

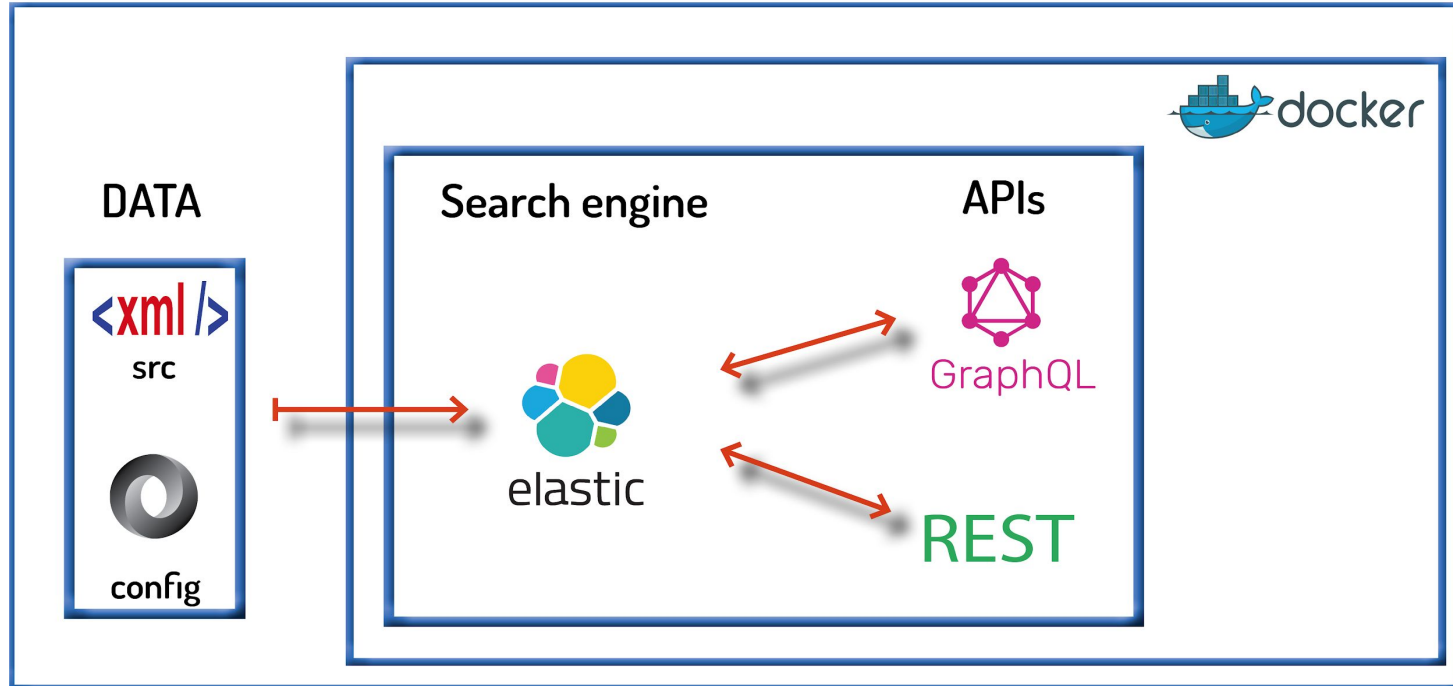


Introducing an Open, Dynamic and Efficient Access for TEI-encoded Dictionaries on the Internet

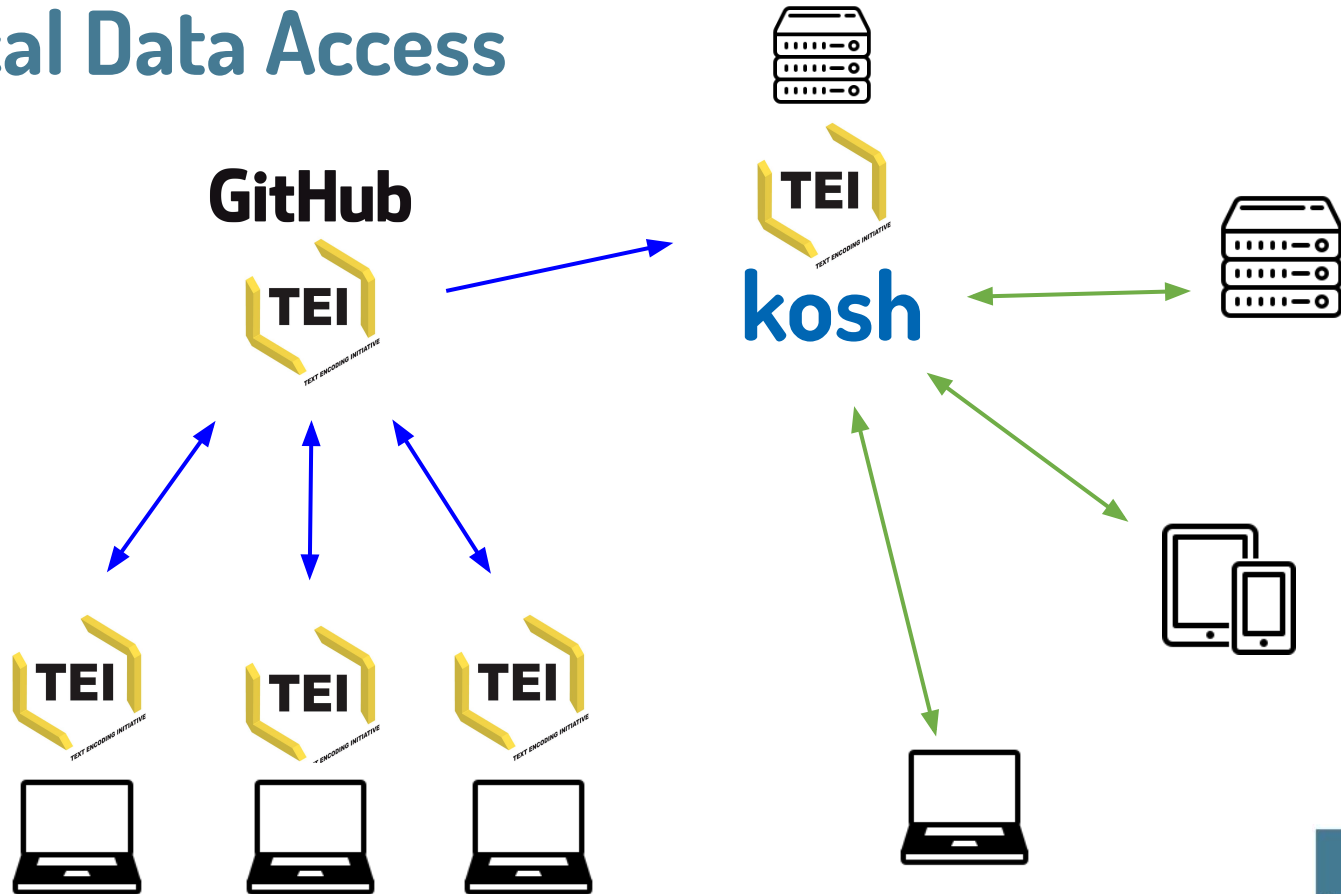
Accessing TEI-encoded lexical data

- Problem:
Multiple dictionaries scattered over the Internet, not **directly** accessible for **computers**
- One of the (best) solutions:
Application Programming Interfaces (APIs)
- How?
With a framework that as input takes **any XML**-encoded lexical dataset and as output provides **APIs** for its access

Kosh - Architecture



Lexical Data Access



Kosh Data – De Alcedo’s Dictionary

- *Diccionario Geografico-Histórico de las Indias Occidentales ó América* (1787-1789), 5 vol., covering geographical descriptions and [Latin] American word-usage (descriptive)
- **XML**-encoded for **Gazetteer** projects (HGIS de las Indias, Pelagios Commons)

Our work:

- **TEI-P5** encoding
- **API** access with **Kosh**

Kosh - Input

```
<entry xml:id="alcedo02424">
  <form>
    <orth type="lemma">CAEUA</orth>
  </form>
  <sense>Rio pequeño de la Provincia y Gobierno de la Guaiana ó Nueva Andalucía Nace en la
    Serranía de Parime corre del S al N y entra en el de Aredato</sense>
</entry>
<entry xml:id="alcedo02425">
  <form>
    <orth type="lemma">CAFAIATE</orth>
  </form>
  <sense>Pueblo de la Provincia y Gobierno del Tucuman en la jurisdicción de Salta anexo del
    Curato de Chiquiana</sense>
</entry>
```

```
{
  "mappings": {
    "_meta": {
      "_xpath": {
        "id": "../@xml:id",
        "root": "//tei:entry",
        "fields": {
          "lemma": "../tei:form/tei:orth",
          "sense": "../tei:sense"
        }
      }
    },
    "properties": {
      "lemma": {
        "type": "keyword"
      },
      "sense": {
        "type": "text"
      }
    }
  }
}
```

Kosh Data – GitHub

alcedo00481 ws removed from lemma

[Browse files](#)

 master



vocabulista committed 16 minutes ago

1 parent [b00e0cb](#)

commit [cfca473334742a20070522a740ea156d84e1cdb5](#)

Showing 1 changed file with 1 addition and 1 deletion.

Unified

Split

2  de_alcedo/alcedo-1.tei 

...

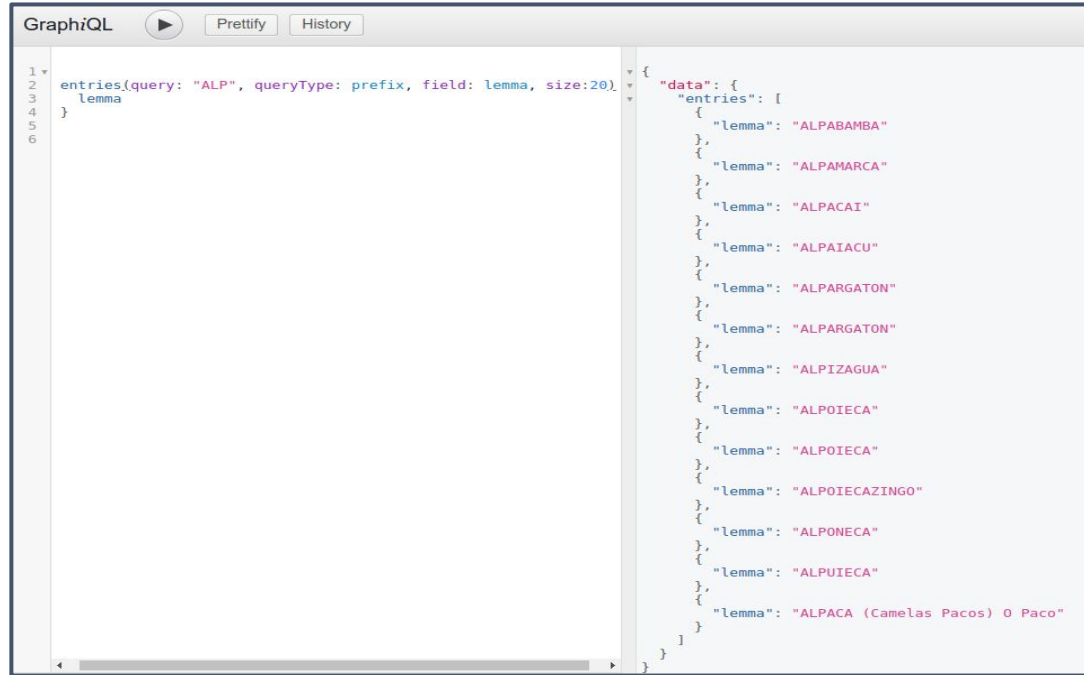
ΣΣ

@@ -4364,7 +4364,7 @@

4364	4364		</entry>
4365	4365		<entry xml:id="alcedo00481">
4366	4366		<form>
4367	-		<orth type="lemma">ALPUIEC A</orth>
4367	+		<orth type="lemma">ALPUIECA</orth>
4368	4368		</form>
4369	4369		<sense>Santa Maria de) Pueblo de la Cabecera de Partido de Huitepec y Alcaldia mayor de
4370	4370		Cuernavaca en Nueva España ; tiene 77 familias de Indios</sense>

ΣΣ

Kosh - GraphQL API



The screenshot shows the GraphQL Playground interface. On the left, a query is entered: `entries(query: "ALP", queryType: prefix, field: lemma, size:20)`. On the right, the JSON response is displayed, showing a list of 20 lemmas starting with 'ALP'.

```
1  entries(query: "ALP", queryType: prefix, field: lemma, size:20)
2  lemma
3  }
4
5
6
```

```
{
  "data": {
    "entries": [
      {
        "lemma": "ALPABAMBA"
      },
      {
        "lemma": "ALPAMARCA"
      },
      {
        "lemma": "ALPACAI"
      },
      {
        "lemma": "ALPAIACU"
      },
      {
        "lemma": "ALPARGATON"
      },
      {
        "lemma": "ALPARGATON"
      },
      {
        "lemma": "ALPIZAGUA"
      },
      {
        "lemma": "ALPOIECA"
      },
      {
        "lemma": "ALPOIECA"
      },
      {
        "lemma": "ALPOIECAZINGO"
      },
      {
        "lemma": "ALPONECA"
      },
      {
        "lemma": "ALPUIECA"
      },
      {
        "lemma": "ALPACA (Camelas Pacos) O Paco"
      }
    ]
  }
}
```


Thank You!

SLIDES

<https://www.doi.org/10.5281/zenodo.3451535>

Kosh Project Website:

kosh.uni-koeln.de

Kosh Data:

cceh.github.io/kosh_data

Cologne Center for eHumanities:

cceh.uni-koeln.de

Data Center for the Humanities:

dch.uni-koeln.de

Cologne South Asian Languages and Texts:

c-salt.uni-koeln.de