Creating a collaborative research platform for Vedic Sanskrit texts

Fischer, Anna

annafischer@uni-wuppertal.de
University of Wuppertal, Germany

Kiss, Börge

b.kiss@uni-koeln.de
University of Cologne, Germany

Casaretto, Antje

antje.casaretto@uni-koeln.de
University of Freiburg, Germany

Kölligan, Daniel

daniel.koelligan@uni-wuerzburg.de
University of Würzburg, Germany

Korobzow, Natalie

natalie.korobzow@uni-wuerzburg.de
University of Würzburg, Germany

Neuefeind, Claes

c.neuefeind@uni-koeln.de
University of Cologne, Germany

Reinöhl, Uta

uta.reinoehl@uni-koeln.de
University of Freiburg, Germany

Sahle, Patrick

sahle@uni-wuppertal.de
University of Wuppertal, Germany

Introduction

Interdisciplinary collaboration is not something that just happens in Digital Humanities projects and deserves more attention (Griffin / Hayler 2018, Kiss et al. 2019). *VedaWeb*[1] is an established research platform for the ancient Vedic text Rigveda and is now evolving into an open collaborative tool for researching and sharing Vedic Sanskrit language data – an initiative carried by historical, general and computational linguists as well as DH specialists. How could this platform stimulate collaborative processes among researchers of Vedic Sanskrit and beyond?

Integrating diverse related resources

Grown out of an oral tradition, the Rigveda is the oldest religious text written in Vedic Sanskrit dated to ca. 1300/1000 BCE. The first DFG-funded "VedaWeb" project [2] (2017-2020) was dedicated to gathering, enriching and preprocessing various digital linguistic resources related to the Rigveda, transforming them into structured formats such as a comprehensive corpus in TEI-XML, and making them available on a web-based platform. Using this platform, they can now be searched, compared and browsed.

In the follow-up project [3] ("VedaWeb 2.0", 2022-2025), we are developing *Tekst (Text Exploration and Knowledge Structuring Tool)[4]*, a software that can be used by corpus or edition projects for creating open collaborative text research platforms. It will be the technical basis for the new *VedaWeb* platform, featuring more Vedic texts, some of which are fundamentally different in structure from the Rigveda. Text versions, translations, linguistic annotations, visual and audio material or links to dictionary entries and secondary literature are aligned with the structure of the respective reference text and can be viewed together and searched in unison (Neuefeind et al. 2022).

Blind collaboration

Regular feedback shows that *VedaWeb* is in active use by researchers from different disciplines, such as General and Historical Linguistics, Indo-European Studies, Indology and History. The platform has improved the scholarly access to the Rigveda: It offers one central place to access and explore numerous related resources.

We frequently receive corrections from users on semi-automatically created annotations and incorporate them into our data. While these users are not anonymous to us, they are unaware of each other's contributions to the same cause. This *blind collaboration* of platform users is far from ideal. At this point, the motivation of the research community to participate outruns our project's infrastructural capabilities: Suggested corrections currently need to be evaluated and executed manually by the linguistic experts in our team.

A perspective for actual collaboration

VedaWeb is meant to become a basis for the collective and participatory creation of an accessible scholarly research environment. As Robinson (2009: 9) puts it: "Through well-constructed scholarly networks over the web, scholars and readers may not only look at materials: they may make them, annotate them, correct them, draw conclusions from them and then contribute to others their conclusions."

For all this, the right conditions must be established: How does the platform make different resources accessible? Own materials can be added either as completely new data layers or based on existing ones. This way, *VedaWeb* allows for different practices to coexist, e.g. orthographic conventions or strategies of syntactic annotation (Biagetti et al. 2021). In addition to on-board functionality for editing, import and export is made possible for externally created data using certain exchange formats. How do corrections find their way into existing data? They can be shared,

commented on and thus negotiated and may exist as stand-alone, alternative variants or they can be merged into the original data.

Our experience as an interdisciplinary project team has shown us that producing (machine-)readable, comparable and searchable data is a collaborative process that keeps changing each team member's perspective on the resources and strategies used, opening up new possibilities for research (Kiss et al. 2019). The new collaborative platform gives rise to new questions, e.g. how it is going to change existing standards and workflows for other researchers of Vedic Sanskrit and beyond. Will it become a representation of diverging methodologies for linguistic analysis or lead to more converging practices and similar resources?

What is evident from the existing *VedaWeb* platform is that there seems to be a need for low-threshold, efficient access to text-related research data for Vedic Sanskrit as an "under-resourced language" (Hellwig et al. 2023: 1). How might the added possibility of participation further change this picture? Can (and should?) collaborative features really turn research platforms into 'one-stop shops' for research data?

Notes

1. VedaWeb Project Website: https://vedaweb.uni-koeln.de
2. https://gepris.dfg.de/gepris/projekt/329358806?language=en
3. https://gepris.dfg.de/gepris/projekt/470926253?language=en
4. *Tekst* documentation and source code: https://github.com/VedaWebProject/Tekst

Bibliography

Biagetti, Erica / **Hellwig, Oliver** / **Scarlata, Salvatore** / **Ackermann, Elia** / **Widmer, Paul** (2021): "Evaluating Syntactic Annotation of Ancient Languages", in: Old World: Journal of Ancient Africa and Eurasia 1, 1: 1-32. DOI: 10.1163/26670755-01010003.

Griffin, Gabriele / **Hayler, Matt S.** (2018): "Collaboration in Digital Humanities Research – Persisting Silences", in: Digital Humanities Quarterly 12, 1 http://digitalhumanities.org/dhq/vol/12/1/000351/000351.html [30.04.23].

Hellwig, Oliver / **Nehrdich, Sebastian** / **Sellmer, Sven** (2023): "Data-driven dependency parsing of Vedic Sanskrit", in: Lang Resources & Evaluation. DOI: 10.1007/s10579-023-09636-5.

Kiss, Börge / **Kölligan, Daniel** / **Mondaca, Francisco** / **Neuefeind, Claes** / **Reinöhl, Uta** / **Sahle, Patrick** (2019): "It Takes a Village: Co-developing VedaWeb, a Digital Research Platform for Old Indo-Aryan Texts", in: Proceedings of the Twin Talks Workshop at DHN 2019 – Understanding Collaboration in Digital Humanities, Copenhagen, Denmark, March 5, 2019: 35-44 http://ceur-ws.org/Vol-2365/ [30.04.23]. URN: urn:nbn:de:0074-2365-1.

Neuefeind, Claes / **Kölligan, Daniel** / **Reinöhl, Uta** / **Sahle, Patrick** (2022): "VedaWeb 2.0: Towards a Collaborative Workspace for Indo-Aryan Texts", in: Book of Abstracts of the Digital Humanities Conference 2022 (DH2022), Tokyo, Japan, July 25-29, 2022: 530-531 https://dh2022.dhii.asia/dh2022bookofabsts.pdf [30.04.23].

Robinson, Peter (2009): "Towards a Scholarly Editing System for the Next Decades", in: Huet, Gérard / Kulkarni, Amba / Scharf, Peter (eds.): Sanskrit Computational Linguistics (ISCLS 2007) (ISCLS 2008): 346-357, Lecture Notes in Computer Science 5402. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-642-00155-0_18.