parent/Child</Relationship_Type></pre>	<p>equipment by context\$http://vocab.getty.edu/aat/300239170, equipment\$http://vocab.getty.edu/aat/300122241, Tools and Equipment (hierarchy name)\$http://vocab.getty.edu/aat/300022238, Furnishings and Equipment (hierarchy name)\$http://vocab.getty.edu/aat/300264551, Objects Facet\$http://vocab.getty.edu/aat/300264092</p>

VERGELIJK AAT- PARENTS MET TERMEN UIT JE EIGEN THESAURUS

A	B	C	D	E	F	G
16	schilderij	http://voc	3E+08	<?xml vers	paintings by form\$	
117	serviesgo	http://voc	3E+08	<?xml vers	sets (groups)\$http:	
154	persoonlij	http://voc	3E+08	<?xml vers	equipment by cont	
155	servetring	http://voc	3E+08	<?xml vers	accessory containe	
157	drijfwerk	http://voc	3E+08	<?xml vers	visual works by ma	
179	strook	http://voc	3E+08	<?xml vers	object genres by fo	
205	jak	http://voc	3E+08	<?xml vers	coats (garments)\$f	
255	merklap	http://voc	3E+08	<?xml vers	embroidery (visual	
276	poppenhu	http://voc	3E+08	<?xml vers	temporary alphabe	
280	klerenkas	http://voc	3E+08	<?xml vers	case furniture\$http	
297	kaas	http://www.wikidata.org/entity/Q10943				
305	vuurtest	http://www.wikidata.org/entity/Q2094867				
313	droogrek	http://voc	3E+08	<?xml vers	drying racks (wupp	
316	stellingka	http://voc	3E+08	<?xml vers	objec	
317	aanrechtk	http://voc	3E+08	<?xml vers	temp	
319	mangeltaf	http://voc	3E+08	<?xml vers	surfa	
324	keukentaf	http://voc	3E+08	<?xml vers	table	
328	messenb	http://voc	3E+08	<?xml vers	cases (containers):	
331	fornuis	http://voc	3E+08	<?xml vers	culinary devices for	
331		http://www.wikidata.org/entity/Q3748418				
337	aanrechtk	http://voc	3E+08	<?xml vers	furniture compone	
350	bouilloire	http://voc	3E+08	<?xml vers	teakettles\$http://vc	
352	wandklok	http://voc	3E+08	<?xml vers	clocks by form\$http	
372	tobbe	http://voc	3E+08	<?xml vers	containers by form	
382	deegroller	http://voc	3E+08	<?xml vers	culinary equipmen	
393	vuurpot	http://voc	3E+08	<?xml vers	heating equipment	
399	oliekruik	http://voc	3E+08	<?xml vers	vessels (containers	
411	karaf	http://voc	3E+08	<?xml vers	vessels for serving	
413	treeft	http://voc	3E+08	<?xml vers	protective and life-	

	A	B	C	D	E	F
1	A	B	C			
2		1 bord (va	http://vocab.getty.edu/aat/300042991			
3		6 diepte	http://vocab.getty.edu/aat/300072633			
4		13 schenking	http://vocab.getty.edu/aat/300138913			
5		16 schilderij	http://vocab.getty.edu/aat/300177435			
6		18 olieverf	http://vocab.getty.edu/aat/300015050			
7		19 doek	http://vocab.getty.edu/aat/300014078			
8		21 hoogte	http://vocab.getty.edu/aat/300055644			
9		22 breedte	http://vocab.getty.edu/aat/300055647			
10		24 brons	http://vocab.getty.edu/aat/300010957			
11		27 meubilair	http://vocab.getty.edu/aat/300037680			
		30 metaal	http://vocab.getty.edu/aat/300010900			
		31 triplex	http://vocab.getty.edu/aat/300012849			
		32 draaiend	http://vocab.getty.edu/aat/300053158			
		36 papier	http://vocab.getty.edu/aat/300014109			
16		37 hout	http://vocab.getty.edu/aat/300011914			
17		39 plastic	http://vocab.getty.edu/aat/300014570			
18		44 glas	http://vocab.getty.edu/aat/300043230			
19		45 lood (gew	http://vocab.getty.edu/aat/300011022			
20		46 verf	http://vocab.getty.edu/aat/300015029			
21		49 lengte	http://vocab.getty.edu/aat/300055645			
22		52 gereedsch	http://vocab.getty.edu/aat/300024841			
23		54 zilver	http://vocab.getty.edu/aat/300011029			
24		67 draagteke	http://vocab.getty.edu/aat/300193994			
25		69 porse	http://vocab.getty.edu/aat/300053136			

✨ THIS IS
WHERE THE
MAGIC
HAPPENS ✨

URI
MATCHING
ONLY!

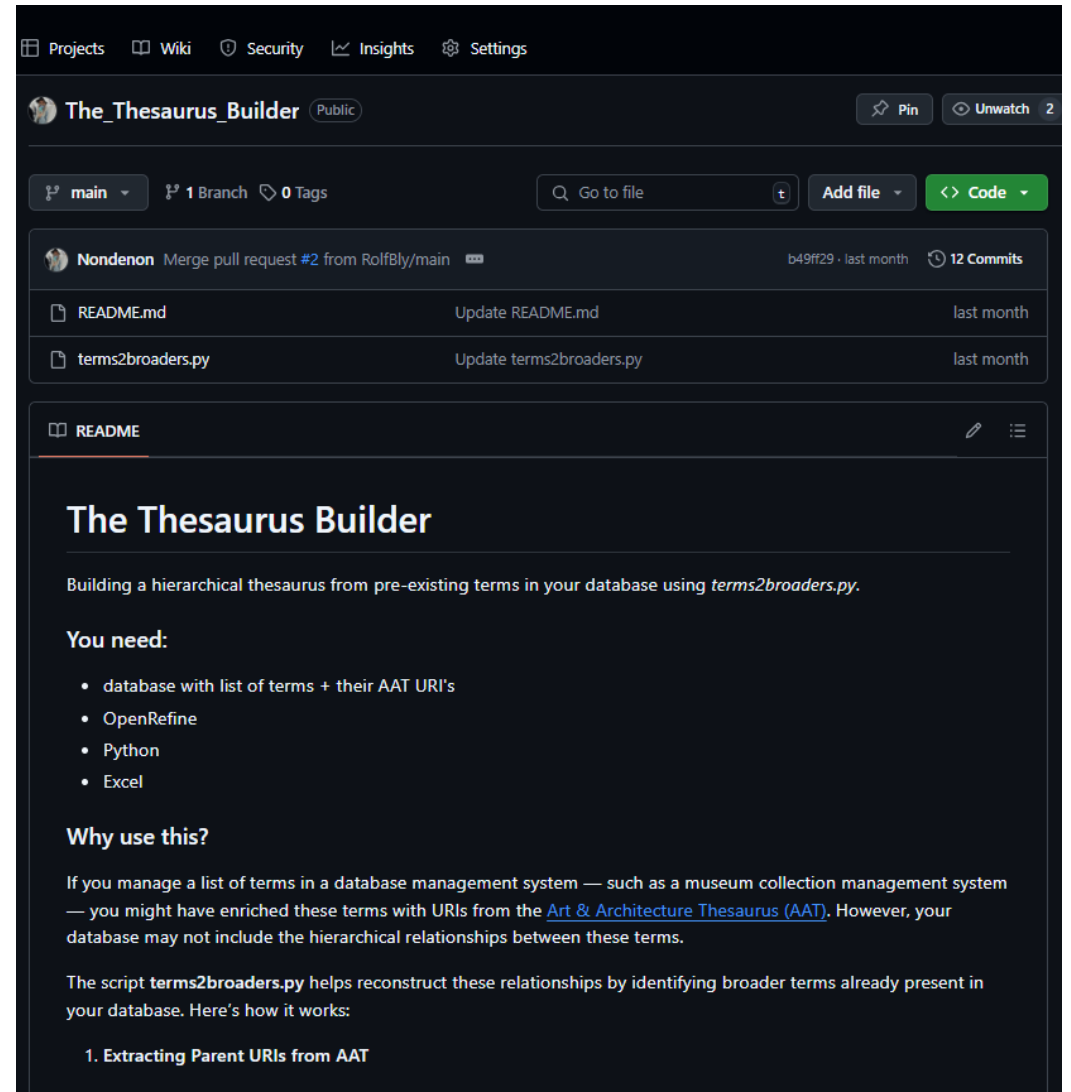
```
terms2broaders.py x 20250110-beschrijver.dat x new 58 x new 60 x new 62 x validate_domains.py x new 56 x new 59 x export_tn_naar_4P.dat x import
11     ... # Ensure necessary columns exist
12     ... required_sheet1_cols = {'B', 'F'}
13     ... required_sheet2_cols = {'A', 'B', 'C'}
14     ...
15     ... if not required_sheet1_cols.issubset(sheet1.columns):
16     ...     raise ValueError("Sheet1 must contain columns B and F.")
17     ... if not required_sheet2_cols.issubset(sheet2.columns):
18     ...     raise ValueError("Sheet2 must contain columns A, B, and C.")
19     ...
20     ... # Extract URIs from Sheet1 Column F
21     ... def extract_uris(cell):
22     ...     if pd.isna(cell):
23     ...         return []
24     ...     return [entry.split('$')[1] for entry in cell.split(',') if '$' in entry]
25     ...
26     ... sheet1['URIs'] = sheet1['F'].apply(extract_uris)
27     ...
28     ... # Match URIs exactly and include all matching terms from Sheet2
29     ... def match_uris(uris):
30     ...     matches = []
31     ...     for uri in uris:
32     ...         matching_rows = sheet2.loc[sheet2['C'] == uri, ['A', 'B']]
33     ...         matches.extend([f"{row['A']}${row['B']}${uri}" for _, row in matching_rows.iterrows()])
34     ...     return matches
35     ...
36     ... sheet1['Matches'] = sheet1['URIs'].apply(match_uris)
37     ...
38     ... # Expand matches into separate columns
39     ... max_matches = sheet1['Matches'].apply(len).max()
40     ... for i in range(max_matches):
41     ...     sheet1[f'Match_{i+1}'] = sheet1['Matches'].apply(lambda x: x[i] if i < len(x) else None)
42     ...
43     ... # Drop intermediate columns and save the resulting dataframe
44     ... result = sheet1.drop(columns=['URIs', 'Matches'])
45     ... output_file = 'terms_with_matches.xlsx'
46     ... result.to_excel(output_file, index=False)
47     ... print(f"You've successfully build a thesaurus! Output saved to {output_file}")
48     ...
49     ... # Run the function
50     ... process_excel('terms_read_broaders.xlsx')
```

RESULTING OPTIONAL HIERARCHY BASED ON TERMS THAT ARE ALREADY IN THE RMA THESAURUS

48194	ambtsketen	kostuum accessoires\$http://vocab.getty.edu/aat/300209273	kostuum\$http://vocab.getty.edu/aat/300178802
67717	ballotagezak	kostuum accessoires\$http://vocab.getty.edu/aat/300209273	kostuum\$http://vocab.getty.edu/aat/300178802
66946	boekethouder	kostuum accessoires\$http://vocab.getty.edu/aat/300209273	kostuum\$http://vocab.getty.edu/aat/300178802
9724	borstlap	kostuum accessoires\$http://vocab.getty.edu/aat/300209273	kostuum\$http://vocab.getty.edu/aat/300178802
9723	borststuk	kostuum accessoires\$http://vocab.getty.edu/aat/300209273	kostuum\$http://vocab.getty.edu/aat/300178802
48183	bretels	kostuum accessoires\$http://vocab.getty.edu/aat/300209273	kostuum\$http://vocab.getty.edu/aat/300178802
550367	bruidssjaal	kostuum accessoires\$http://vocab.getty.edu/aat/300209273	kostuum\$http://vocab.getty.edu/aat/300178802
6588	corsage	kostuum accessoires\$http://vocab.getty.edu/aat/300209273	kostuum\$http://vocab.getty.edu/aat/300178802
5322	dasslip	kostuum accessoires\$http://vocab.getty.edu/aat/300209273	kostuum\$http://vocab.getty.edu/aat/300178802
7215	devant-de-gorge	kostuum accessoires\$http://vocab.getty.edu/aat/300209273	kostuum\$http://vocab.getty.edu/aat/300178802
35304	epaulet	kostuum accessoires\$http://vocab.getty.edu/aat/300209273	kostuum\$http://vocab.getty.edu/aat/300178802
5681	fontaneldoek	kostuum accessoires\$http://vocab.getty.edu/aat/300209273	kostuum\$http://vocab.getty.edu/aat/300178802

GITHUB

The Thesaurus Builder github repository



The screenshot shows the GitHub interface for the repository 'The_Thesaurus_Builder'. At the top, there are navigation links for Projects, Wiki, Security, Insights, and Settings. The repository name 'The_Thesaurus_Builder' is displayed with a 'Public' label, along with 'Pin' and 'Unwatch' buttons. Below this, the current branch is 'main', with '1 Branch' and '0 Tags' indicated. A search bar for files and buttons for 'Add file' and 'Code' are visible. A recent commit by 'Nondenon' is shown, including a merge pull request and a list of files updated: 'README.md' and 'terms2broaders.py'. The 'README' file is selected, showing its content. The README title is 'The Thesaurus Builder'. The text describes the project's purpose: 'Building a hierarchical thesaurus from pre-existing terms in your database using terms2broaders.py.' It lists requirements under 'You need:' (database with terms and AAT URIs, OpenRefine, Python, Excel) and explains the project's value under 'Why use this?'. The first step of the process, '1. Extracting Parent URIs from AAT', is partially visible at the bottom.

Projects Wiki Security Insights Settings

The_Thesaurus_Builder Public Pin Unwatch 2

main 1 Branch 0 Tags Go to file Add file Code

Nondenon Merge pull request #2 from RolfBly/main b49ff29 · last month 12 Commits

README.md Update README.md last month

terms2broaders.py Update terms2broaders.py last month

README

The Thesaurus Builder

Building a hierarchical thesaurus from pre-existing terms in your database using `terms2broaders.py`.

You need:

- database with list of terms + their AAT URI's
- OpenRefine
- Python
- Excel

Why use this?

If you manage a list of terms in a database management system — such as a museum collection management system — you might have enriched these terms with URIs from the [Art & Architecture Thesaurus \(AAT\)](#). However, your database may not include the hierarchical relationships between these terms.

The script `terms2broaders.py` helps reconstruct these relationships by identifying broader terms already present in your database. Here's how it works:

1. Extracting Parent URIs from AAT

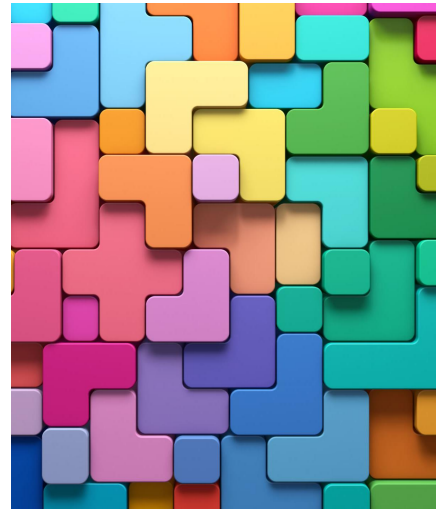
GEDACHTES HIEROVER...

- Bespaart een hoop tijd: je hoeft niet zelf een thesaurus uit te vinden
 - Kies zelf in hoeverre je de hiërarchie wil overnemen (hoeveel stappen). Wat past bij jouw collectie?
 - Wat te doen met termen die geen broader term match hebben gekregen? Deze termen hebben broader terms nodig die nu niet voorkomen in de eigen thesaurus.
-

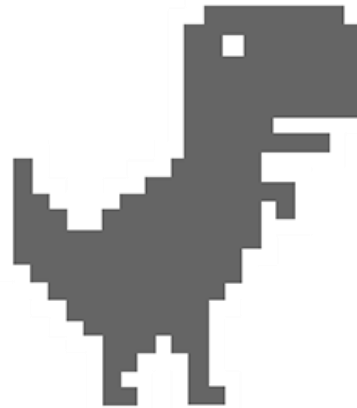
BRUIKBAARHEID

- Betere zoekfunctionaliteiten intern en op de website:
 - ‘geef me alle wapens en alles gerelateerd aan wapentuig in de collectie van het Rijksmuseum’
 - ‘geef me alle objecten met ‘kant’ als techniek. Levert nu 709 objecten op, maar zou 4236 objecten opleveren als deze term ingezet wordt als broader term voor andere kanttechnieken.





BEDANKT!
VRAGEN?



Shannon van Muijden
Informatiespecialist
Rijksmuseum
