D2.5
Validation and Benchmarking Services

Author(s): Lenka Bajčetić, Seung-Bin Yim, Tanja Wissik, John P. McCrae, Atul Kumar Ojha
Date: 29/7/2022

european lexicographic infrastructure

elexis european lexicographic infrastructure

H2020-INFRAIA-2016-2017
Grant Agreement No. 731015
ELEXIS - European Lexicographic Infrastructure

D2.5 Validation and benchmarking services

Deliverable Number:	D2.5
Dissemination Level:	PU
Delivery Date:	29/July/2022
Version:	1.0
Author(s):	Lenka Bajčetić, Seung-Bin Yim, Tanja Wissik, John P. McCrae, Atul Kumar Ojha

Project Acronym: ELEXIS

Project Full Title: European Lexicographic Infrastructure

Grant Agreement No.:   731015

Deliverable/Document Information

Project Acronym: ELEXIS

Project Full Title: European Lexicographic Infrastructure

Grant Agreement No.:   731015

Document History

Version Date	Changes/Approval	Author(s)/Approved by
V0.1 14/June/2022	First Draft	Lenka Bajčetić, Tanja Wissik
V0.5 25/June/2022	Second Draft	Lenka Bajčetić, Seung-Bin Yim, Tanja Wissik, John P. McCrae, Atul Kumar Ojha
V1.0 26/June/2022	Final Version	Incorporating feedback from all WP2 members.

D2.5 Validation and benchmarking services

D2.5 Validation and benchmarking services

Executive Summary

In this deliverable we describe the work on the Validation and Benchmarking Service within ELEXIS. The tool is called EDiE, which stands for ELEXIS Dictionary Evaluation Tool and is aimed to assist users with context-dependent qualitative assessment of linguistic resources by creating "dictionary profiles" which can be easily compared and evaluated by the end user (Bajčetić et al., 2022). The tool is available as open-source software on GitHub[1].

1 Introduction

The Validation and Benchmarking Service within ELEXIS is called EDiE which stands for ELEXIS Dictionary Evaluation Tool. In the following we describe the motivation behind the development of EDiE and the architecture and implementation of the tool. Since there are a multitude of types of users and user scenarios, we needed to create a generic dictionary assessment tool which would work best under these ambiguous circumstances. Since we cannot make any definitive assumptions regarding the goal of the end users and their priorities regarding dictionary quality, we have decided to create a tool which creates something similar to a "dictionary profile" and would leave the final evaluation to the end users based on these "dictionary profiles". In order to provide information within context, EDiE allows the user to compare the "dictionary profile" with other relevant (by language or type) dictionary within the ELEXIS Infrastructure as an "aggregated" evaluation.

2 Architecture and Implementation

EDiE is created as a Python project and deployed as a REST service. Within the ELEXIS infrastructure, it is interconnected with Lexonomy[2], an open-source dictionary writing and dictionary publishing system: in order to run the assessment on a particular dictionary, the user must first upload it to Lexonomy. Additionally, EDiE relies on all the dictionaries stored within Lexonomy to create the aggregated evaluation.
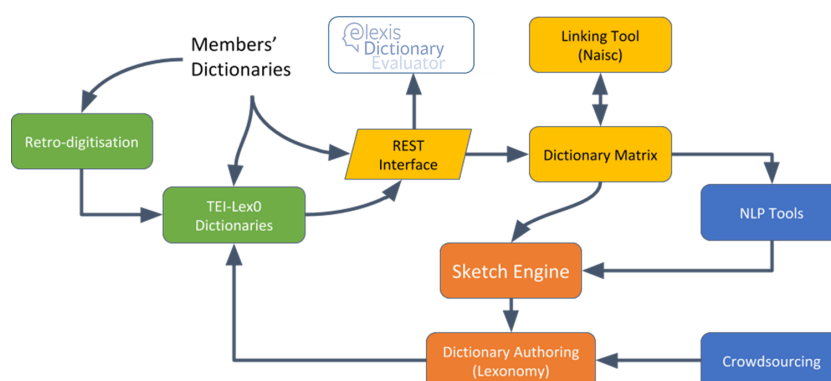

Fig. 1: EDiE within the ELEXIS Infrastructure (adapted from McCrae et al., 2019[3])

---

[1] https://github.com/elexis-eu/edie

[2] More information on Lexonomy can be found in the publication Michal Měchura. 2017. Introducing Lexonomy: an open-source dictionary writing and publishing system. In proceedings of eLex2017, https://www.lexonomy.eu/docs/elex2017.pdf or on www.elex.is.

[3] McCrae, J. P., Tiberius, C., Khan, A. F., Kernerman, I., Declerck, T., Krek, S., Monachini, M., and Ahmadi,S. (2019). The ELEXIS interface for interoperable lexical resources. In Proceedings of the sixth biennial conference on electronic lexicography (eLex) ,pages 642–659, Sintra, Portugal.

When a new dictionary is uploaded to Lexonomy, the evaluation is run automatically and stored alongside the dictionary on the cloud. If the content or metadata of the dictionary are changed, the evaluation is re-run. In this way we make sure the evaluations can be accessed quickly on-demand, because creating an assessment profile can take a while, especially for large dictionaries because the model iterates through all the entries. Creating aggregated evaluations beforehand is not viable because they depend on the user's input regarding which dictionaries to compare with. However these evaluations are not as time-consuming because they are done using the pre-made dictionary profiles.

The main component of EDiE are the evaluators, separated on three levels of evaluation. The microstructure of dictionary entries is evaluated using the Entry evaluator module. This module iterates through all the entries and creates a statistical overview of a typical/average dictionary entry. Most of the implemented metrics are within the Entry evaluator. Our rationale behind this decision was that reading through all the entries is too impractical for the end user, so we thought that such a statistical preview would prove the most helpful.

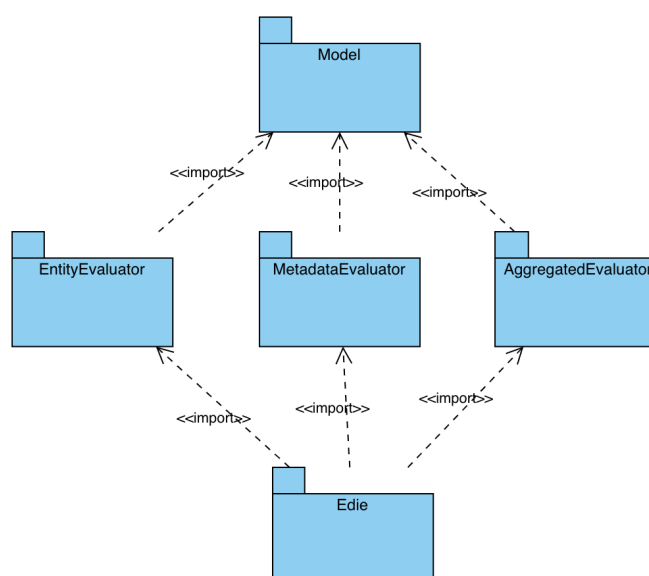In comparison, analysing the metadata can be done quite quickly by the user.

Fig. 2: EDiE package diagram (Yim et al., 2022)

Once a user selects a specific resource, they can send a parameterized request to EDIE using the REST API, and quickly get an end report in JSON format which consists of the resource's content statistics, metadata with the missing data pointed out, formatting errors, and the aggregation profile if requested. In Fig. 3 you can see an example end report.

```
{
  "endpoint": "http://lexonomy.elex.is/",
  "available": true,
  "dictionaries": {
    "elexis-dsl-moth": {
      "entry_report": {
        "errors": [
          "Part of speech value was invalid: ['sb.']",
          "Part of speech value was invalid: ['adv.']",
          "Part of speech value was invalid: ['pr\u00e6p.']",
          "Part of speech value was invalid: ['sb.']",
          "Part of speech value was invalid: ['udr\u00e5bsord']",
        ]
      },
      "metadata_report": {
        "errors": [
          "License not specified"
        ],
        "metric count": 18,
        "total metrics": 112,
        "sizeOfDictionary": 93832
      }
    },
    "elexis-oeaw-jakob": {
      "entry_report": {
        "errors": [
          "No type of entry",
          "No type of entry",
```

Fig. 3: Excerpt of an EDiE End Report (Bajčetić et al., 2022)

3 Metrics

There are multiple levels that are considered for evaluation and each has a different metrics: entry level (entry metrix), single dictionary level (metadata metrix) and at the level of all relevant dictionaries within the ELEXIS infrastructure (aggregated metrics) (see Fig. 3).

Metadata Metrics	Entry Metrics	Aggregated Metrics
SizeOfDictionary	formsPerEntry	DictionarySize:
Recency	sensesPerEntry	min
Digital Availability	definitionPerSense	max
Licence	definitionPerEntry	mean
	DefinitionLengthPerEntryBaCharacter	median
	DefinitionLengthPerEntryByToken	
	DefinitionLengthPerSenseByCharacter	
	DefinitionLengthPerSenseByToken	
	FormatsPerEntry	
	JsonSupportedEntries	
	TeiSupportedEntries	
	OntolexSupportedEntries	
	JsonCoverage	
	TeiCoverage	
	OntolexCoverage	

Tab. 1: Different metrics used for EDiE (Yim et al., 2022)

3.1 Entry Metrics

The entry metics comprises the following:
- Number of forms per entry
- Number of senses per entry
- Number of definitions per sense
- Number of definitions per entry
- Average definition length per entry (in characters)
- Average definition length per entry (in tokens)
- Average definition length per sense (in characters)
- Average definition length per sense (in tokens)
- Average number of formats per entry
- Number of JSON supported entries
- Number of TEI supported entries
- Number of OntoLex supported entries
- Percentage of JSON coverage
- Percentage of TEI coverage
- Percentage of OntoLex coverage

As shown in the list of entry metrics above, most of them focus on providing a statistical overview of the average structure, sense granularity and type of information which can be found in the entries providing the user a quick insight into the respective resources, because it is hard for the user to go through all the entries in order to assess, if a certain resource is suitable for a specific use-case. Additionally, these metrics show to what extent a certain format is supported within a particular dictionary. Due to the possibilities provided by the ELEXIS toolkit, the user can easily transform a part of a dictionary to a different format, hence we provide information for all three supported formats as a percentage, in the case where only some entries are provided in a particular format.

3.2 Metadata Metrics

At the single dictionary level the metadata of the dictionary is used for evaluation e.g. licence, size (see Tab. 1). The metadata is defined by Dublin Core and those fields used by the whole Elexis infrastructure. When evaluating dictionary metadata, it is important to keep in mind the limits of automatic metadata assessment. Creating a tool which can automatically assess the quality and accuracy of metadata is almost impossible, but we can create a completeness check (see Yim et al. 2022, Bajčetić et al., 2022) . Not all information within the metadata is considered equally important, which means that some pieces of information missing will be shown as errors in the assessment, while others will not. For example, while the metadata can contain the full address of the publisher, this information is not considered as important as the publisher's name itself.
With the metadata metrics, the user can check the completeness of the metadata provided, for example if the licence is specified. For example, In the end report example provided in Fig. 3, the licence is not specified, therefore this specific resource might not be usable for several use cases.

3.3 Aggregated Metrics

The aggregated metrics is used for the context specific comparative evaluation based on a specific aspect of a dictionary, e.g., size of the dictionary in question compared to the other relevant dictionaries within the ELEXIS infrastructure. Language and genre of dictionaries is used to filter out non-relevant dictionaries.
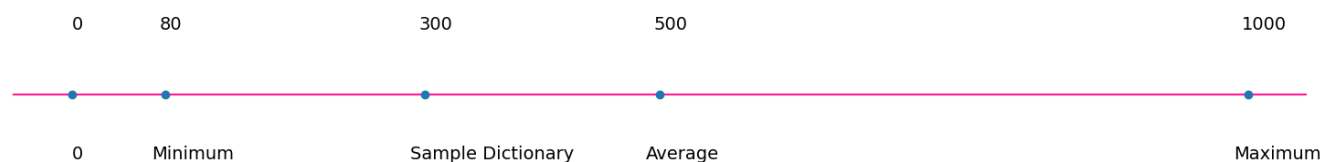

Fig. 4: Visualisation output for number of entries within the aggregated comparison (Yim et al., 2022)

4 Conclusion

In this deliverable we have described EDiE, an evaluation tool for dictionaries, that allows users to assess different aspects of dictionaries based on their metadata and entries. Furthermore, it provides aggregated metrics over dictionaries of interests/contexts within the ELEXIS Infrastructure and enables users to compare different dictionaries for their specific use cases.

5 Repository

The code can be found on github here: https://github.com/elexis-eu/edie

6 Publications

The following publications resulted from this task:

Bajčetić, Lenka, Yim, Seung-Bin, & Declerck, Thierry. (2022, June 29). Towards the Profiling of Linked Lexicographic Resources. In Proceedings of the Globalex Workshop on Linked Lexicography @LREC2022, https://doi.org/10.5281/zenodo.6777880

Seung-Bin Yim, Lenka Bajčetić, Thierry Declerck and John P. McCrae. EDIE - Elexis DIctionary Evaluation Tool. In Workshop on PROfiling LINGuistic KNOWledgE gRaphs (ProLingKNOWER), 23 May 2022. https://doi.org/10.5281/zenodo.6900760