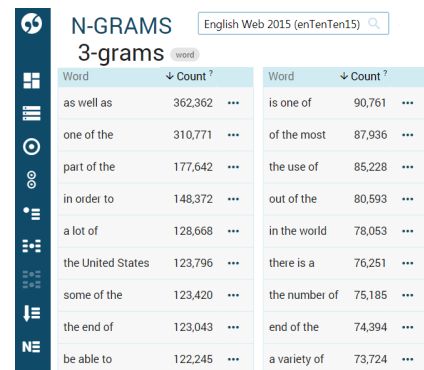# SKETCH ENGINE

for IT and NLP

Sketch Engine is software for analysing large amounts of texts (text corpora) and for generating linguistic databases. It contains data of billions of words in 90+ languages and lets users upload their own data for analysis.

◖ language corpora and linguistic databases

Enormous corpora in 90+ languages with a size of up to 40 billion words can be used to produce various linguistic databases: dictionaries, frequency lists, n-gram databases, collocations databases, thesauri and many others for use in IT and NLP applications.

►List of corpora bit.ly/2Ve1mN1

►List of languages bit.ly/2Li3v60

N-GRAMS 3-grams word — English Web 2015 (enTenTen15)			
Word	↓Count?	Word	↓Count?
as well as	362,362	is one of	90,761
one of the	310,771	of the most	87,936
part of the	177,642	the use of	85,228
in order to	148,372	out of the	80,593
a lot of	128,668	in the world	78,053
the United States	123,796	there is a	76,251
some of the	123,420	the number of	75,185
the end of	123,043	end of the	74,394
be able to	122,245	a variety of	73,724

◖ text analytics API

API is included with each user account and provides access to the complete functionality and all corpora in Sketch Engine.

►Learn bit.ly/2LjpGst

◖ language models

Language models and word embeddings in dozens of languages can be generated to your specifications.

►View embeddings bit.ly/2IUrA0u

◖ language tools

Familiarize yourself with the complete Sketch Engine functionality in the web interface. All features are also available via our API.

►Learn bit.ly/2IuGy8Z

learn

Quick Start Guide www.sketchengine.eu/quick-start-guide
Training videos bit.ly/2HY0gys   Face-to-face training bit.ly/2Vr8K3m

log in

Use your institutional login details on
www.sketchengine.eu

Institutional login

works on mobiles