

Creating and Using Virtual Corpora at English-Corpora.org (June 2021)

In 5-10 seconds, you can search through millions of texts to create “Virtual Corpora” (VC) on virtually any topic – *nuclear power, thyroid, vaccine, basketball, muggles, fake news, ransomware*, etc. In 2-3 seconds more, you can create “keywords” from this corpus, and search for words, phrases, collocates, etc in the VC.

Creating Virtual Corpora

Search interface showing the search term "nuclear power" and the "Find matching strings" button. The "Texts/Virtual" option is selected under "Search by date". The "FIND TEXTS" button is highlighted, and the "Create corpus" link is also visible.

(From main Search page)

1. Click Select Texts/Virtual
2. Click Find texts
3. Enter the desired word or phrase
4. Click "Find matching strings"

On the following “results” page

SAVE LIST (DON'T SHOW BY TEXT)

HELP	TEXT	IN 1
1	<input checked="" type="checkbox"/>	IN 1
2	<input checked="" type="checkbox"/>	US 1

1. Click Save List and then 2. Name the list
- (Optional) 3. Choose the number of texts in the VC, or (de)select specific texts
- A. See information on each of the texts, including relevance

(Optional) SAVE AS: nuclear OR ADD TO: --SELECT-- SUBMIT RESET

HELP	TEXT	# WORDS	# HITS ↓	RELEVANCE ↓	PE
1	11-12-15: COUNTERCURRENTS.ORG	9791	66	6,740.9	
2	US 13-02-11: EXTREMETECH	29288	49	1,673.0	
3	US 19-08-05: MARGINAL REVOLUTION	10584	49	4,629.6	
4	US 20-03-11: REUTERS	1947	46	23,626.1	
5	ALL 12-03-10: LINKS INTERNATIONAL	3180	41	12,803.1	

Using and searching your Virtual Corpora

You can see a list of all of the Virtual Corpora you have created (potentially hundreds or thousands of texts). Just one corpus is shown in the screenshot below.

12 [trash] [lock] NUCLEAR 100 375,166 NOUN VERB ADJ ADV N+N ADJ+N 0 h

1. Click on a part of speech to see Keywords from the Virtual Corpus
2. See the number of texts and the number of words in each Virtual Corpus
3. Delete corpora, temporarily deactivate them, rename them, move/add texts between corpora, etc

In 1-2 seconds, you can see a list of hundreds of keywords in your Virtual Corpus, ranked by relevance.

NUCLEAR [375,166 WORDS, 100 TEXTS] (3.0% OF TOTAL) **NOUN** VERB ADJ ADV N+N ADJ+N [ALL CORPORA] [SAVE LIST](#)

HELP	WORD (CLICK FOR CONTEXT)	FREQ	# TEXTS	SPECIFIC		ENTIRE CORPUS	EXPECTED
				FREQ	TEXTS		
				45	10		
1	REACTOR	1395	87	1.7	3	28,609	845.1
2	THORIUM	69	17	1.4		1,623	47.9
3	PLUTONIUM	142	24	1.4		3,505	103.5
4	FISSION	62	21	1.0		2,018	59.6
5	RADIOACTIVITY	57	24	0.9		2,066	61.0

1. Click to see the words in context, within your Virtual Corpus (see below)
2. Information on the number of hits and number of texts in the VC that have the word
3. Adjust the "specificity" of the keywords; lower Frequency or Texts for more specific words, increase for more general, and then click 'Specific'.

Keyword in Context (from list of Keywords, above)

CLICK FOR MORE CONTEXT				NEW	SAVE	TRANSLATE	ANALYZE
1	12-07-30 US	Yale Environment 360	main types of breeding cycles: * The fast breeder reactor's fast neutrons can fission even actinides with even neutron numbers. Even numbered actinid				
2	17-06-03 MY	The Star Online	of teaching nuclear science, a field that is wider than just utilising the nuclear fission process for electricity generation. # Even prior to 2009, Universiti K				
3	13-02-11 US	ExtremeTech	a fission reactor in my life. I agree with my country's stance on fission . But I think my government would also refuse to use fusion power, if				
4	13-06-04 US	Scientific America	hand, these proponents urge a new look at other technologies for generating electricity from fission : the fast-breeder reactors that work like a perpetua				
5	12-07-30 US	Yale Environment 360	In principle, breeder fuel cycles can recycle and consume all actinides, leaving only fission products. So, after several hundred years, the waste's radioac				
6	11-03-13 US	ProPublica	or with liquid fuel). The fission products actually go away relatively rapidly. Fission product inventory due to 1 MW operation for one year would be abou				
7	19-09-09 IN	downtoearth.org.in	the protest by the People's Movement Against Nuclear Energy (PMANE) caused high fission and forced most staff to leave, leaving a skeleton force to ov				

Searching within your Virtual Corpora

[List](#) [Chart](#) [Collocates](#) [Compare KWIC](#)

ADJ material 1

Search by date

Sections Texts/Virtual Sort/Limit

covid
electron
immigrant
investment
nuclear 2
poverty
refugee

[Create corpus](#)
[Edit corpora](#)
[Find keywords](#)
[Refresh list](#)

1. Search for words, phrases, construction, collocates, synonyms, etc
2. Select the Virtual Corpus

HELP	ALL FORMS (SAMPLE): 100 200 500	FREQ
1	RADIOACTIVE MATERIAL	20
2	NUCLEAR MATERIAL	13
3	FISSILE MATERIAL	8
4	RAW MATERIAL	3
5	WEAPONS-GRADE MATERIAL	2

You can also compare the frequency of words and phrases in the different Virtual Corpora that you have created. And you can also use metadata for the texts (source, title of text, date, author, genre, sub-genre, etc) as you create the Virtual Corpora. [Click here](#) for more details.

With the corpora from English-Corpora.org, you can create Virtual Corpora on virtually any topic within 4-5 seconds, and then search within that corpus or find keywords in 2-3 seconds more – even with corpora containing billions of words and tens of millions of texts.