

Automatic Terminology Extraction with Sketch Engine

Sketch Engine is one of the best out-of-the-box terminology extraction tools available on the market. Where tools like AntConc and the term extraction feature of memoQ work best with languages that are based on the Latin alphabet, [Sketch Engine](#) works with over [90 languages and 30 writing systems](#), though the development of features like part-of-speech tagging, lemmatization, and terms remains outstanding for many of these languages.

Sketch Engine has a somewhat overwhelming amount of training materials available, including videos via their [YouTube channel](#) and a great deal of [user guides](#).

For the purposes of this seminar, review the user guides associated with these corpus management and terminology extraction features:

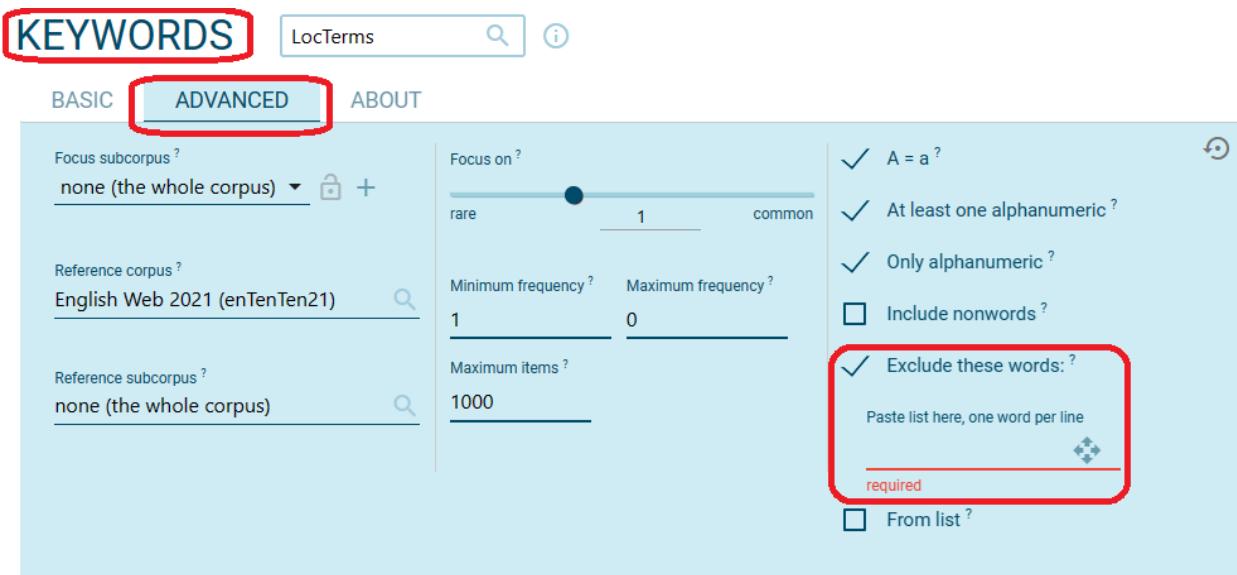
- [11 - Manage a corpus](#)
- [8 - Keywords](#)
- [Build a corpus from the web](#)

The screenshot shows the Sketch Engine User Guide dashboard. At the top, there's a navigation bar with links for Home, News & Events, Pricing, Guide (which is currently selected), About us, Contact, and a search icon. Below the navigation, a breadcrumb trail says "You are here: Home / Guide". The main area is titled "Sketch Engine User Guide" and contains two main sections: "general screens" and "corpus tools". The "general screens" section includes links for dashboard, select corpus, corpus info page, manage corpus, and settings. The "corpus tools" section includes links for word sketch, sketch difference, thesaurus, concordance, parallel concordance, wordlist, n-grams, keywords & terms, bilingual term extraction, trends, and OneClick Dictionary. On the right side of the dashboard, there's a sidebar titled "RECENTLY USED CORPORA" showing a list of recently used corpora, and a footer with a link to "Master the interface in 2 days!".

Sketch Engine screenshot showing the Sketch Engine User Guide dashboard

Keywords (Terminology Extraction)

The Keywords (terminology extraction) feature extracts both frequent single-words and multi-word terms for human validation. A stop list can be added to Keyword extraction by clicking **Keywords > Advanced**, then selecting the checkbox for **Exclude these words**, and finally copying a list of stop words into the dialogue box.



Sketch Engine screenshot showing the Advanced Keywords dialogue where stopwords can be added

Once keyword results have been generated, the context for the term can be accessed by clicking into the 3 dots beside each candidate term. You can access the context for that term in both the focus corpus (your corpus) and the reference corpus (used to determine relative frequency).

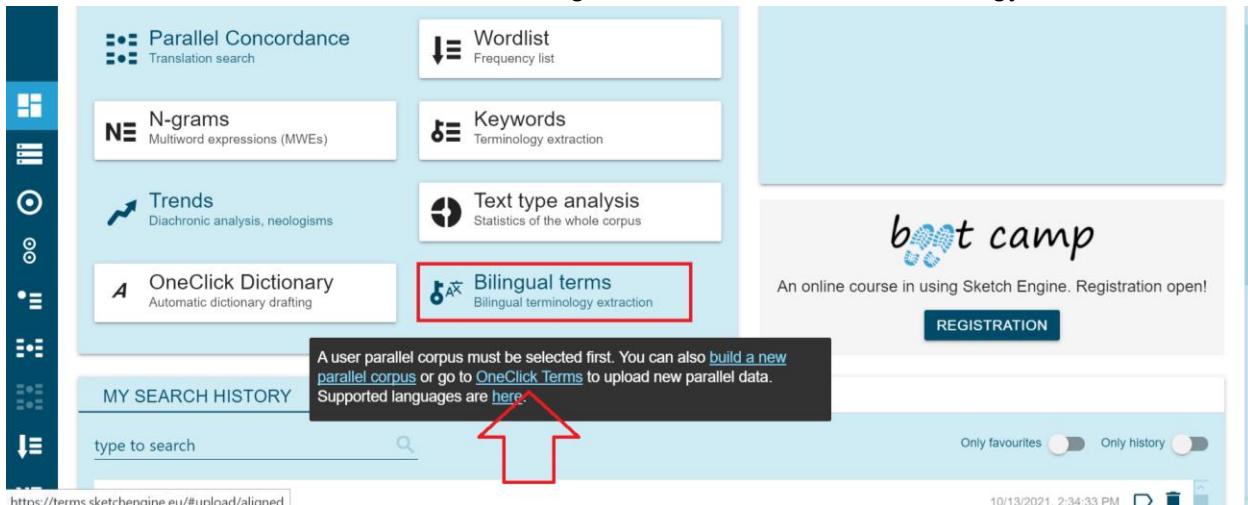
The screenshot shows the 'Keywords' page with the 'MULTI-WORDS TERMS ✓' tab selected. It displays a list of terms with their frequencies and contexts. One term, 'lipstick composition', has its context menu highlighted with a red box. The page also shows navigation controls for rows per page (50), page number (1-50 of 291), and page navigation.

Word	Word	Word	Word
1 lipstick composition	14 wheat germ oil	27 fruit extract	40 microcrystalline wax
2 cosmetic composition	15 germ oil	28 long term stability	41 ascorbic acid
3 glycine soja	16 corn oil	29 term stability	42 pdf version
4 germ extract	17 possible use	30 lemon oil	43 canola oil
5 inventive composition	18 wheat germ	31 seed extract	44 useful life
6 ascorbyl palmitate	19 prunus oil	32 boric acid	45 sodium chloride
7 liquid cosmetic enhancer	20 euphorbia cerifera	33 citric acid	46 sesame oil
8 cosmetic enhancer	21 tocopherol acetate	34 vegetatum equiline	47 castor oil
9 liquid enhancer	22 capric triglyceride	35 comprising vegetatum equiline	48 beneficial moisturizing effect
10 hydroxylated lanolin	23 soybean oil	36 virgin prunus oil	49 moluccana seed
11 lauroyl lysine	24 sodium borate	37 virgin prunus	50 soft rich feel
12 liquid carrier	25 sorbic acid	38 comprising vegetatum	
13 total weight	26 cream base	39 lip composition	

Sketch Engine screenshot showing the generated Keywords page with tabs for single words and multi-word terms

OneClick Terms

Check out the OneClick Terms feature that generates a nice list of terminology too.



Sketch Engines screenshot showing where to find the Bilingual Terms button

Getting Logged Onto Sketch Engine

Subscriptions to Sketch Engine are affordable and can be canceled any time. You can also access a [free 30-day trial](#) to try out Sketch Engine's features.