

## Full-text corpus data

introduction format/samples corpora related sites get data

You can now download the **NOW corpus** for offline use, including monthly updates via a subscription. In total, this is about **14.7 billion words** of data that you can have on your own machine. There's nothing else like it.

The NOW corpus contains data from **25,785,733** texts from online magazines and newspapers in 20 different English-speaking countries from 2010 to the current time (see sources). At **14.7 billion words**, it is by far the largest corpus (of any language) that is available in full-text format. Most importantly, the corpus grows by 7-8 million words of data *each day*. This translates to about 190-220 million words each month and more than 2 billion words each year. (See totals by month.) If you're interested in what's going on in English up to and including right now, this is by far the best corpus available.

When you purchase the full-text data from NOW, you get all of the data up through the previous month. You can also purchase an annual subscription, which will give you data for the next 12 months (typically about 2.2 billion words each year).

For example, if you purchase both datasets on 15 March 2022, you would have the data from January 2010 - February 2022 (which was released on 1 March 2022; see the samples in #2 below), and an annual subscription would give you the data for one more year: March 2022 - February 2023.

		Time period	Size	Samples
1	One-time purchase	2010 - month of purchase	(Currently) About <b>14.7</b> billion words	Database, WordLemPoS, Text, Sources, Lexicon
2	Annual subscription	The 12 month period after	month of purchase	Database, WordLemPoS, Text, Sources, Lexicon

If you purchase just #1 above, it would be the price of one corpus, and there would be a discount for purchasing both corpora (#1 and #2) at the same time.

Note also that the monthly updates will be released at the beginning of the following month. You will be notified by email as soon as the update is available, and you will have ten days to download the data.