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, ransomeware, 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

(From main Search page)
1. Click Select Texts/Virtual
2. Click Find texts
3. Enter the desired word or phrase, e.g., etc
4. Click Find matching strings

On the following “results” page

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

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 corpora). Just one corpus is shown in the screenshot below.

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.

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

Keyword in Context (from list of Keywords, above)

- 12-07-30 US: Yale Environment 360
- 17-06-03 NY: The Star Online
- 13-02-11 US: ExtremeTech
- 13-06-04 US: Scientific America
- 12-07-30 US: Yale Environment 360
- 11-03-13 US: ProPublica
- 19-09-09 IN: downtownearth.org.in

Searching within your Virtual Corpora

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

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.