Manual for GDEX

Read How to sort sentences by GDEX in Sketch Engine? if you want to quickly start using GDEX.

See Syntax of GDEX configuration files if you want to write your own configuration.

Introduction

GDEX is an abbreviation for "Good Dictionary EXamples". It is a system for evaluation of sentences with respect to their suitability to serve as dictionary examples. Its typical usage is in sorting sentences so that good examples do not have to be searched for in hundreds of unusable sentences. Especially in web-based corpora it can effectively rule out sentences that are poor candidates as dictionary examples and it offers the lexicographers a selected set of sentences with a higher chance of containing a good sentence.

The exact way of sorting of the sentences can be adapted for various languages or even various purposes by changing parameters in a GDEX configuration file. Custom based configurations can be created and evaluated partly with tools directly provided with GDEX or some external applications.

For fees for a user-specific GDEX configuration, please contact inquiries@sketchengine.co.uk

GDEX in Sketch Engine

Sketch Engine uses GDEX to sort sentences in Concordances and in TickBox Lexicography (TBL). Sorting of concordances using GDEX has to be activated in View Options otherwise the concordance is shown in corpus order. Sorting in TickBox Lexicography is always activated and the number of sorted sentences is 300 (if available) for each collocation.

Currently only the default GDEX configuration is available for all users. It is, however, possible to create and use custom GDEX configurations.

Adding user GDEX configurations

The online interface provides a special page for uploading user configurations to Sketch Engine. Local installations need to register gdex configurations manually. Currently the upload page is only accessible in beta, as user configurations can cause errors if not set up properly. Once the GDEX configuration is uploaded that version of GDEX becomes available in the View Options dialog to be selected. Since the configurations do not have to be corpus/language dependent, it is up to the user to use them with correct corpora.

Uploaded user configurations can also be shared with other users or user groups.

Selecting from a list of GDEX configurations

If more than one GDEX configuration is available a drop-down list appears in View Options. The selected configuration is used for sorting in both Concordance View and TickBox Lexicography.
Comparing two different GDEX configurations

Similarly, if more than one GDEX configuration is available another drop-down list appears at the TBL result page, where the user can select an alternative configuration that will be used for sorting the same set of sentences side-by-side with the first GDEX configuration.

GDEX Configuration Files

Technically, GDEX assigns the sentences with a score and sorts them from the best to the worst. The assigned value is composed from results of a variety of classifiers that measure various features. The exact set of measured features and the way they are combined together is specified by the GDEX configuration files. Each configuration file is a description of the sentence evaluation function.

See Syntax of GDEX configuration files for full reference on how to write GDEX configuration.