

VedaWeb – eine webbasierte Plattform für die Erforschung altindischer Texte

Reinöhl, Uta

uta.reinoehl@uni-koeln.de
Universität zu Köln

Kölligan, Daniel

d.koelligan@uni-koeln.de
Universität zu Köln

Kiss, Borge

kiss@mailbox.org
Universität zu Köln

Mondaca, Francisco

f.mondaca@uni-koeln.de
Universität zu Köln

Neuefeind, Claes

c.neuefeind@uni-koeln.de
Universität zu Köln

Sahle, Patrick

sahle@uni-koeln.de
Universität zu Köln

Einleitung

Das hier beschriebene Poster thematisiert den technischen Entwurf und den funktionalen Aufbau von *VedaWeb*, einer webbasierten Plattform für die sprachwissenschaftliche Erforschung altindischer Texte (siehe <http://vedaweb.uni-koeln.de>). Das 2017 begonnene Vorhaben wird als Kooperationsprojekt an der Universität zu Köln durchgeführt, in enger Zusammenarbeit zwischen Fachwissenschaftlern des Instituts für Linguistik (Allgemeine Sprachwissenschaft, Historisch-Vergleichende Sprachwissenschaft und Sprachliche Informationsverarbeitung) sowie des Cologne Center for eHumanities (CCeH). Das Projekt wird von der Deutschen Forschungsgemeinschaft (DFG) in der LIS-Förderlinie „eResearch-Technologien“ gefördert.

Über die *VedaWeb*-Plattform werden altindische, in Sanskrit verfasste Texte morphologisch und metrisch annotiert, sowie nach lexikographischen und korpuslinguistischen Kriterien durchsuchbar zur

Verfügung gestellt. Als Pilottext dient zunächst der *Rigveda*, einer der ältesten und wichtigsten Texte der indogermanischen Sprachfamilie. Der *Rigveda* ist in der ältesten Sprachform des Altindischen, dem Vedischen, verfasst. Seine Entstehung kann auf das späte zweite Jahrtausend v. Chr. datiert werden. Mit einem Umfang größer als Homers *Ilias* und *Odyssee* zusammen stellt er eine überaus reiche Datengrundlage dar. Perspektivisch sollen auch weitere Texte wie etwa der *Atharvaveda*, *Yajurveda* und vedische Prosatexte in die *VedaWeb*-Plattform integriert werden. Das Projekt wird Forschungen in allen Bereichen des Vedischen erleichtern und voranbringen, beispielsweise in Bezug auf Fragestellungen der Syntax (siehe z.B. zu referentiellen Null-Objekten Keydana/Luraghi 2012, zu Nicht-Konfiguralität Reinöhl 2016), der Morphologie (siehe z.B. zum vedischen *vr#k#*-Typus Widmer 2007, zu *ya*-Präsentien Kulikov, 2012) oder der Wortbildung (siehe z.B. zu Komposita Scarlata/Widmer 2015). Es wird angestrebt, die *VedaWeb*-Plattform längerfristig zur zentralen Anlaufstelle für die internationale Fachgemeinschaft, die mit altindischen Primärtexten arbeitet, auszubauen, um den in Köln bestehenden Schwerpunkt auf südasiatische Sprachen weiter zu stärken.

Projektziele

Ausgangspunkt und Grundlage des Projekts ist eine vollständige morphologische (d.h. wortstrukturelle) Annotation des *Rigveda*, die im Vorfeld an der Universität Zürich durchgeführt und dem Projekt zur Verfügung gestellt wurde. Hinzu kommen metrische Informationen (Kevin Ryan, Harvard University, siehe <http://www.meluhha.com/rv/>) sowie perspektivisch auch syntaktische Informationen aus verschiedenen abgeschlossenen und andauernden Forschungsprojekten als weitere Annotationsebenen. Anhand dieser Annotationen werden im Projekt verschiedene Recherche- und Analysewerkzeuge entwickelt und sukzessive in die *VedaWeb*-Plattform integriert. Hierzu gehören eine kombinierte Suchfunktion nach linguistischen Parametern (u.a. Lemmata, Wortformen, morphologische und metrische Informationen), die Verknüpfung mit dem Standardwörterbuch zum *Rigveda* von Hermann Grassmann (1873), die Anzeige von Übersetzungen (u.a. Grassmann 1876, Geldner 2003, Griffith 1896) und von Kommentaren (Oldenberg 1909/1912), sowie die Möglichkeit des Exports von annotierten Textabschnitten nach vom Nutzer gewählten Kriterien.

Von zentraler Bedeutung ist die Verknüpfung des *Rigveda* mit der am CCeH angesiedelten Portalseite für Sanskritwörterbücher (*Cologne Digital Sanskrit Dictionaries*, siehe <http://www.sanskrit-lexicon.uni-koeln.de>), einer zentralen Anlaufstelle für die internationale Sanskritforschung. Auf Basis einer Modellierung der Daten in TEI werden die Wortformen über die jeweiligen Lemmata mit dem digital erfassten

Wörterbuch von Grassmann verknüpft, so dass sowohl vom Text auf das Wörterbuch verwiesen wird als auch umgekehrt vom Wörterbuch aus Textstellen aufgesucht werden können. Auf diese Weise kann über das Lemma gleichzeitig auch eine Verknüpfung zu den weiteren Sanskrit-Wörterbüchern hergestellt werden, etwa um vergleichende, wörterbuchübergreifende Recherchen zu ermöglichen. In der direkten Verknüpfung von Text und Wörterbuch und der damit verbundenen Erweiterung der Analysemöglichkeiten besteht das wesentliche Alleinstellungsmerkmal des Projekts gegenüber bestehenden Ressourcen des Altindischen wie z.B. dem *Thesaurus Indogermanischer Text- und Sprachmaterialien* (TITUS, siehe <http://titus.uni-frankfurt.de>).

Technische Aspekte

Der Schwerpunkt des Posters liegt auf der Präsentation des Systementwurfs der *VedaWeb*-Plattform sowie der dort eingesetzten Technologien. Dies umfasst zum einen eine Beschreibung der funktionalen Elemente der Nutzerschnittstelle, die dem Anwender als Forschungsumgebung dient, indem sie verschiedene Werkzeuge bereitstellt (z.B. Suche, Verlinkung, Export in TEI-Format). Zum anderen werden die Systemarchitektur und die für deren Umsetzung verwendeten Technologien thematisiert. Die *VedaWeb*-Plattform wird als Webanwendung auf Basis des *Spring-Frameworks* umgesetzt (siehe <https://spring.io>). Durch dessen weite Verbreitung und den großen Funktionsumfang sind sowohl die langfristige Wartbarkeit als auch zahlreiche Möglichkeiten zur späteren Erweiterung der Plattform gewährleistet. Für die Umsetzung der Suchlogik wurde zunächst der Einsatz von etablierten, auf *Lucene* (siehe <http://lucene.apache.org>) aufbauenden Such-Servern wie *Solr* oder *Elasticsearch* geprüft, jedoch wieder verworfen, da ihre Stärken v.a. im Durchsuchen sehr umfangreicher, dabei aber verhältnismäßig einfach strukturierter Textsammlungen liegen. Die *VedaWeb*-Suche hingegen soll mittels komplexer, kombinierbarer Suchkriterien über die reine Volltextsuche hinaus auch linguistisch motivierte Suchanfragen auf Grundlage der verschiedenen Annotationsebenen zulassen. Anstelle der genannten Such-Server wird deshalb eine eigene, auf den gegebenen Anwendungsfall zugeschnittene Suchlogik direkt mit *Lucene* implementiert. Das Frontend wird unter Verwendung einer Template Engine (*Thymeleaf*, siehe <http://www.thymeleaf.org>) und Client-seitigen JavaScript-Technologien umgesetzt, um eine möglichst flexible und effiziente Arbeitsweise zu ermöglichen. Mit der Fokussierung auf den Systementwurf möchten wir mit dem Poster vor allem einen kompakten Überblick über die Nutzungsmöglichkeiten sowie über die technische Funktionsweise der *VedaWeb*-Plattform geben.

Bibliographie

- Geldner, Karl Friedrich** (2003) [1951-57]: *Der Rig-Veda. Aus dem Sanskrit ins Deutsche übersetzt und mit einem laufenden Kommentar versehen von Karl Friedrich Geldner*. Cambridge (Mass.): Harvard University Press.
- Grassmann, Hermann** (1873): *Wörterbuch zum Rig-veda*. Wiesbaden, O. Harrassowitz.
- Grassmann, Hermann** (1876): *Rig-veda. Übersetzt und mit kritischen und erläuternden Anmerkungen versehen von Hermann Grassmann*. Leipzig: F.A. Brockhaus.
- Griffith, Ralph T. H.** (1896): *The Hymns of the Rigveda*. Benares: Lazarus.
- Keydana, Götz / Luraghi, Silvia** (2012): *Definite referential null objects in Vedic Sanskrit and Ancient Greek*. *Acta Linguistica Hafniensia* 44 (2):116–28. <https://doi.org/10.1080/03740463.2013.776245> .
- Kulikov, Leonid** (2012): *The Vedic -ya-Presents: Passives and Intransitivity in Old Indo-Aryan*. Amsterdam, Netherlands: Rodopi.
- Oldenberg, Hermann** (1909/1912): *R#gveda. Textkritische und exegetische Noten*. Berlin: Weidmann.
- Reinöhl, Uta** (2016): *Grammaticalization and the Rise of Configurationality in Indo-Aryan*. Oxford: Oxford University Press
- Scarlata, Salvatore / Widmer, Paul** (2015): *Vedische exozentrische Komposita mit drei Relationen*. *Indo-Iranian Journal*, 58(1):26-47.
- Widmer, Paul** (2007): *Der altindische vrk##-Typus und hethitisch nakki-: Der indogermanische Instrumental zwischen Syntax und Morphologie*. *Zeitschrift für Sprachwissenschaft*, (1-2):190-208.