C-SALT APIs for Sanskrit Dictionaries: a novel approach for accessing digital lexical resources online

Introduction

This poster presents the C-SALT (Cologne South Asian Languages and Texts)[1] APIs for Sanskrit Dictionaries (https://api.c-salt.uni-koeln.de), whose aim is to grant direct access to the most relevant Sanskrit dictionaries hosted at the University of Cologne (http://www.sanskrit-lexicon.uni-koeln.de). The motivation for developing these APIs was, first, to allow our own applications, such as VedaWeb[2] (Fig. 2), access some of these Sanskrit dictionaries, and second, to create a workflow (Fig. 1) with open-access, reliable and well-documented technologies that allows third parties to access them.

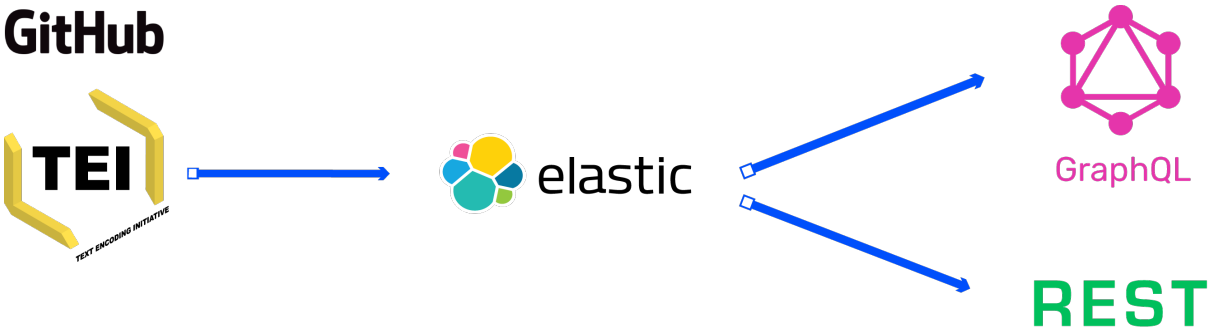

Fig 1. Simplified Workflow of the C-SALT APIs for Sanskrit Dictionaries

Workflow

The base of our workflow are TEI-encoded (Text Encoding Initiative) dictionaries that have been structured with a scheme (https://github.com/cceh/c-salt_dicts_schema) developed for transforming the most complex Sanskrit dictionaries that we host in C-SALT ([1][2][5]). In a second stage, the entries of the TEI file are indexed into elasticsearch[3]. Here, we index an entry's headword in Devanagari[4] and in three different Devanagari transliterations (ISO 15919[5], SLP1[6], and plain ASCII[7]

[1] http://c-salt.uni-koeln.de/
[2] http://vedaweb.uni-koeln.de/
[3] https://www.elastic.co/products/elasticsearch
[4] https://en.wikipedia.org/wiki/Devanagari
[5] https://en.wikipedia.org/wiki/ISO_15919
[6] https://en.wikipedia.org/wiki/SLP1
[7] https://en.wikipedia.org/wiki/ASCII

from ISO 15919). We index also the complete entry in ISO 15919 (TEI tags are not analyzed), which allows us to access TEI data in its original form.

For offering clients access to the elasticsearch server, we employ two APIs: a REST (Representational State Transfer) API ([3]) and a GraphQL API (https://graphql.org/). Both are implemented in Python with Flask, a web framework[8]. The reason for offering two APIs was to leave the clients decide which technology they want to employ and also to test how GraphQL fits our demands. The dictionaries in TEI format can be accessed by anyone in Github (https://github.com/cceh/c-salt_sanskrit_data), where changes can be proposed via pull requests that must be approved by a core team of scholars before taking effect. After a pull request has been accepted in Github, the modification is communicated automatically to the Git repository in our server through a Webhook[9], which triggers an update of the dictionary's respective index. This way, the APIs offer their clients the current status of the lexical data hosted in Github.

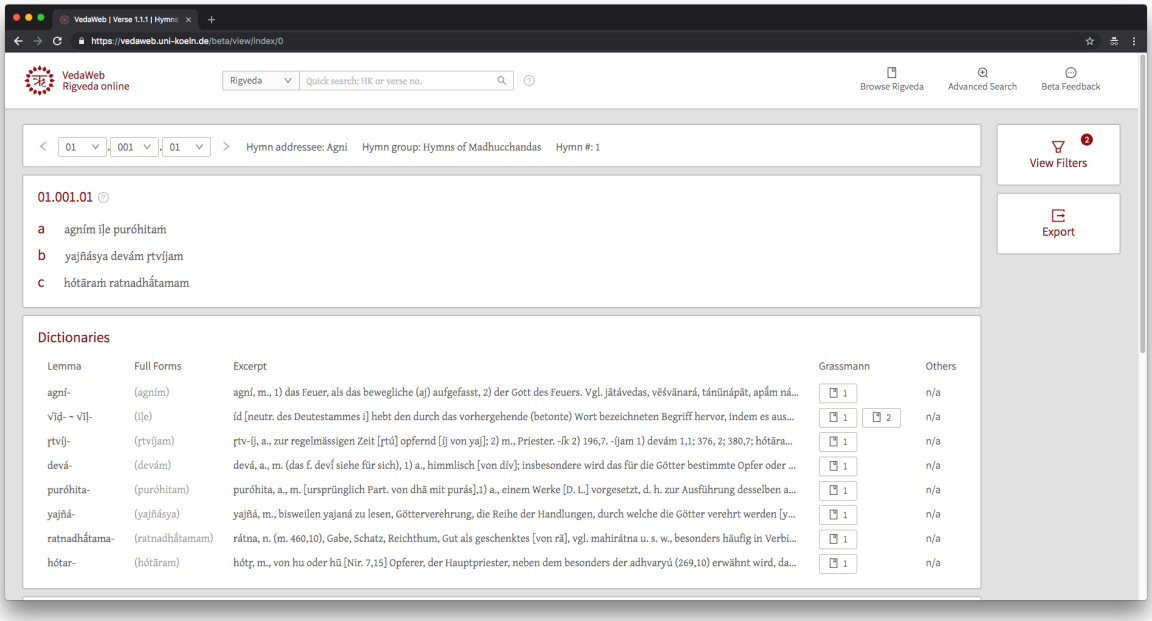

Fig 2. VedaWeb - Main layer (Rigveda) and Dictionaries layer, containing lexical data ([4]) fetched from the C-SALT GraphQL API for Sanskrit Dictionaries (https://api.c-salt.uni-koeln.de/dicts/sa/graphql)

Current status

As shown in Fig 2. in VedaWeb, a web application soon to be publicly available, we fetch data from Grassmann's dictionary for the Rigveda ([4]), but plan to add other dictionaries to this layer. This is an example of how our APIs can be integrated into a digital edition, but its usage is not reduced to this

8 http://flask.pocoo.org/

9 https://developer.github.com/webhooks/

purpose. Any application capable of establishing HTTP connections can access them. We hope that the different projects that in the last years have hosted their own instances of some of the C-SALT Sanskrit dictionaries will use our APIs and thus contribute to centralize efforts in order to provide high quality Sanskrit lexical data.

Literature

[1] APTE, Vama Shivaran (1920)3: *The student's English-Sanskrit dictionary*. Pune.

[2] BÖHTLINGK, Otto von, and Rudolf Roth (1855-1875). *Sanskrit-Wörterbuch*. St. Petersburg: Kaiserliche Akademie der Wissenschaften.

[3] FIELDING, Roy Thomas (2009): "Architectural Styles and the Design of Network-based Software Architectures". Dissertation. Irvine: University of California. URL: https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf

[4] GRASSMANN, Hermann (1873): *Wörterbuch zum Rig-veda*. Wiesbaden, O. Harrassowitz.

[5] MONIER-WILLIAMS, Monier (1899). *A Sanskrit-English dictionary, new edition*. Oxford: Clarendon.