

GRETIL – ein Überblick

Allgemeines

Das *Göttingen Register of Electronic Texts in Indian Languages* (GRETIL) ist eine fachkundig verwaltete Ressourcen-Plattform ...

- mit dem derzeit größten Angebot an einheitlich aufgearbeiteten maschinenlesbaren Texten, die mit dem Kulturraum Indien in Verbindung stehen,
- die philologisch arbeitenden Forschern weltweit schnellen Zugriff und selbstbestimmtes Arbeiten an diesen Texten erlaubt und
- die die Bemühungen vieler, Texte beisteuernder Individuen und Institutionen vergleichbar macht, bündelt und sichert.

Die Entwicklung

2001 Initiation im Rahmen des Fachreferats für Süd- und Südostasiatische Philologien; Listung von im World Wide Web zu findenden Texten und Konvertierung dieser in zwei damals übliche Formate:

CSX Classical Sanskrit Extended (auf CP437-Zeichensatz basierend)

REE Ronald E. Emmerick's Encoding (auf CP437-Zeichensatz basierend)

2003 Hinzufügung eines dritten Formats:

UTF-8 IAST-Transliteration in UTF-8 (heutiger Unicode-Standard)

2016 Vereinfachung und Aktualisierung der Prüf- und Konvertierungsvorgänge im Rahmen eines Technologietransfers

Derzeitiger Umfang

Insgesamt stehen auf GRETIL 1 054 (Stand November 2017) verschiedene Texte in diesen drei Formatierungen zur Verfügung. Auf die vertretenen Sprachen verteilen sich diese wie folgt:

Sprache/Sprachfamilie	Textanzahl
Sanskrit	771
Mittelindoarische Sprachen	71
Neuindoarische Sprachen	8
Dravidische Sprachen	182
Altjavanisch	6
Sekundärquellen	16

Prominent vertretene Kernthemen und der ungefähre Zeitraum, den die auf GRETIL vorhandenen Texte in diesen Themengebieten umreißen, sind dabei:

Themengebiet	Textanzahl
Vedische Religion (ca. 1500 v. Chr.–200 v. Chr.)	43
Einheimische Wissenschaften (ca. 400 v. Chr.–1600 n. Chr.)	99
Buddhismus (ca. 200 v. Chr.–900 n. Chr.)	386
Schöne Literatur (ca. 200–1900 n. Chr.)	266
Hinduismus (ca. 250–1600 n. Chr.)	190

Diese Texte wurden von über 260 Forschern und insgesamt 11 Institutionen beigesteuert, wobei der Pāli-Textbestand maßgeblich durch die Initiative der *Dhammakaya Foundation* und des *Sri Lanka Tripitaka Project* aufgebaut wurde. Derzeit wird das Angebot über Kooperationen mit *Colo-gne Digital Sanskrit Dictionaries* (CDSL) und dem *Thesaurus Indogermanischer Text- und Sprachmaterialien* (TITUS) ergänzt.

Weiteres Vorgehen

Kontinuität garantieren

Das bisherige, hohe Nutzeraufkommen – mit durchschnittlich 2 900 Zugriffen pro Monat (Stand Februar 2016) – und die Bereitschaft der wissenschaftlichen Gemeinde, eigene Texte beizutragen, liegt in erster Linie in der regelmäßigen Versorgung der Plattform mit neuen Texten begründet. Daher ist diese zu erhalten, indem die Texte in Qualität und Formatierung an das bisherige Corpus anschlussfähig bleiben. Folgende Aspekte von GRETEL bleiben daher unverändert:

- Konzentration auf ein möglichst breit gefächertes, umfangreiches Angebot an Texten in Plain Text,
- Erschließung und Sicherung weiterer, bereits vorhandener Texte durch Kooperation mit anderen Plattformen und Portalen,
- Ermutigung zur Beisteuerung individueller Forschungsdaten durch Aufrechterhaltung guter Kontakte in die wissenschaftliche Gemeinde.

Gezielt modernisieren

Ohne diese Kontinuität zu gefährden, wird der 2016 begonnene Technologietransfer nicht nur zur Modernisierung der Versorgung, sondern auch zu Erweiterung des philologischen Nutzens und zur Erhöhung der Nutzerfreundlichkeit fortgesetzt werden. Folgende Maßnahmen sind geplant:

- Überführung in ein dauerhaft verfügbares Repositorium,
- Erhöhung der Transparenz der Revisionsgeschichte einzelner Texte,
- Verbesserung der Möglichkeiten zur Versionskontrolle,
- Modernisierung und Dynamisierung des Frontends,
- Überarbeitung der Dokumentation und Bereitstellung zusätzlicher Anleitungen, um auch neue Forscher auf das Angebot aufmerksam zu machen.