JOST GIPPERT,
The TITUS Project. 25 years of corpus building in ancient languages,

in:
Perspektiven einer corpusbasierten historischen Linguistik und Philologie. Internationale Tagung des Akademienvorhabens „Altägyptisches Wörterbuch" an der Berlin-Brandenburgischen Akademie der Wissenschaften, 12. – 13. Dezember 2011, herausgegeben von Ingelore Hafemann, Berlin 2013, S. 169-192.

BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN

Thesaurus Linguae Aegyptiae 4

Perspektiven einer corpusbasierten historischen Linguistik und Philologie. Internationale Tagung des Akademienvorhabens „Altägyptisches Wörterbuch“ an der Berlin-Brandenburgischen Akademie der Wissenschaften, 12. – 13. Dezember 2011

herausgegeben von Ingelore Hafemann

BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN

Thesaurus Linguae Aegyptiae

4

BERLIN 2013

BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN

Perspektiven einer corpusbasierten historischen Linguistik und Philologie

Internationale Tagung des Akademienvorhabens „Altägyptisches Wörterbuch“ an der Berlin-Brandenburgischen Akademie der Wissenschaften, 12. – 13. Dezember 2011

herausgegeben von Ingelore Hafemann

BERLIN

2013

berlin-brandenburgische
AKADEMIE DER WISSENSCHAFTEN

Dieser Band wurde durch die gemeinsame Wissenschaftskonferenz im Akademienprogramm mit Mitteln des Bundes (Bundesministerium für Bildung und Forschung) und des Landes Berlin (Senatsverwaltung für Wirtschaft, Technologie und Forschung) gefördert

URN: urn:nbn:de:kobv:b4-opus-24310

INHALTSVERZEICHNIS

WORTGESCHICHTE - TEXTGESCHICHTE - SPRACHGESCHICHTE: TRADITION UND INNOVATION BEI DER TEXTPRODUKTION

VORWORT

Die internationale Tagung „Perspektiven einer corpusbasierten historischen Linguistik und Philologie“ vom 12. – 13. Dezember 2011 am Akademienvorhaben „Altägyptisches Wörterbuch“ der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) war dem Thema des Aufbaus und der Nutzungsperspektiven elektronischer Textcorpora und Wörterbücher in den historischen Sprachen gewidmet. Die Teilnehmer, Vertreter der Ägyptologie, der Hethitologie, Indogermanistik sowie Referenten aus der historischen Lexikographie des Mittel- und Frühneuhochdeutschen und des Altfranzösischen diskutierten vor allem über die Veränderungen, die mit dem Einsatz elektronischer Erfassungs- und Verarbeitungsprozeduren einhergehen. Vertreter der Computerlinguistik vom „Zentrum Sprache“ der BBAW wurden in die Diskussionen einbezogen. Dort beschäftigt man sich seit Jahren mit dem Aufbau großer elektronischer Textcorpora (DWDS), darunter auch solcher, die historische Texte (DTA) für die elektronische Nutzung ermöglichen.

Die größte Herausforderung dieser neuen elektronischen Corpora und Wörterbücher ist es, sowohl den Methoden und damit den wissenschaftlichen Ansprüchen der traditionellen Philologie und Lexikographie unbedingt verpflichtet zu bleiben als auch neue Gebiete wie die Corpus- und Computerlinguistik für die historischen Sprachen zu öffnen. Die Teilnehmer haben gemeinsam und disziplinenübergreifend die Möglichkeiten und Grenzen der Datenerfassung, ihrer Präsentation und den Nutzen neuer Auswertungsprozeduren diskutiert.

Unter dem ersten Thema „Historische Corpusprojekte – synchron und diachron“ wurden elektronische Corpora vorgestellt und ein intensiver Austausch darüber geführt, welche Datenstrukturen die linguistischen Inhalte in adäquater Weise abbilden. Wichtig war die Frage, auf welche Resonanz diese elektronischen Corpora bei den Nutzern gestoßen sind und welche Erwartungen und Anforderungen aus den verschiedenen Fachdisziplinen an die Projekte herangetragen werden. Der Austausch über Nutzungsperspektiven elektronischer Corpora schloss auch die Diskussion über die Erarbeitung projektübergreifend einsetzbarer Standards der Codierung und Strukturierung historischer Textdaten mit ein. Hinsichtlich einer mittel- und langfristigen Nutzbarkeit sowie einer langfristigen Datensicherheit stehen solche Fragen zunehmend im Focus und einige aktuelle Initiativen dazu wurden vorgestellt. Spezielle technische Aspekte elektronischer Datenerfassung und automatischer Analyse- und Speicherungsverfahren elektronischer Textdaten konnten am letzten Tag als ein Themenschwerpunkt mit den Programmierern diskutiert werden.

Ein zweiter Schwerpunkt waren konkrete Fragstellungen aus der historischen Lexikographie und diachronen Textanalyse. Für das Ägyptische ist der diachrone Ansatz auf Grund der über viertausendjährigen Textüberlieferung von großer Relevanz. Themen wie historischer und/oder textgattungsspezifischer Wortgebrauch, die Erarbeitung diachroner Wortlisten und Aspekte des kontaktindizierten Sprachwandels konnten disziplinübergreifend zwischen den Ägyptologen und den Kollegen der historischen Lexikographie des Mittel- und Frühneuhochdeutschen und des Altfranzösischen behandelt werden.

Mit dem Abendreferenten Gregory Crane, dem Begründer der „Perseus Digital Library", wurde ein breites Publikum angesprochen. In seinem Vortrag hat er noch einmal die hohe Relevanz und die neuen Möglichkeiten der Einbeziehung zahlreicher Wissenschaftler und einer interessierten Öffentlichkeit in die Projektarbeit demonstriert, die das Internet auf völlig neue Weise eröffnet hat. Die Herausgeberin ist sehr froh, seinen programmatischen Beitrag zu diesem Thema, dessen schriftliche Form er gemeinsam mit Alison Babeu erarbeitet hat, ebenfalls in diesem Band präsentieren zu können.

Wir danken der Berlin-Brandenburgischen Akademie der Wissenschaften für die umfassende Unterstützung unserer Projektarbeit und ganz speziell der Vorbereitung dieser Konferenz sowie der Möglichkeit, die Akten auf dem E-Doc-Server der Akademie veröffentlichen zu können.

Der Hermann und Elise geborene Heckmann Wentzel-Stiftung sei hiermit ausdrücklich für die unbürokratische und großzügige finanzielle Unterstützung dieser erfolgreichen Tagung gedankt.

Das Akademienvorhaben „Altägyptisches Wörterbuch" konnte sich als aktives Mitglied des Weiteren auf das „Zentrum Grundlagenforschung Alte Welt" stützen, dem alle altertumswissenschaftlichen Vorhaben der BBAW angehören. Dem Zentrum ist es zu danken, dass der Abendvortrag von Gregory Crane einem breiteren Publikum dargeboten werden konnte.

Allen Autoren dankt die Herausgeberin für ihre anregenden Diskussionen und die qualitätvollen Beiträge in diesem Band.

Auf eine Gesamtbibliographie wurde verzichtet und die Abkürzungen der in den ägyptologischen Beiträgen erwähnten Zeitschriften und Reihen folgen dem Lexikon der Ägyptologie, herausgegeben von Wolfgang Helck und Wolfhart Westendorf, Band VII: Nachträge, Korrekturen, Indices, Wiesbaden 1992, XIV-XIX.

Ganz besonders sei schließlich Frau Angela Böhme für die gewissenhafte redaktionelle Bearbeitung der Manuskripte gedankt sowie Dr. Simon Schweitzer für seine Hilfe beim Erstellen des Layouts.

Berlin, Mai 2013 Ingelore Hafemann

THE TITUS PROJECT
25 YEARS OF CORPUS BUILDING IN ANCIENT LANGUAGES

JOST GIPPERT

The article summarizes the contents and the structurtal premises of the "Thesaurus Indogermanischer Text- und Sprachmaterialien" (TITUS), focussing on search functions and facilities and questions of the encoding of ancient languages written in various scripts. Examples are taken from Tocharian, Greek, Vedic Sanskrit, and other ancient Indo-European languages covered by TITUS.

In September 1987, a group of Indo-Europeanists decided to join efforts in the digitization of primary sources that are essential for their research, by creating a common pool of the electronic texts to be prepared. Eversince,[1] the text pool has developed into a comprehensive retrieval system covering a large amount of relevant materials. The scope, the contents and the structural premises of the "Thesaurus Indogermanischer Text- und Sprachmaterialien" (TITUS) are summarized in the following pages.[2]

1. Since its foundation, the primary goal of the TITUS project consisted in the compilation of a comprehensive text database of ancient Indo-European languages that were not covered by concurrent projects such as the *Thesaurus Linguae Graecae*.[3] To reach this aim, a practical way of cooperation was decided upon: everybody who was able to contribute to the database was granted, as a member of the TITUS team, access to the complete database. In the 1980ies, this still presupposed data exchange via floppy or, later, compact disks, as internet facilities were not yet available in a sufficient way. Nevertheless, as early as 1988 the complete text of the Old Indic Rigveda Samhita, which had been electronically prepared as a text file of ca. 1.5 MB by H.S. Ananthanarayana under the supervision of W.P. Lehmann in Austin/Texas, was successfully transferred via a data line from the USA to the Berlin Free University, which hosted

[1] The project was announced under the title "Thesaurus altindogermanischer Textcorpora auf Datenträgern" in: *Die Sprache* 32/2, 1987 [1988], 151t.

[2] For previous accounts of the TITUS project cf. GIPPERT (1995a; 1995b; 1996; 2001; 2010).

[3] Project of the University of California at Irvine; cf. http://stephanus.tlg.uci.edu/.

the data pool then. By 1994, when the facilities of the internet emerged, the exchange of data was put on an online basis by establishing an FTP server at the University of Frankfurt, and soon after, the first web pages of the project were launched under the new name of "TITUS" which had meanwhile been agreed upon by the members.[4] Since 1996, the TITUS project has been promoting the use of Unicode to ensure a reliable encoding of its data, and the independent web server of the project established then[5] was one of the first sites world-wide to make a considerable amount of textual data available in this way of encoding. Thanks to a generous grant of the WordCruncher Company, the project was able in 1997 to install, along with its web site, a special "WordCruncher" server for the search and retrieval of data from the database. This service has been maintained until recently but has now been given up as most of the facilities it provides have meanwhile been implemented in an SQL-based online retrieval engine that has been publicly accessible since 2000.[6] Today, the TITUS database comprises not only corpora of ancient Indo-European languages such as Avestan, Vedic Sanskrit, Phrygian, or Umbrian, many of them covering the complete textual heritage of the languages involved, but also materials in more recent Indo-European as well as neighbouring languages, among them the largest corpus of Old and Middle Georgian available world-wide.[7] Many of the TITUS corpora have been the basis for more specialized corpus projects such as, e.g., the Referenzkorpus Altdeutsch project,[8] which aims at a full annotation of all textual materials in Old High German and Old Saxon; the Sanskrit Library project at Brown University, which aims at providing grammatical and other information pertinent to Sanskrit texts;[9] or the National Corpus of the Georgian Language, an international project aiming to cover the complete

[4] The clumsy URL was http://www.rz.uni-frankfurt.de/home/ftp/pub/titus/public_html/.

[5] URL: http://titus.uni-frankfurt.de/.

[6] URL: http://titus.fkidg1.uni-frankfurt.de/search/query.htm.

[7] Cf. http://titus.uni-frankfurt.de/texte/texte.htm for a full account of available corpora and texts.

[8] A common project of the universities of Berlin (Humboldt), Frankfurt and Jena, financed by the Deutsche Forschungsgemeinschaft since 2009 and part of the initiative "Deutsch-diachron-digital"; cf. http://www.deutschdiachrondigital.de.

[9] Cf. http://sanskritlibrary.org.

written history of Georgian from the 5th century A.D. up to the present day.[10]

2. With the establishment of the WordCruncher server in 1997, the TITUS project has moved far away from its original concept of being a mere exchange base of text files. Instead, the focus has shifted towards providing sophisticated search facilities within and across the text corpora, thus supporting online research into the languages and literatures in question. A few examples may suffice to illustrate the facilities that have been developed meanwhile.

2.1 One of the Indo-European languages for which TITUS may claim to cover the complete textual heritage in its corpus, is Tocharian, a language that was spoken in two different varieties in East Turkestan in the first millennium of our era. The textual remnants of the two Tocharian varieties (East- or A- and West- or B-Tocharian) are contained in ca. 5,000 manuscripts written in a "Northern" type of Brahmi script that were found in a region extending from Kucha to Turfan and Dunhuang along the Silk Road.[11] The largest part of these manuscripts is preserved in the Turfan Collection of the Berlin-Brandenburg Academy of Sciences (BBAW) today (ca. 4,000 manuscripts);[12] other major collections are hosted in London, Paris, and St. Petersburg. Within the TITUS project, work on the Tocharian manuscripts started in 1996 with the digitization of the printed editions of A- and B-Tocharian texts of the Berlin collection (in Romanized transcription), which formed the foundation of the emerging corpus. In the same year still, TITUS and the BBAW agreed upon preparing a complete set of digital images of the Tocharian manuscripts from Berlin to provide them online along the transcribed texts;[13] this endeavour, which was kindly supported by T. Tamai, resulted, in 2000, in one of the first frame-based online editions providing images, transcribed texts, and metadata as to each manuscript side by side. Today, this online-edition comprises the complete set of Berlin manuscripts including the ca. 3,000 hitherto unpublished

[10] Cf. http://georgiannationalcorpus.ac.ge.

[11] Cf. http://titus.uni-frankfurt.de/didact/karten/turkstan/turkst.htm for a map showing the locations.

[12] Cf. http://www.bbaw.de/en/research/turfanforschung as to the Turfan Studies project of the BBAW.

[13] Cf. GIPPERT (1997 and 1998) as to the technical foundations of the digitization project.

fragments, all manually transliterated by T. Tamai (cf. Fig. 1 showing a screen-shot of the site).[14]

2.2 In parallel to the online edition of the Berlin collection, which provides access to the corpus only via the catalogue number of a given manuscript,[15] the Tocharian data of all major collections have been prepared for a word-form based retrieval via the TITUS search engine.[16] This is built upon a more fine-grained referencing system where every single line of a manuscript (page) can be addressed directly (cf. Fig. 2 showing line 3 of the recto of the A-Tocharian fragment THT 634, which is part no. 1a in the edition by SIEG & SIEGLING 1921). To facilitate investigations into the paleography of the Brahmi script used for Tocharian, each line is further provided with an *akṣara*-based transliteration alongside the "normal" word-based transcription (as visible in Fig. 2). On the basis of a preindexation of the complete corpus, this allows for searching for both word-forms and individual *akṣaras*, either by clicking upon an item as displayed in the text or by using a query form. E.g., clicking upon the word-form *kumseñc* 'they come' in the given text line invokes (via a javascript underlying the word) a query for all (eight) occurrences of the same word-form throughout the A-Tocharian corpus,[17] which is output as a list of keyword-in-context entries with full referentiation of the text passages in question (cf. Figs. 3 and 4). Each entry is linked to the corresponding text passage so that the wider context can be accessed at will (cf. Fig. 5 showing the context of THT 935 = 302a, line 5). Note that in the transliteration, "Ä\" stands for the combination of the diaeresis-like vowel mark (which usually stands

[14] Cf. http://titus.fkidg1.uni-frankfurt.de/texte/tocharic/thtframe.htm; most elements of the edition were prepared in cooperation by K. Kupfer and T. Tamai.

[15] In the TITUS edition, the Berlin manuscripts are referenced according to their catalogue number in the Turfan Archive ("THT"). Of the 4074 manuscripts listed there, nos. 1–633 are B-Tocharian, and nos. 634–1099 are A-Tocharian (numbered 1–467 in the printed edition by SIEG & SIEGLING 1921). Several manuscripts have been missing since the Second World War; in some cases, digitized images could be provided from existing photographs.

[16] Cf. http://titus.uni-frankfurt.de/texte/texte2.htm#toch; for the time being, access to the B-Tocharian corpus, which is still under construction, is restricted to TITUS members and other registered users (cf. http://titus.uni-frankfurt.de/titusstd.htm for a form to apply for registration).

[17] The javascript causes an SQL-query to be sent to the following ASP script: http://titus.fkidg1.uni-frankfurt.de/database/titusinx/titusinx.asp?LXLANG=58285&LXWORD=kumseF100c&LCPL=0&TCPL=0&C=H&PF=26.

for the shewa vowel rendered as *ä* in transcriptions of Tocharian) with a *virāma* in word-final position (i.e., *kumseñc*$^{\ddot{a}}_{\backslash}$ in the transcription system applied in the editions by Sieg and Siegling); in the word-based search, it is ignored. In a similar way, the so-called “Fremdzeichen” are represented by capital letters, with “A” standing for their inherent vowel (i.e., KA etc. stands for *ḵa̱* etc. in Sieg/Siegling’s transcription); in the word-based search, *A* is treated as an equivalent of *ä* and the difference between “Fremdzeichen” and “Indian” akṣaras is ignored. The unsyllabic *u* vowel indicated by subscript *u* with a bent line above in the editions is represented by *ù* in the corpus; in the word-search, this is treated as being equivalent with plain *u*. In the *akṣara*-based transliteration, * stands for a (missing or unreadable) complete *akṣara* and +, for a (vocalic or consonantal) element of an *akṣara*; ^ stands for a word boundary within an *akṣara* (ignored in the search). All this is warranted by a specific structure of the underlying relational database, which contains “normalised” variants of the word-forms wherever applicable (cf. Fig. 6).[18]

2.3 A more flexible and powerful query method than the hyperlink-based retrieval is provided via special input forms. In general, the TITUS “Search Engine” comprises two different methods of input-based access to its data, one yielding a list of word-forms matching the query input, and one, the keyword-in-context output of occurrences as shown above (cf. Fig. 7).[19] In both input forms, the language of the search must be determined first, either specifically (e.g., “Tocharian A” as in Fig. 8) or more generally (e.g., “Tocharian” as in Fig. 9). The word-form to be searched can then be entered either in toto or partially, in exact Unicode encoding or in a substitutional plain-ASCII representation (or in a mixed representation), and with two types of “wild cards” replacing explicit characters: the question mark, “?”, stands for one single character, and the asterisk, “*”, for any sequence of characters (including zero). E.g., the word-form *kumseñc* can be entered as such or in the form *kumsen~c* (with the diacritic adscripted to match ASCII-based keyboards, cf. Fig. 10), and *kumseñc* will also be found if the query string is reduced to

[18] The database used at present is IBM DB-2 Express version 9.1, a powerful and yet free SQL-based system with full Unicode support.

[19] Both query types are accessible via http://titus.fkidg1.uni-frankfurt.de/search/query.htm. – A third query type, styled “unspecified”, consists of a mere link to a Google search over the TITUS site.

_k?ms*c_, i.e., with one character between _k_ and _m_ and any sequence of characters between _s_ and _c_ (cf. Fig. 11), or kums*, i.e., with any sequence of characters following _kums_, as the resulting word-list shows (cf. Fig. 12). Similarly, the word-form _NAmseñc_ ‘they revere’, which occurs in the same line of THT 634 as _kumseñc_, can be retrieved by entering _NAmseñc_, _nämseñc_, _n*ms??c_, _näm*_, etc. (cf. Fig. 13 showing the variant spellings in parentheses). In the word-list output, every word-form is provided with a hyperlink to the relevant context query; this means that by clicking upon _NAmseñc_ or _nämseñc_ in the list, all 3 occurrences of the word-form (the number of occurrences is indicated in square brackets for each list entry) will be listed in an extra window (cf. Fig. 14).[20] Of course, the same result can also be achieved with the “context output” form, entering, e.g., _n?mse?c_ (cf. Fig. 15). If the query string has more than one match as, e.g., in the case of _kums*_ (cf. above), the occurrences of the different matching forms will be output in alphabetic order (cf. Fig. 16 showing first an occurrence of _kumsanträ_, 3rd pers.pl.pres.ind.med., then one of _kumse_, 3rd pers.pl.pres.ind.act., shortened form). In addition, in the case of verb forms, the header of the output list indicates the underlying root (if determinable), again provided with a hyperlink (cf. Fig. 17); this leads to a special table which illustrates for every Tocharian verb which of its paradigm forms are attested in the two dialects (cf. Fig. 18).

2.4 As was stated above, the input of search strings in the query forms can be exactly “as is”, i.e., in Unicode encoding, or in a substitutional plain-ASCII format with adscript diacritics. This is true not only for the input of Latin-based scripts (or transcriptions) but also for other scripts. Thus, e.g., to search for the attestations of Greek ἄνδρα (acc.sg. of ἀνήρ ‘man’), both the Greek spelling and Latin _andra_ can be entered (cf. Figs. 19 to 21). Note that the entry of Greek diacritics is not necessary as unaccented variants are stored in the database for all word-forms; this means that all occurrences of ἄνδρα will also be found in a search for (less specified) ανδρα. This is even true for the same word spelt with an initial capital (Ἄνδρα) or with an acute on the word-final vowel (ἄνδρά, to be expected in the

[20] This is again invoked by a javascript which sends the SQL-query to the following URL: http://titus.fkidg1.uni-frankfurt.de/database/titusinx/titusinx.asp?lxlang=941&lxword=N4100mseF100c&LCPL=1&TCPL=1&C=H.

position preceding a clitic), which are matched by ανδρα (and *andra*) but not by ἄνδρα (cf. Figs. 21 and 22).

2.5 A special feature of the context-related search is the "combined search" function. Up to four query patterns (word-forms, stems, etc.) can be entered in parallel for a search of co-occurrences in a given context; cf., e.g., Figs. 24 and 25 showing a combined search for *thaz* 'the' and *uuort* 'word' in the Old High German corpus. The amount of context envisaged here can be determined by the user. If the "distance" is set to "0" (the default setting), the context in question is the lowest reference level of a given text (usually a sentence, a verse or a line); in the given example, this yields 111 co-occurrences of *thaz* and *uuort*, irrespective of the order of the two words (and including spelling variants such as *tház* and *uuórt*). Setting the distance to "1 - exact" (cf. Fig. 26) yields but 33 co-occurrences, with *uuort* immediately following *thaz*. (cf. Fig. 27).

2.6 A feature that has only been implemented for Old Indic (Sanskrit) and Avestan so far is the "thesaurus search" function. Different from the word-form based queries illustrated above, this function admits of searching for complete paradigms of words irrespective of a common ("matching") string structure of the individual word-forms. Starting, e.g., from *r̥tvíjo* as the genitive singular case form of the Vedic noun *r̥tvíj-* 'priest' (cf. Fig. 28), the output displays all occurrences of all case forms of this word as met with in the corpus[21] beginning with the nominative singular *r̥tvík*, provided with a grammatical analysis of each form[22] and a German translation of the respective lemmata[23] (cf. Fig. 29).

3. It is obvious that the latter type of retrieval presupposes a thorough modelling of the morphology of the language concerned. To implement similar facilities for all languages covered by TITUS

[21] Including spelling variants (caused by sandhi and accentuation), the list comprises the following forms: *r̥tvik*, *r̥tvík*, *r̥tvig*, *r̥tvíg*, *r̥tvijam*, *r̥tvíjam*, *r̥tvijaṃ*, *r̥tvíjaṃ*, *r̥tvijā*, *r̥tvíjā*, *r̥tvijaḥ*, *r̥tvíjaḥ*, *r̥tvijas*, *r̥tvíjas*, *r̥tvijaś*, *r̥tvíjaś*, *r̥tvijo*, *r̥tvíjo*, *r̥tvija*, *r̥tvíja*, (*r̥tvijaẖ*, *r̥tvijaḫ*), *r̥tvijām*, *r̥tvíjām*, *r̥tvijāṃ*, *r̥tvíjāṃ*.

[22] The analysis was worked out by R. Gehrke in the course of the AUREA project ("Avesta und Rigveda: Elektronische Analyse") financed by the Deutsche Forschungsgemeinschaft in 1997-1999; cf. http://titus.uni-frankfurt.de/curric/aurea/aurea.htm.

[23] Based upon the dictionary by K. MYLIUS (1992).

therefore means an immense task for the project that still has to be undertaken. Another task for the future that can be envisaged today concerns improvements in the rendering of non-Latin scripts as in the case of the cuneiform inscriptions of Old Persian for which a Unicode-based encoding in the original script has recently been provided by A. Sarhadi and M. Esnaashari (cf. Fig. 30).[24] In some cases, this is still hampered by the fact that the corresponding code-points of the Unicode standard are not yet available; e.g., it would be possible now to encode the Avestan texts in the original script[25] but for Middle Persian (Pahlavi) passages that are often met with in Avestan contexts, a Unicode rendering is not yet available. As in former cases, the members of the TITUS project are ready to support the standardisation process with their expertise.

[24] Cf. http://titus.uni-frankfurt.de/texte/etcs/iran/airan/apers/apers.htm.

[25] Cf. GIPPERT (forthcoming) as to details.

BIBLIOGRAPHY

GIPPERT, J., 1995a: TITUS. Das Projekt eines indogermanistischen Thesaurus, in: *LDV-Forum* 12/1, 35-47.

GIPPERT, J., 1995b: TITUS. Von der Keilschrifttafel zur Textdatenbank, in: *Forschung Frankfurt* 4/1995, 46-56.

GIPPERT, J., 1996: TITUS – Alte und neue Perspektiven eines indogermanistischen Thesaurus, in: *Studia Iranica, Mesopotamica et Anatolica* 2, 1996 [1997], 49-76.

GIPPERT, J., 1997: Digitization of Tocharian Manuscripts. Short notice about a new project, in: *Tocharian and Indo-European Studies* 7, 265-266.

GIPPERT, J., 1998: Digitization of Tocharian Manuscripts from the Berlin Turfan Collection, in: *Manuscripta Orientalia. International Journal for Oriental Manuscript Research* 4/1, 49-57.

GIPPERT, J., 2001: Der TITUS-Server: Grundlagen eines multilingualen Online-Retrieval-Systems, in: WILLÉE, G. *et al.* (eds.), *Computerlinguistik. Was geht, was kommt? / Computational Linguistics. Achievements and Perspectives. Festschrift für Wilhelm Lenders*, Bonn 2002, 81-85.

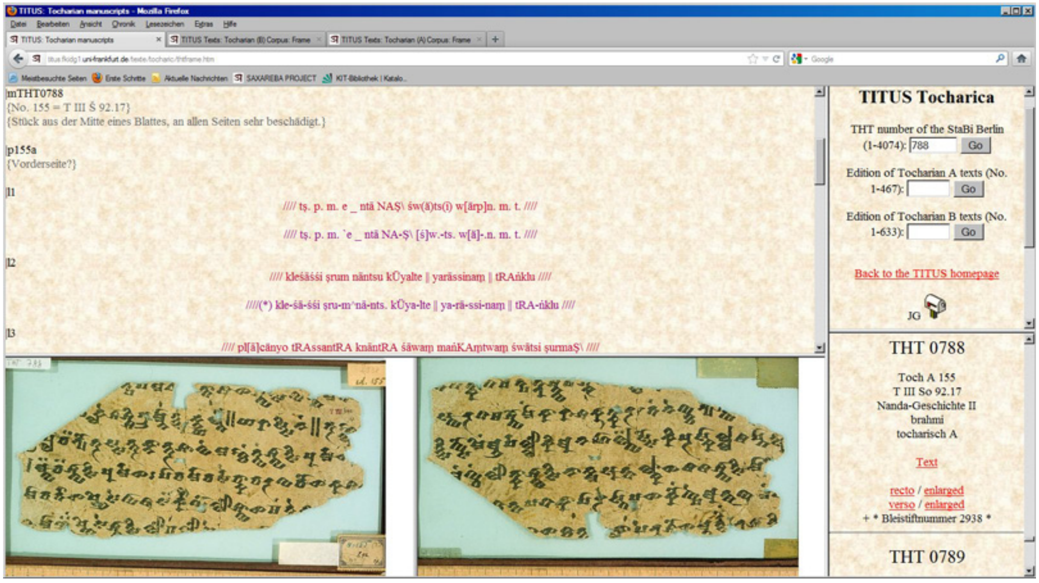
GIPPERT, J., 2010: Manuscript Related Data in the TITUS Project, in: *Comparative Oriental Manuscript Studies Newsletter* 1, 2011, 7-8.

GIPPERT, J., forthcoming: The Encoding of Avestan: Problems and Solutions, to appear in: *Journal for Language Technology and Computational Linguistics*, 2012.

MYLIUS, K., 1992: *Wörterbuch Sanskrit-Deutsch*, 4. Auflage, Leipzig [u.a.].

SIEG, E. & W. SIEGLING, 1921: *Tocharische Sprachreste*, Band I: *Die Texte*, A: *Transcription*, B: *Tafeln*, Berlin [u.a.].

FIGURES


TITUS: Tocharian manuscripts - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian manuscripts
TITUS Texts: Tocharian (B) Corpus: Frame
TITUS Texts: Tocharian (A) Corpus: Frame
titus.fkidg1.uni-frankfurt.de/texte/tocharic/thtframe.htm
Google
Meistbesuchte Seiten Erste Schritte Aktuelle Nachrichten SAXAREBA PROJECT KIT-Bibliothek | Katalo...

|mTHT0788
{No. 155 = T III Š 92.17}
{Stück aus der Mitte eines Blattes, an allen Seiten sehr beschädigt.}

|p155a
{Vorderseite?}

|l1

//// tṣ. p. m. e _ ntā NAṢ\ św(ā)ts(i) w[ārp]n. m. t. ////

//// tṣ. p. m. `e _ ntā NA-Ṣ\ [ś]w.-ts. w[ā]-.n. m. t. ////

|l2

//// kleśāśśi ṣrum nāntsu kÜyalte || yarāssinaṃ || tRAṅklu ////

////(*) kle-śā-śśi ṣru-m^nā-nts. kÜya-lte || ya-rā-ssi-naṃ || tRA-ṅklu ////

|l3

//// pl[ā]cānyo tRAssantRA knāntRA śāwaṃ maṅKAṃtwaṃ śwātsi ṣurmaṢ\ ////

TITUS Tocharica

THT number of the StaBi Berlin (1-4074): 788 Go

Edition of Tocharian A texts (No. 1-467): Go

Edition of Tocharian B texts (No. 1-633): Go

Back to the TITUS homepage

JG

THT 0788

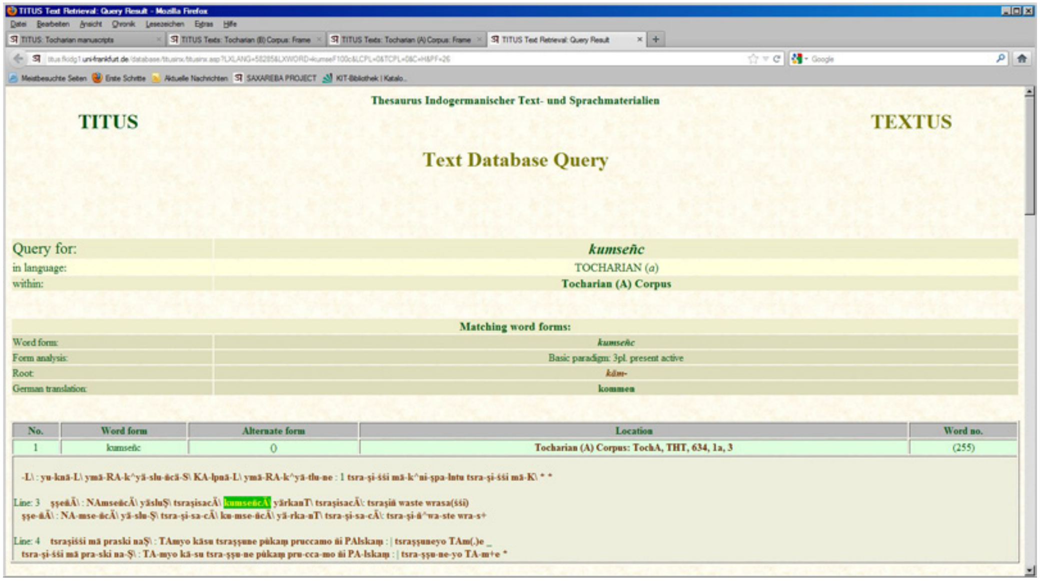
Toch A 155
T III So 92.17
Nanda-Geschichte II
brahmi
tocharisch A

Text

recto / enlarged
verso / enlarged
+ * Bleistiftnummer 2938 *

THT 0789

Figure 1


TITUS Text Retrieval: Query Result - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian manuscripts
TITUS Texts: Tocharian (B) Corpus: Frame
TITUS Texts: Tocharian (A) Corpus: Frame
TITUS Text Retrieval: Query Result
titus.fkidg1.uni-frankfurt.de/database/titusinx/titusinx.asp?LXLANG=582858LXWORD=kumseF100c&LCPL=0&TCPL=0&C=H&PF=26
Google
Meistbesuchte Seiten Erste Schritte Aktuelle Nachrichten SAXAREBA PROJECT KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

Text Database Query

Query for:	***kumseñc***
in language:	TOCHARIAN (*a*)
within:	**Tocharian (A) Corpus**

Matching word forms:	
Word form:	***kumseñc***
Form analysis:	Basic paradigm: 3pl. present active
Root:	***käm-***
German translation:	**kommen**

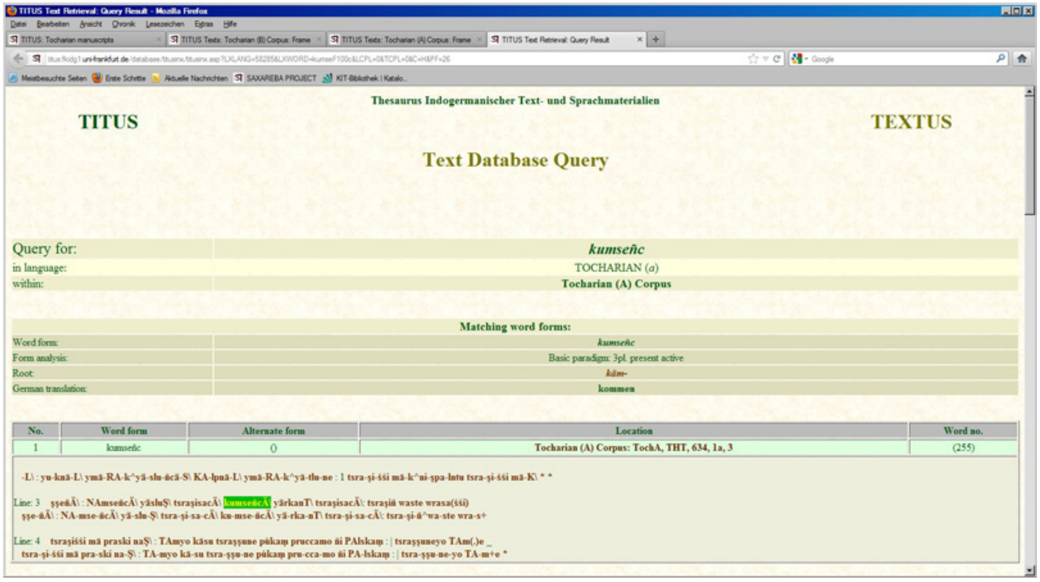
No.	Word form	Alternate form	Location	Word no.
1	kumseñc	()	**Tocharian (A) Corpus: TochA, THT, 634, 1a, 3**	(255)

-L\ : yu-knā-L\ ymā-RA-k^yā-slu-ñcā-S\ KA-lpnā-L\ ymā-RA-k^yā-tlu-ne : 1 tsra-ṣi-śśi mā-k^ni-ṣpa-lntu tsra-ṣi-śśi mā-K\ * *

Line: 3 ṣṣeñĀ\ : NAmseñcĀ\ yāsluṢ\ tsraṣisacĀ\ kumseñcĀ\ yārkanT\ tsraṣisacĀ\: tsraṣiñ waste wrasa(śśi)
ṣṣe-ñĀ\ : NA-mse-ñcĀ\ yā-slu-Ṣ\ tsra-ṣi-sa-cĀ\ ku-mse-ñcĀ\ yā-rka-nT\ tsra-ṣi-sa-cĀ\: tsra-ṣi-ñ^wa-ste wra-s+

Line: 4 tsraṣiśśi mā praski naṢ\ : TAmyo kāsu tsraṣṣune pùkaṃ pruccamo ñi PAlskaṃ : | tsraṣṣuneyo TAm(.)e _
tsra-ṣi-śśi mā pra-ski na-Ṣ\ : TA-myo kā-su tsra-ṣṣu-ne pùkaṃ pru-cca-mo ñi PA-lskaṃ : | tsra-ṣṣu-ne-yo TA-m+e *

Figure 2


TITUS Text Retrieval: Query Result - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian manuscripts
TITUS Texts: Tocharian (B) Corpus: Frame
TITUS Texts: Tocharian (A) Corpus: Frame
TITUS Text Retrieval: Query Result
titus.fkidg1.uni-frankfurt.de/database/titusinx/titusinx.asp?LXLANG=58285&LXWORD=kumseF100c&LCPL=0&TCPL=0&C=H&PF=26
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

Text Database Query

Query for:	*kumseñc*
in language:	TOCHARIAN (*a*)
within:	**Tocharian (A) Corpus**

Matching word forms:

Word form:	***kumseñc***
Form analysis:	Basic paradigm: 3pl. present active
Root:	***käm-***
German translation:	**kommen**

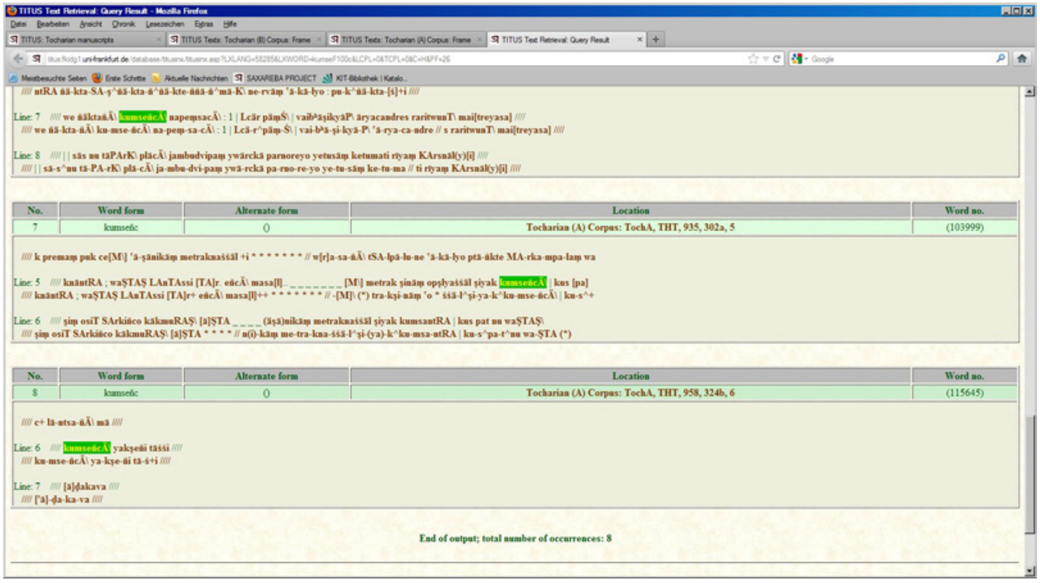
No.	Word form	Alternate form	Location	Word no.
1	kumseñc	()	Tocharian (A) Corpus: TochA, THT, 634, 1a, 3	(255)

-L\ : yu-knā-L\ ymā-RA-k^yā-slu-ñcā-S\ KA-lpnā-L\ ymā-RA-k^yā-tlu-ne : 1 tsra-ṣi-śśi mā-k^ni-ṣpa-lntu tsra-ṣi-śśi mā-K\ * *

Line: 3 ṣṣeñĀ\ : NAmseñcĀ\ yāsluṢ\ tsraṣisacĀ\ kumseñcĀ\ yārkanT\ tsraṣisacĀ\: tsraṣiñ waste wrasa(śśi)
ṣṣe-ñĀ\ : NA-mse-ñcĀ\ yā-slu-Ṣ\ tsra-ṣi-sa-cĀ\ ku-mse-ñcĀ\ yā-rka-nT\ tsra-ṣi-sa-cĀ\: tsra-ṣi-ñ^wa-ste wra-s+

Line: 4 tsraṣiśśi mā praski naṢ\ : TAmyo kāsu tsraṣṣune pùkaṃ pruccamo ñi PAlskaṃ : | tsraṣṣuneyo TAm(.)e _
tsra-ṣi-śśi mā pra-ski na-Ṣ\ : TA-myo kā-su tsra-ṣṣu-ne pùkaṃ pru-cca-mo ñi PA-lskaṃ : | tsra-ṣṣu-ne-yo TA-m+e *

Figure 3


TITUS Text Retrieval: Query Result - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian manuscripts
TITUS Texts: Tocharian (B) Corpus: Frame
TITUS Texts: Tocharian (A) Corpus: Frame
TITUS Text Retrieval: Query Result
titus.fkidg1.uni-frankfurt.de/database/titusinx/titusinx.asp?LXLANG=58285&LXWORD=kumseF100c&LCPL=0&TCPL=0&C=H&PF=26
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

//// ntRA ñā-kta-SA-ṣ^ñā-kta-ñ^ñā-kte-ññā-ñ^mā-K\ ne-rvāṃ 'ā-kā-lyo : pu-k^ñā-kta-[ś]+i ////

Line: 7 //// we ñāktañĀ\ kumseñcĀ\ napeṃsacĀ\ : 1 | Lcār pāṃŚ\ | vaibʰāṣikyāP\ āryacandres raritwunT\ mai[treyasa] ////
//// we ñā-kta-ñĀ\ ku-mse-ñcĀ\ na-peṃ-sa-cĀ\ : 1 | Lcā-r^pāṃ-Ś\ | vai-bʰā-ṣi-kyā-P\ 'ā-rya-ca-ndre // s raritwunT\ mai[treyasa] ////

Line: 8 //// | | sās nu tāPArK\ plācĀ\ jambudvipaṃ ywārckā parnoreyo yetusāṃ ketumati rīyaṃ KArsnāl(y)[i] ////
//// | | sā-s^nu tā-PA-rK\ plā-cĀ\ ja-mbu-dvi-paṃ ywā-rckā pa-rno-re-yo ye-tu-sāṃ ke-tu-ma // ti rīyaṃ KArsnāl(y)[i] ////

No.	Word form	Alternate form	Location	Word no.
7	kumseñc	()	Tocharian (A) Corpus: TochA, THT, 935, 302a, 5	(103999)

//// k premaṃ puk ce[M\] 'ā-ṣānikāṃ metraknaśśāl +i * * * * * * * * // w[r]a-sa-ñĀ\ tSA-lpā-lu-ne 'ā-kā-lyo ptā-ñkte MA-rka-mpa-laṃ wa

Line: 5 //// knāntRA ; waṢTAṢ LAnTAssi [TA]r. eñcĀ\ masa[l].. _ _ _ _ _ _ _ _ [M\] metrak ṣināṃ opṣlyaśśāl ṣiyak kumseñcĀ\ | kus [pa]
//// knāntRA ; waṢTAṢ LAnTAssi [TA]r+ eñcĀ\ masa[l]++ * * * * * * * * // -[M]\ (*) tra-kṣi-nāṃ 'o * śśā-l^ṣi-ya-k^ku-mse-ñcĀ\ | ku-s^+

Line: 6 //// ṣiṃ osiT SArkiñco kākmuRAṢ\ [ā]ṢTA _ _ _ _ (āṣā)nikāṃ metraknaśśāl ṣiyak kumsantRA | kus pat nu waṢTAṢ\
//// ṣiṃ osiT SArkiñco kākmuRAṢ\ [ā]ṢTA * * * * // n(i)-kāṃ me-tra-kna-śśā-l^ṣi-(ya)-k^ku-msa-ntRA | ku-s^pa-t^nu wa-ṢTA (*)

No.	Word form	Alternate form	Location	Word no.
8	kumseñc	()	Tocharian (A) Corpus: TochA, THT, 958, 324b, 6	(115645)

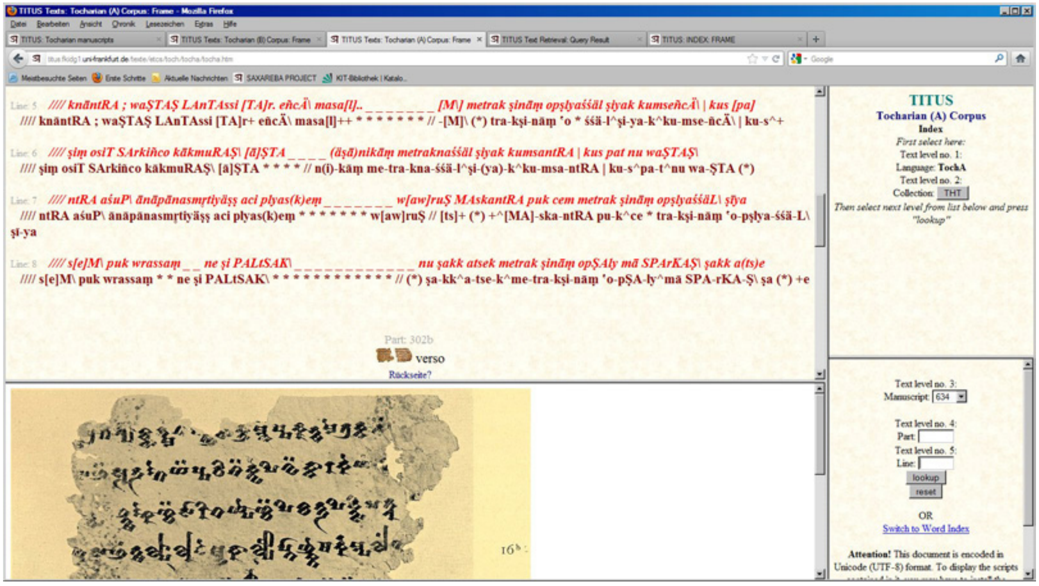
//// c+ lā-ntsa-ñĀ\ mā ////

Line: 6 //// kumseñcĀ\ yakṣeñi tāśśi ////
//// ku-mse-ñcĀ\ ya-kṣe-ñi tā-ś+i ////

Line: 7 //// [ā]ḍakava ////
//// ['ā]-ḍa-ka-va ////

End of output; total number of occurrences: 8

Figure 4


TITUS Texts: Tocharian (A) Corpus: Frame - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian manuscripts
TITUS Texts: Tocharian (B) Corpus: Frame
TITUS Texts: Tocharian (A) Corpus: Frame
TITUS Text Retrieval: Query Result
TITUS: INDEX: FRAME
titus.fkidg1.uni-frankfurt.de/texte/etcs/toch/tocha/tocha.htm
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...
16b

Line: 5 //// knāntRA ; waṢTAṢ LAnTAssi [TA]r. eñcĂ\ masa[l].. _ _ _ _ _ _ _ _ [M\] metrak şināṃ opşlyaśśāl şiyak kumseñcĂ\ | kus [pa]
//// knāntRA ; waṢTAṢ LAnTAssi [TA]r+ eñcĂ\ masa[l]++ * * * * * * * * // -[M]\ (*) tra-kşi-nāṃ 'o * śśā-l^şi-ya-k^ku-mse-ñcĂ\ | ku-s^+

Line: 6 //// şiṃ osiT SArkiñco kākmuRAṢ\ [ā]ṢTA _ _ _ _ (āşā)nikāṃ metraknaśśāl şiyak kumsantRA | kus pat nu waṢTAṢ\
//// şiṃ osiT SArkiñco kākmuRAṢ\ [ā]ṢTA * * * * // n(i)-kāṃ me-tra-kna-śśā-l^şi-(ya)-k^ku-msa-ntRA | ku-s^pa-t^nu wa-ṢTA (*)

Line: 7 //// ntRA aśuP\ ānāpānasmṛtiyāşş aci plyas(k)eṃ _ _ _ _ _ _ _ _ w[aw]ruṢ MAskantRA puk cem metrak şināṃ opşlyaśśāL\ şīya
//// ntRA aśuP\ ānāpānasmṛtiyāşş aci plyas(k)eṃ * * * * * * * w[aw]ruṢ // [ts]+ (*) +^[MA]-ska-ntRA pu-k^ce * tra-kşi-nāṃ 'o-pşlya-śśā-L\ şi-ya

Line: 8 //// s[e]M\ puk wrassaṃ _ _ ne şi PALtSAK\ _ _ _ _ _ _ _ _ _ _ _ _ _ nu şakk atsek metrak şināṃ opṢAly mā SPArKAṢ\ şakk a(ts)e
//// s[e]M\ puk wrassaṃ * * ne şi PALtSAK\ * * * * * * * * * * * * // (*) şa-kk^a-tse-k^me-tra-kşi-nāṃ 'o-pṢA-ly^mā SPA-rKA-Ṣ\ şa (*) +e

Part: 302b
verso
Rückseite?

TITUS
Tocharian (A) Corpus
Index
First select here:
Text level no. 1:
Language: **TochA**
Text level no. 2:
Collection: THT
Then select next level from list below and press "lookup"

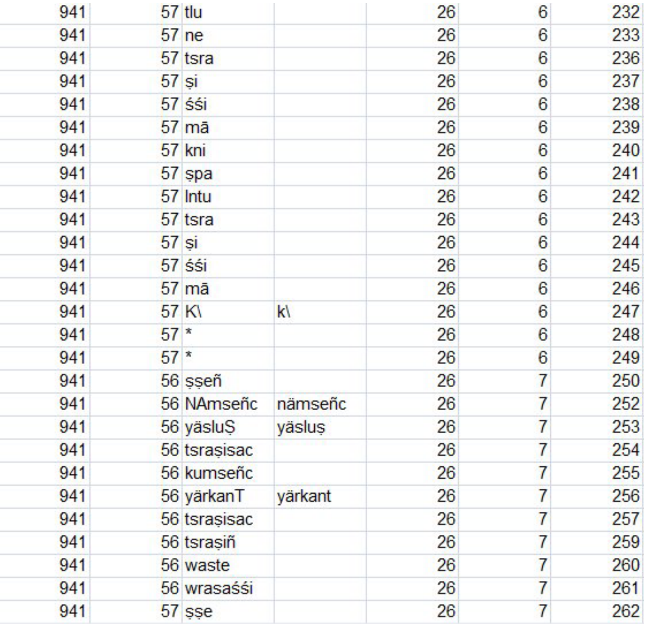
Text level no. 3:
Manuscript: 634

Text level no. 4:
Part:
Text level no. 5:
Line:
lookup
reset

OR
Switch to Word Index

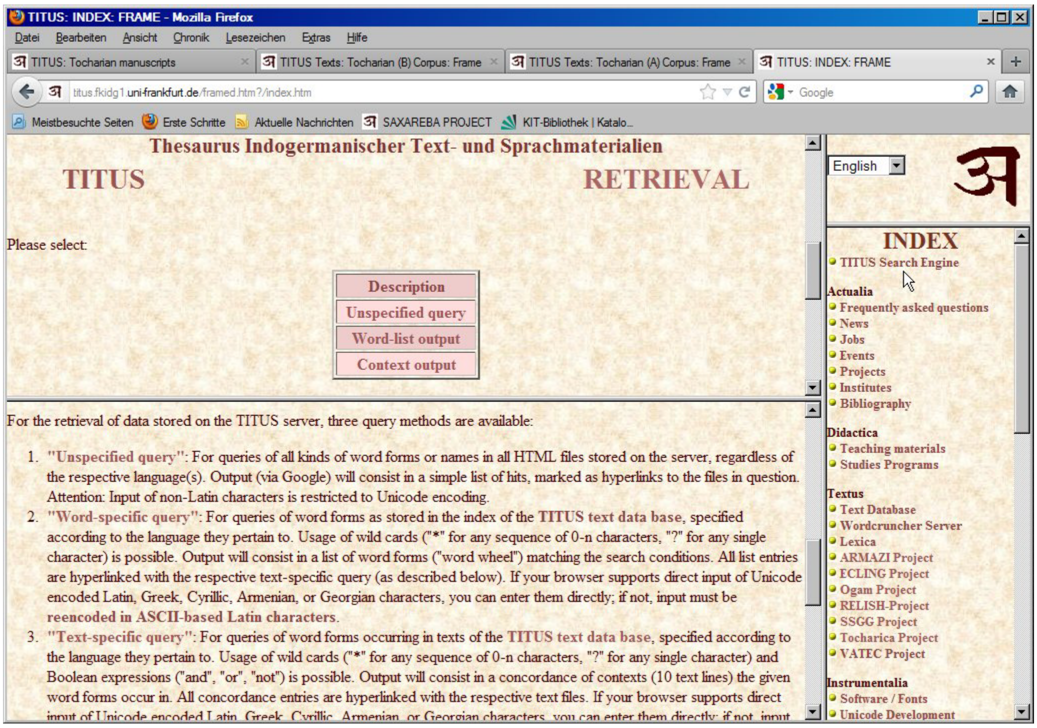
Attention! This document is encoded in Unicode (UTF-8) format. To display the scripts

Figure 5



941	57	tlu		26	6	232
941	57	ne		26	6	233
941	57	tsra		26	6	236
941	57	ṣi		26	6	237
941	57	śśi		26	6	238
941	57	mā		26	6	239
941	57	kni		26	6	240
941	57	ṣpa		26	6	241
941	57	lntu		26	6	242
941	57	tsra		26	6	243
941	57	ṣi		26	6	244
941	57	śśi		26	6	245
941	57	mā		26	6	246
941	57	K\	k\	26	6	247
941	57	*		26	6	248
941	57	*		26	6	249
941	56	ṣṣeñ		26	7	250
941	56	NAmseñc	nämseñc	26	7	252
941	56	yäsluṢ	yäsluṣ	26	7	253
941	56	tsraṣisac		26	7	254
941	56	kumseñc		26	7	255
941	56	yärkanT	yärkant	26	7	256
941	56	tsraṣisac		26	7	257
941	56	tsraṣiñ		26	7	259
941	56	waste		26	7	260
941	56	wrasaśśi		26	7	261
941	57	ṣṣe		26	7	262

Figure 6


TITUS: INDEX: FRAME - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian manuscripts
TITUS Texts: Tocharian (B) Corpus: Frame
TITUS Texts: Tocharian (A) Corpus: Frame
TITUS: INDEX: FRAME
titus.fkidg1.uni-frankfurt.de/framed.htm?/index.htm
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...
English
अ

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS RETRIEVAL

Please select:

Description
Unspecified query
Word-list output
Context output

For the retrieval of data stored on the TITUS server, three query methods are available:

1. "Unspecified query": For queries of all kinds of word forms or names in all HTML files stored on the server, regardless of the respective language(s). Output (via Google) will consist in a simple list of hits, marked as hyperlinks to the files in question. Attention: Input of non-Latin characters is restricted to Unicode encoding.
2. "Word-specific query": For queries of word forms as stored in the index of the TITUS text data base, specified according to the language they pertain to. Usage of wild cards ("*" for any sequence of 0-n characters, "?" for any single character) is possible. Output will consist in a list of word forms ("word wheel") matching the search conditions. All list entries are hyperlinked with the respective text-specific query (as described below). If your browser supports direct input of Unicode encoded Latin, Greek, Cyrillic, Armenian, or Georgian characters, you can enter them directly; if not, input must be reencoded in ASCII-based Latin characters.
3. "Text-specific query": For queries of word forms occurring in texts of the TITUS text data base, specified according to the language they pertain to. Usage of wild cards ("*" for any sequence of 0-n characters, "?" for any single character) and Boolean expressions ("and", "or", "not") is possible. Output will consist in a concordance of contexts (10 text lines) the given word forms occur in. All concordance entries are hyperlinked with the respective text files. If your browser supports direct input of Unicode encoded Latin, Greek, Cyrillic, Armenian, or Georgian characters, you can enter them directly; if not, input

INDEX
- TITUS Search Engine

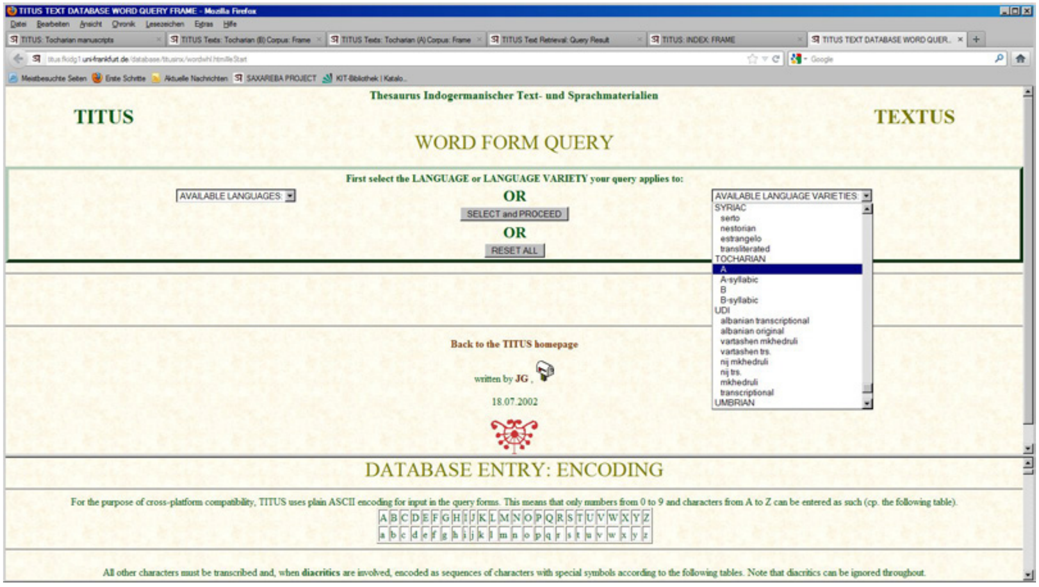
Actualia
- Frequently asked questions
- News
- Jobs
- Events
- Projects
- Institutes
- Bibliography

Didactica
- Teaching materials
- Studies Programs

Textus
- Text Database
- Wordcruncher Server
- Lexica
- ARMAZI Project
- ECLING Project
- Ogam Project
- RELISH-Project
- SSGG Project
- Tocharica Project
- VATEC Project

Instrumentalia
- Software / Fonts
- Unicode Development

Figure 7


TITUS TEXT DATABASE WORD QUERY FRAME - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian manuscripts
TITUS Texts: Tocharian (B) Corpus: Frame
TITUS Texts: Tocharian (A) Corpus: Frame
TITUS Text Retrieval: Query Result
TITUS: INDEX: FRAME
TITUS TEXT DATABASE WORD QUER...
titus.fkidg1.uni-frankfurt.de/database/titusinx/wordwhl.htm#Start
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

WORD FORM QUERY

First select the LANGUAGE or LANGUAGE VARIETY your query applies to:

AVAILABLE LANGUAGES:

OR

SELECT and PROCEED

OR

RESET ALL

AVAILABLE LANGUAGE VARIETIES:
SYRIAC
serto
nestorian
estrangelo
transliterated
TOCHARIAN
A
A-syllabic
B
B-syllabic
UDI
albanian transcriptional
albanian original
vartashen mkhedruli
vartashen trs.
nij mkhedruli
nij trs.
mkhedruli
transcriptional
UMBRIAN

Back to the TITUS homepage

written by JG ,

18.07.2002

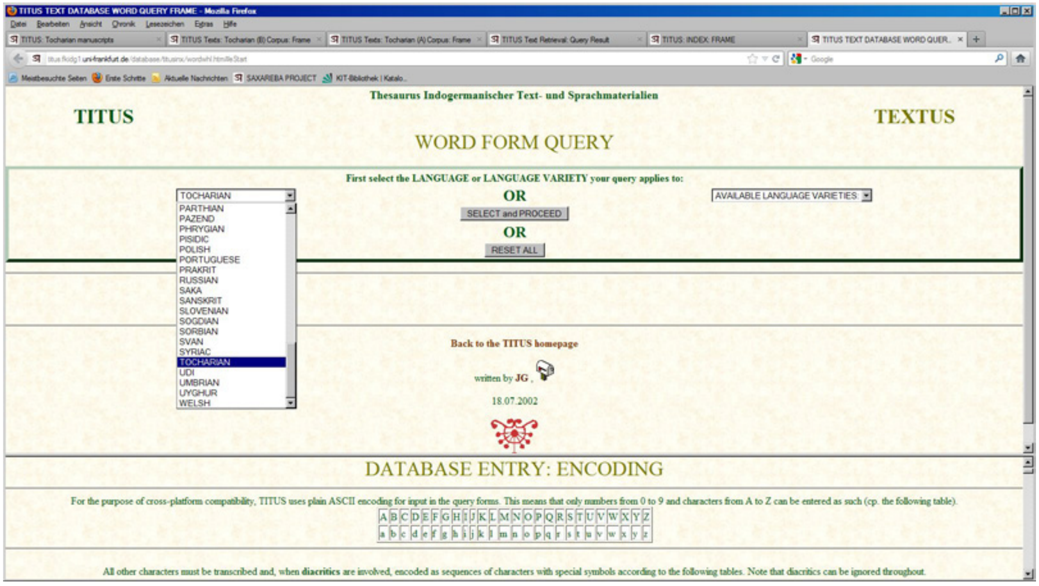
DATABASE ENTRY: ENCODING

For the purpose of cross-platform compatibility, TITUS uses plain ASCII encoding for input in the query forms. This means that only numbers from 0 to 9 and characters from A to Z can be entered as such (cp. the following table).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z

All other characters must be transcribed and, when **diacritics** are involved, encoded as sequences of characters with special symbols according to the following tables. Note that diacritics can be ignored throughout.

Figure 8


TITUS TEXT DATABASE WORD QUERY FRAME - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian manuscripts
TITUS Texts: Tocharian (B) Corpus: Frame
TITUS Texts: Tocharian (A) Corpus: Frame
TITUS Text Retrieval: Query Result
TITUS: INDEX: FRAME
TITUS TEXT DATABASE WORD QUER...
titus.fkidg1.uni-frankfurt.de/database/titusinx/wordwhl.htm#eStart
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...
TOCHARIAN
PARTHIAN
PAZEND
PHRYGIAN
PISIDIC
POLISH
PORTUGUESE
PRAKRIT
RUSSIAN
SAKA
SANSKRIT
SLOVENIAN
SOGDIAN
SORBIAN
SVAN
SYRIAC
TOCHARIAN
UDI
UMBRIAN
UYGHUR
WELSH
AVAILABLE LANGUAGE VARIETIES
SELECT and PROCEED
RESET ALL

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

WORD FORM QUERY

First select the LANGUAGE or LANGUAGE VARIETY your query applies to:

OR

OR

Back to the TITUS homepage

written by JG ,

18.07.2002

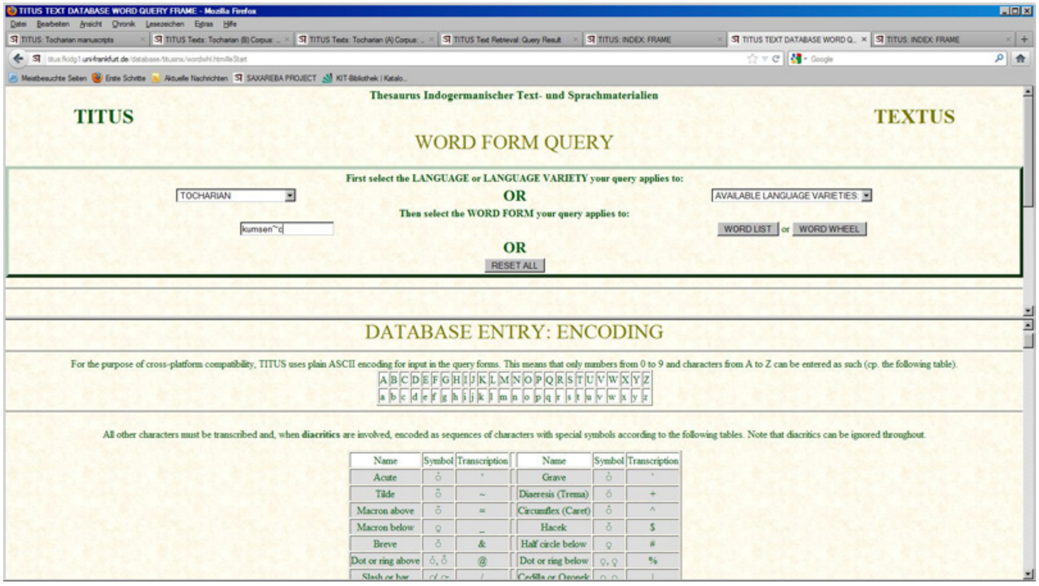
DATABASE ENTRY: ENCODING

For the purpose of cross-platform compatibility, TITUS uses plain ASCII encoding for input in the query forms. This means that only numbers from 0 to 9 and characters from A to Z can be entered as such (cp. the following table).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z

All other characters must be transcribed and, when **diacritics** are involved, encoded as sequences of characters with special symbols according to the following tables. Note that diacritics can be ignored throughout.

Figure 9


TITUS TEXT DATABASE WORD QUERY FRAME - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian manuscripts
TITUS Texts: Tocharian (B) Corpus: ...
TITUS Texts: Tocharian (A) Corpus: ...
TITUS Text Retrieval: Query Result
TITUS: INDEX: FRAME
TITUS TEXT DATABASE WORD Q...
TITUS: INDEX: FRAME
titus.fkidg1.uni-frankfurt.de/database/titusinx/wordwhl.htm#eStart
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...
TOCHARIAN
AVAILABLE LANGUAGE VARIETIES
WORD LIST
or
WORD WHEEL
RESET ALL

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

WORD FORM QUERY

First select the LANGUAGE or LANGUAGE VARIETY your query applies to:

OR

Then select the WORD FORM your query applies to:

kumsen~c

OR

DATABASE ENTRY: ENCODING

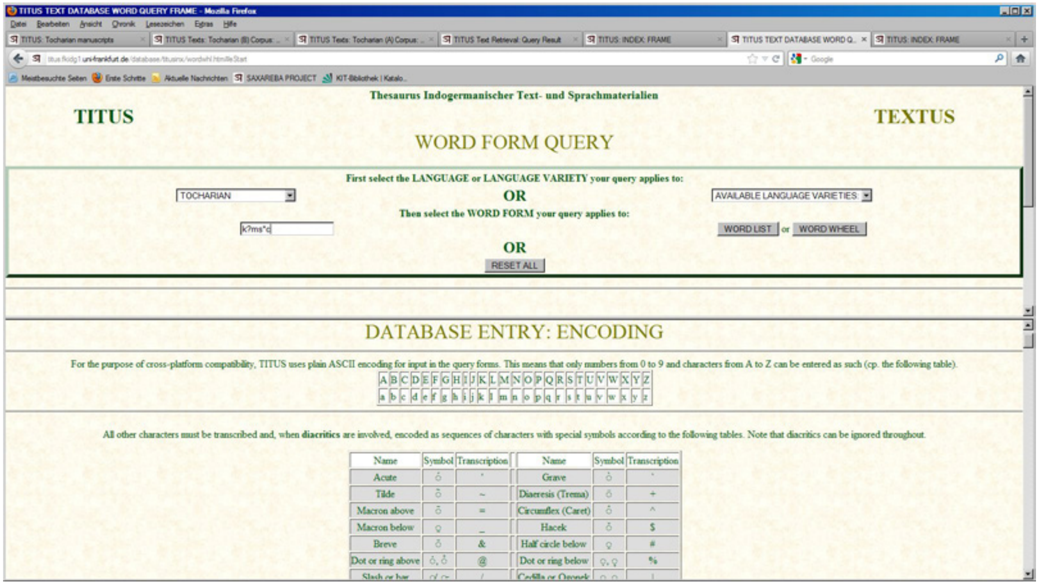
For the purpose of cross-platform compatibility, TITUS uses plain ASCII encoding for input in the query forms. This means that only numbers from 0 to 9 and characters from A to Z can be entered as such (cp. the following table).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z

All other characters must be transcribed and, when **diacritics** are involved, encoded as sequences of characters with special symbols according to the following tables. Note that diacritics can be ignored throughout.

Name	Symbol	Transcription	Name	Symbol	Transcription
Acute	ó	'	Grave	ò	`
Tilde	õ	~	Diaeresis (Trema)	ö	+
Macron above	ō	=	Circumflex (Caret)	ô	^
Macron below	o̱	_	Hacek	ǒ	$
Breve	ŏ	&	Half circle below	o̮	#
Dot or ring above	ȯ, o̊	@	Dot or ring below	ọ, o̥	%
Slash or bar	[illegible]	/	Cedilla or Ogonek	[illegible]	|

Figure 10


TITUS TEXT DATABASE WORD QUERY FRAME - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian manuscripts
TITUS Texts: Tocharian (B) Corpus: ...
TITUS Texts: Tocharian (A) Corpus: ...
TITUS Text Retrieval: Query Result
TITUS: INDEX: FRAME
TITUS TEXT DATABASE WORD Q...
TITUS: INDEX: FRAME
titus.fkidg1.uni-frankfurt.de/database/titusinx/wordwhl.htm#eStart
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

WORD FORM QUERY

First select the LANGUAGE or LANGUAGE VARIETY your query applies to:

TOCHARIAN OR AVAILABLE LANGUAGE VARIETIES

Then select the WORD FORM your query applies to:

k?ms*c

WORD LIST or WORD WHEEL

OR

RESET ALL

DATABASE ENTRY: ENCODING

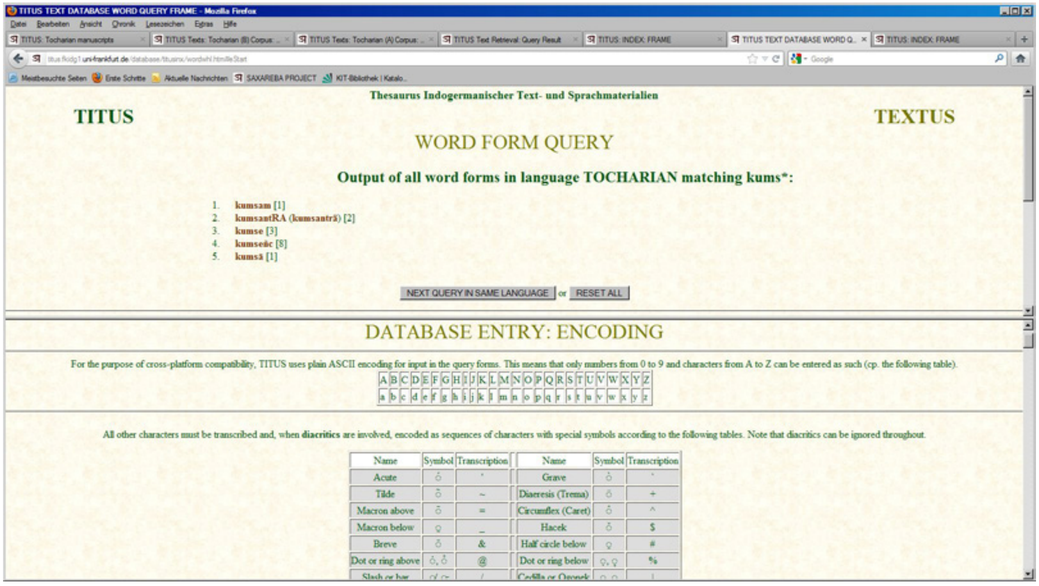
For the purpose of cross-platform compatibility, TITUS uses plain ASCII encoding for input in the query forms. This means that only numbers from 0 to 9 and characters from A to Z can be entered as such (cp. the following table).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z

All other characters must be transcribed and, when **diacritics** are involved, encoded as sequences of characters with special symbols according to the following tables. Note that diacritics can be ignored throughout.

Name	Symbol	Transcription	Name	Symbol	Transcription
Acute	ó	'	Grave	ò	`
Tilde	õ	~	Diaeresis (Trema)	ö	+
Macron above	ō	=	Circumflex (Caret)	ô	^
Macron below	o̱	_	Hacek	ǒ	$
Breve	ŏ	&	Half circle below	o̯	#
Dot or ring above	ȯ, o̊	@	Dot or ring below	ọ, o̥	%
Slash or bar	[illegible]	/	Cedilla or Ogonek	[illegible]	[illegible]

Figure 11


TITUS TEXT DATABASE WORD QUERY FRAME - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian manuscripts
TITUS Texts: Tocharian (B) Corpus: ...
TITUS Texts: Tocharian (A) Corpus: ...
TITUS Text Retrieval: Query Result
TITUS: INDEX: FRAME
TITUS TEXT DATABASE WORD Q...
TITUS: INDEX: FRAME
titus.fkidg1.uni-frankfurt.de/database/titusinx/wordwhl.htm#eStart
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

WORD FORM QUERY

Output of all word forms in language TOCHARIAN matching kums*:

1. **kumsam** [1]
2. **kumsantRA** (**kumsantrā**) [2]
3. **kumse** [3]
4. **kumseñc** [8]
5. **kumsā** [1]

NEXT QUERY IN SAME LANGUAGE or RESET ALL

DATABASE ENTRY: ENCODING

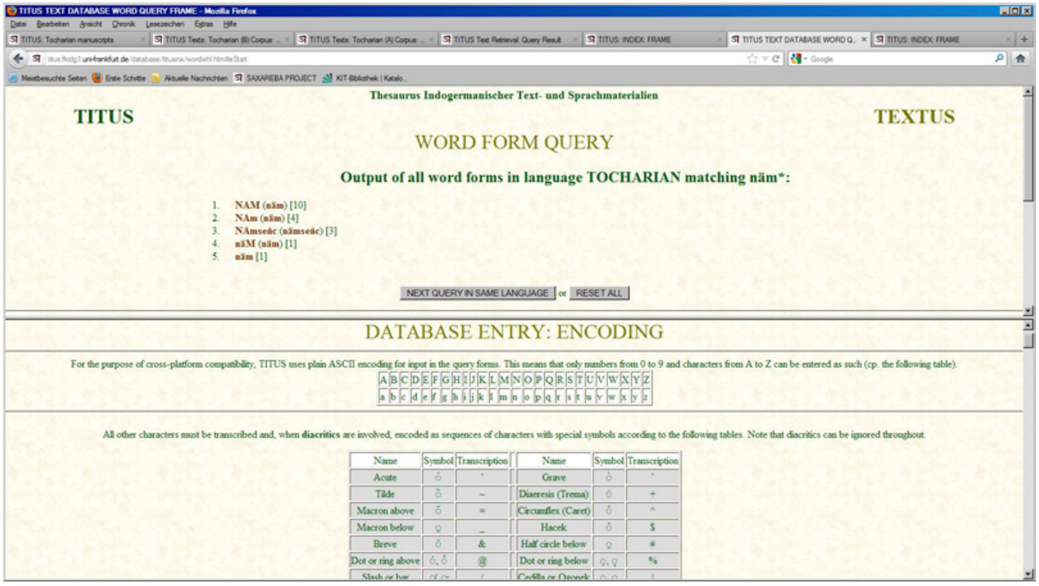
For the purpose of cross-platform compatibility, TITUS uses plain ASCII encoding for input in the query forms. This means that only numbers from 0 to 9 and characters from A to Z can be entered as such (cp. the following table).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z

All other characters must be transcribed and, when **diacritics** are involved, encoded as sequences of characters with special symbols according to the following tables. Note that diacritics can be ignored throughout.

Name	Symbol	Transcription	Name	Symbol	Transcription
Acute	ó	'	Grave	ò	`
Tilde	õ	~	Diaeresis (Trema)	ö	+
Macron above	ō	=	Circumflex (Caret)	ô	^
Macron below	o̱	_	Hacek	ǒ	$
Breve	ŏ	&	Half circle below	o̯	#
Dot or ring above	ȯ, o̊	@	Dot or ring below	ọ, o̥	%
Slash or bar	[illegible]	/	Cedilla or Ogonek	[illegible]	[illegible]

Figure 12


TITUS TEXT DATABASE WORD QUERY FRAME - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian manuscripts
TITUS Texts: Tocharian (B) Corpus: ...
TITUS Texts: Tocharian (A) Corpus: ...
TITUS Text Retrieval: Query Result
TITUS: INDEX: FRAME
TITUS TEXT DATABASE WORD Q...
TITUS: INDEX: FRAME
titus.fkidg1.uni-frankfurt.de/database/titusinx/wordwhl.htm#Start
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

WORD FORM QUERY

Output of all word forms in language TOCHARIAN matching näm*:

1. NAM (näm) [10]
2. NAm (näm) [4]
3. NAmseñc (nämseñc) [3]
4. näM (näm) [1]
5. näm [1]

NEXT QUERY IN SAME LANGUAGE or RESET ALL

DATABASE ENTRY: ENCODING

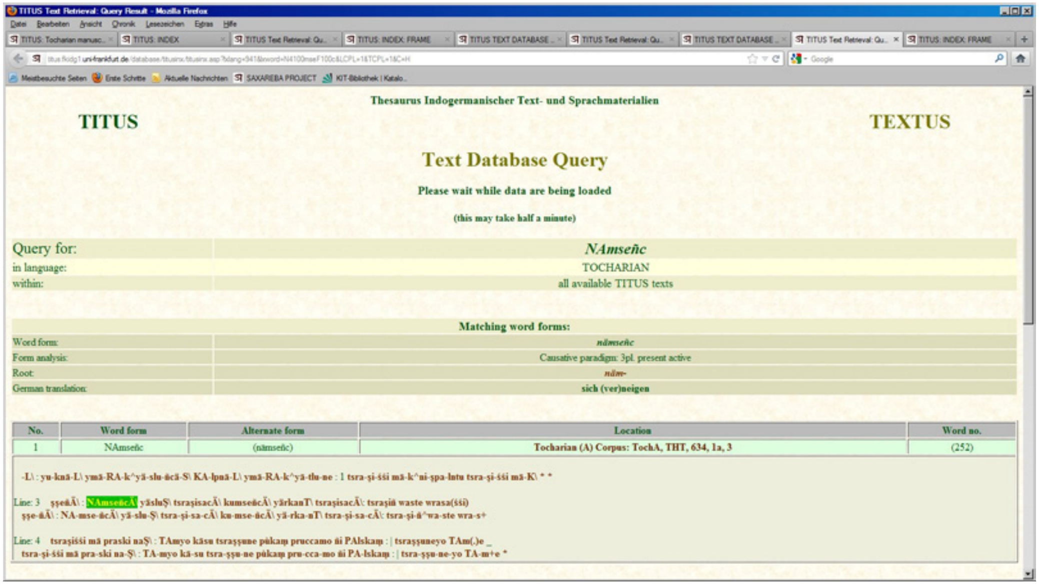
For the purpose of cross-platform compatibility, TITUS uses plain ASCII encoding for input in the query forms. This means that only numbers from 0 to 9 and characters from A to Z can be entered as such (cp. the following table).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z

All other characters must be transcribed and, when **diacritics** are involved, encoded as sequences of characters with special symbols according to the following tables. Note that diacritics can be ignored throughout.

Name	Symbol	Transcription	Name	Symbol	Transcription
Acute	ó	'	Grave	ò	`
Tilde	õ	~	Diaeresis (Trema)	ö	+
Macron above	ō	=	Circumflex (Caret)	ô	^
Macron below	o̱	_	Hacek	ǒ	$
Breve	ŏ	&	Half circle below	o̯	#
Dot or ring above	ȯ, o̊	@	Dot or ring below	ọ, o̥	%
Slash or bar	[illegible]	/	Cedilla or Oronek	[illegible]	[illegible]

Figure 13


TITUS Text Retrieval: Query Result - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian manusc...
TITUS: INDEX
TITUS Text Retrieval: Qu...
TITUS: INDEX: FRAME
TITUS TEXT DATABASE ...
TITUS Text Retrieval: Qu...
TITUS TEXT DATABASE ...
TITUS Text Retrieval: Qu...
TITUS: INDEX: FRAME
titus.fkidg1.uni-frankfurt.de/database/titusinx/titusinx.asp?tdang=941&word=N4100mseF100c&LCPL=1&TCPL=1&C=H
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

Text Database Query

Please wait while data are being loaded

(this may take half a minute)

Query for:	*NAmseñc*
in language:	TOCHARIAN
within:	all available TITUS texts

Matching word forms:	
Word form:	*nāmseñc*
Form analysis:	Causative paradigm: 3pl. present active
Root:	*nām-*
German translation:	sich (ver)neigen

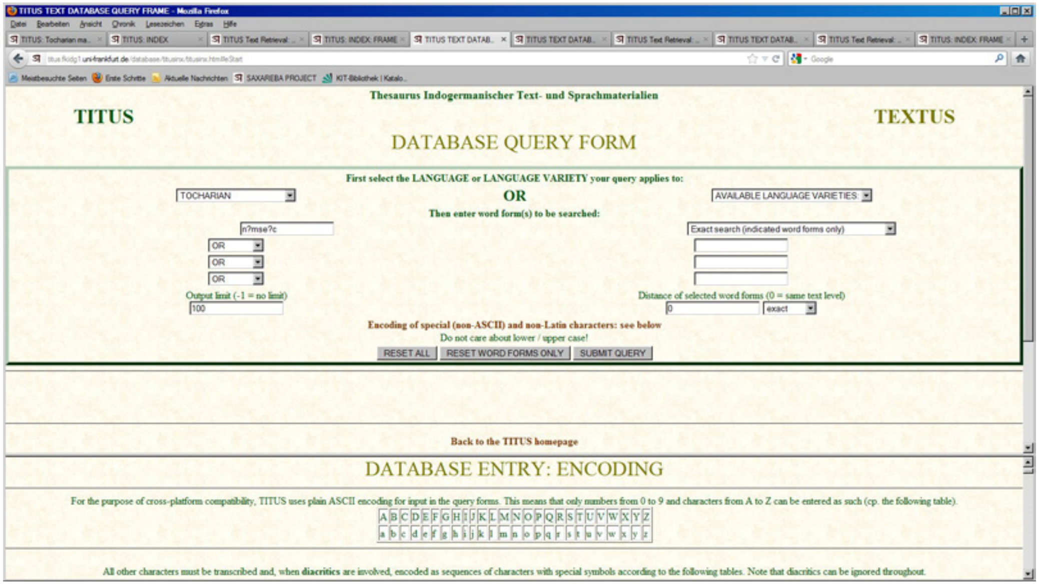
No.	Word form	Alternate form	Location	Word no.
1	NAmseñc	(nämseñc)	Tocharian (A) Corpus: TochA, THT, 634, 1a, 3	(252)

-L\ : yu-knā-L\ ymā-RA-k^yā-slu-ñcā-S\ KA-lpnā-L\ ymā-RA-k^yā-tlu-ne : 1 tsra-şi-śśi mā-k^ni-şpa-ltu tsra-şi-śśi mā-K\ * *

Line: 3 şşeñÃ\ : NAmseñcÃ\ yāsluŞ\ tsraşisacÃ\ kumseñcÃ\ yārkanT\ tsraşisacÃ\: tsraşiñ waste wrasa(śśi)
şşe-ñÃ\ : NA-mse-ñcÃ\ yā-slu-Ş\ tsra-şi-sa-cÃ\ ku-mse-ñcÃ\ yā-rka-nT\ tsra-şi-sa-cÃ\: tsra-şi-ñ^wa-ste wra-s+

Line: 4 tsraşiśśi mā praski naŞ\ : TAmyo kāsu tsraşşune pūkaṃ pruccamo ñi PAlskaṃ : | tsraşşuneyo TAm(.)e _
tsra-şi-śśi mā pra-ski na-Ş\ : TA-myo kā-su tsra-şşu-ne pūkaṃ pru-cca-mo ñi PA-lskaṃ : | tsra-şşu-ne-yo TA-m+e *

Figure 14


TITUS TEXT DATABASE QUERY FRAME - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian ma... TITUS: INDEX TITUS Text Retrieval: ... TITUS: INDEX: FRAME TITUS TEXT DATAB... TITUS TEXT DATAB... TITUS Text Retrieval: ... TITUS TEXT DATAB... TITUS Text Retrieval: ... TITUS: INDEX: FRAME
titus.fkidg1.uni-frankfurt.de/database/titusinx/titusinx.htm#eStart
Google
Meistbesuchte Seiten Erste Schritte Aktuelle Nachrichten SAXAREBA PROJECT KIT-Bibliothek | Katalo...
RESET ALL
RESET WORD FORMS ONLY
SUBMIT QUERY

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

DATABASE QUERY FORM

First select the LANGUAGE or LANGUAGE VARIETY your query applies to:

TOCHARIAN OR AVAILABLE LANGUAGE VARIETIES

Then enter word form(s) to be searched:

n?mse?c

Exact search (indicated word forms only)

OR

OR

OR

Output limit (-1 = no limit) 100

Distance of selected word forms (0 = same text level) 0 exact

Encoding of special (non-ASCII) and non-Latin characters: see below

Do not care about lower / upper case!

Back to the TITUS homepage

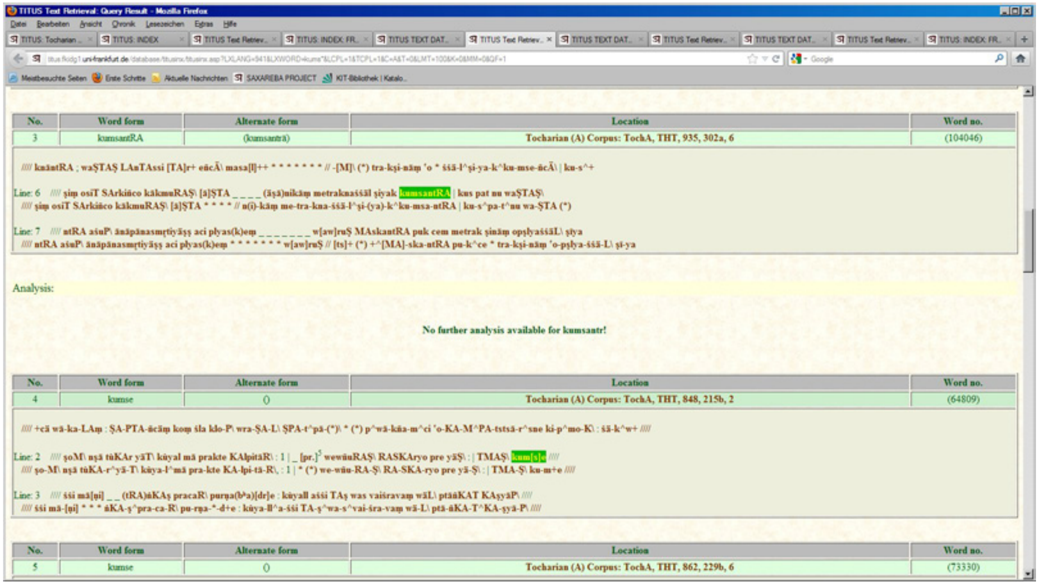
DATABASE ENTRY: ENCODING

For the purpose of cross-platform compatibility, TITUS uses plain ASCII encoding for input in the query forms. This means that only numbers from 0 to 9 and characters from A to Z can be entered as such (cp. the following table).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z

All other characters must be transcribed and, when **diacritics** are involved, encoded as sequences of characters with special symbols according to the following tables. Note that diacritics can be ignored throughout.

Figure 15


TITUS Text Retrieval: Query Result - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian ... TITUS: INDEX TITUS Text Retriev... TITUS: INDEX: FR... TITUS TEXT DAT... TITUS Text Retriev... TITUS TEXT DAT... TITUS Text Retriev... TITUS TEXT DAT... TITUS Text Retriev... TITUS: INDEX: FR...
titus.fkidg1.uni-frankfurt.de/database/titusinx/titusinx.asp?LXLANG=941&LXWORD=kums*&LCPL=1&TCPL=1&C=A&T=0&LMT=100&K=0&MM=0&QF=1
Google
Meistbesuchte Seiten Erste Schritte Aktuelle Nachrichten SAXAREBA PROJECT KIT-Bibliothek | Katalo...

No.	Word form	Alternate form	Location	Word no.
3	kumsantRA	(kumsantrā)	Tocharian (A) Corpus: TochA, THT, 935, 302a, 6	(104046)

//// knāntRA ; waṢTAṢ LAnTAssi [TA]r+ eñcĀ\ masa[l]++ * * * * * * * // -[M]\ (*) tra-kṣi-nām 'o * śśā-l^ṣi-ya-k^ku-mse-ñcĀ\ | ku-s^+

Line: 6 //// ṣim osiT SArkiñco kākmuRAṢ\ [ā]ṢTA _ _ _ _ (āṣā)nikām metraknaśśāl ṣiyak kumsantRA | kus pat nu waṢTAṢ\
//// ṣim osiT SArkiñco kākmuRAṢ\ [ā]ṢTA * * * * // n(i)-kām me-tra-kna-śśā-l^ṣi-(ya)-k^ku-msa-ntRA | ku-s^pa-t^nu wa-ṢTA (*)

Line: 7 //// ntRA aśuP\ ānāpānasmṛtiyāṣṣ aci plyas(k)em _ _ _ _ _ _ _ w[aw]ruṢ MAskantRA puk cem metrak ṣinām opṣlyaśśāL\ ṣiya
//// ntRA aśuP\ ānāpānasmṛtiyāṣṣ aci plyas(k)em * * * * * * * w[aw]ruṢ // [ts]+ (*) +^[MA]-ska-ntRA pu-k^ce * tra-kṣi-nām 'o-pṣlya-śśā-L\ ṣi-ya

Analysis:

No further analysis available for kumsantr!

No.	Word form	Alternate form	Location	Word no.
4	kumse	()	Tocharian (A) Corpus: TochA, THT, 848, 215b, 2	(64809)

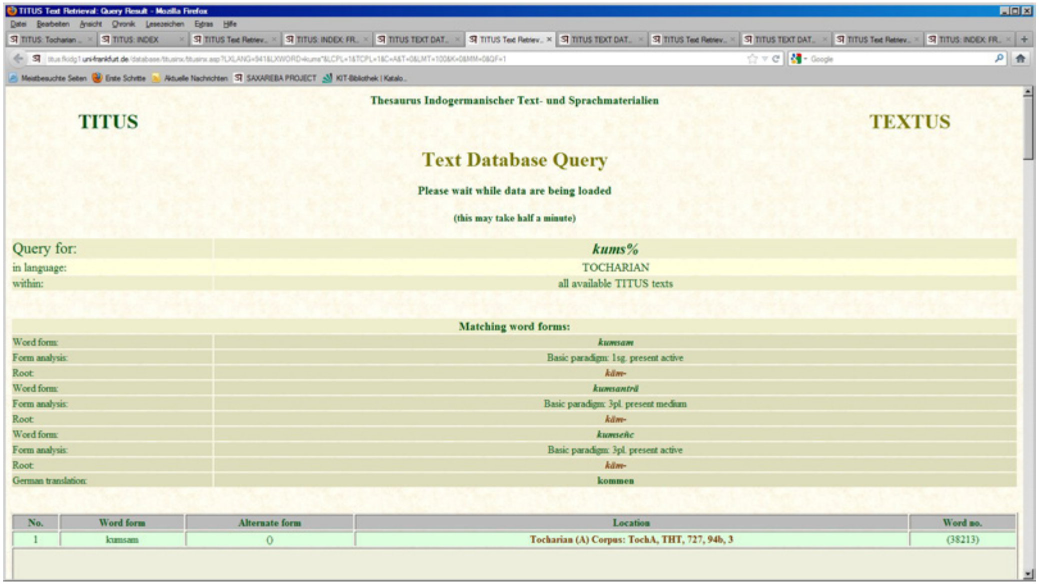
//// +cā wā-ka-LAm : ṢA-PTA-ñcām kom śla klo-P\ wra-ṢA-L\ ṢPA-t^pā-(*)\ * (*) p^wā-kña-m^ci 'o-KA-M^PA-tstsā-r^sne ki-p^mo-K\ : śā-k^w+ ////

Line: 2 //// ṣoM\ nṣā tūKAr yāT\ kuyal mā prakte KAlpitāR\ : 1 | _ [pr.]^5 wewñuRAṢ\ RASKAryo pre yāṢ\ : | TMAṢ\ kum[s]e ////
//// ṣo-M\ nṣā tūKA-r^yā-T\ kuya-l^mā pra-kte KA-lpi-tā-R\ : 1 | * (*) we-wñu-RA-Ṣ\ RA-SKA-ryo pre yā-Ṣ\ : | TMA-Ṣ\ ku-m+e ////

Line: 3 //// śśi mā[ṇi] _ _ (tRA)ñKAṣ pracaR\ purṇa(bʰa)[dr]e : kuyall aśśi TAṣ was vaiśravam wāL\ ptāñKAT KAṣyāP\ ////
//// śśi mā-[ṇi] * * * ñKA-ṣ^pra-ca-R\ pu-rṇa-*-d+e : kuya-ll^a-śśi TA-ṣ^wa-s^vai-śra-vam wā-L\ ptā-ñKA-T^KA-ṣyā-P\ ////

No.	Word form	Alternate form	Location	Word no.
5	kumse	()	Tocharian (A) Corpus: TochA, THT, 862, 229b, 6	(73330)

Figure 16


TITUS Text Retrieval: Query Result - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian ... TITUS: INDEX TITUS Text Retriev... TITUS: INDEX: FR... TITUS TEXT DAT... TITUS Text Retriev... TITUS TEXT DAT... TITUS Text Retriev... TITUS TEXT DAT... TITUS Text Retriev... TITUS: INDEX: FR...
Google
Meistbesuchte Seiten Erste Schritte Aktuelle Nachrichten SAXAREBA PROJECT KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

Text Database Query

Please wait while data are being loaded

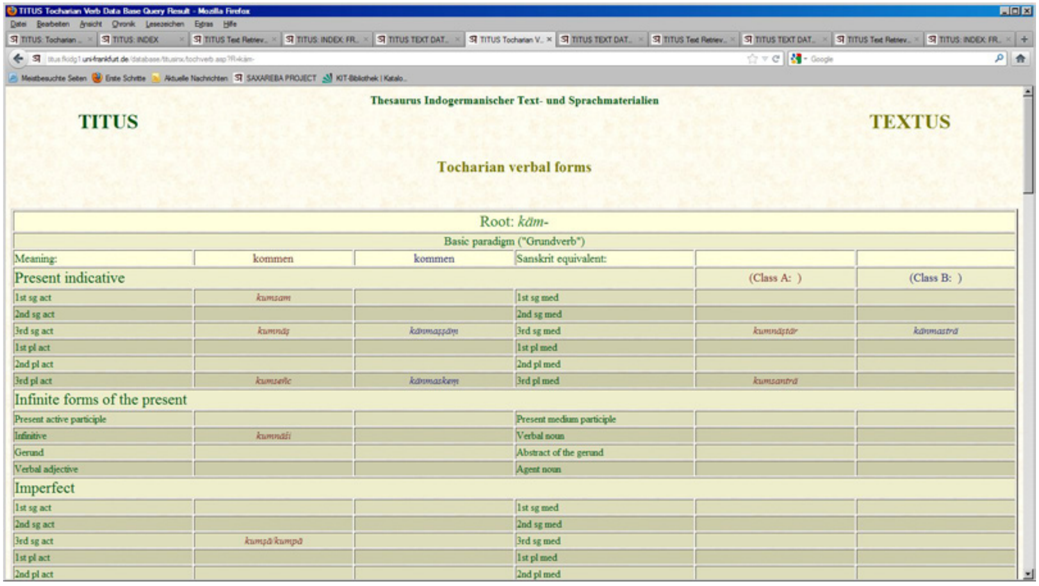
(this may take half a minute)

Query for:	*kums%*
in language:	TOCHARIAN
within:	all available TITUS texts

Matching word forms:	
Word form:	*kumsam*
Form analysis:	Basic paradigm: 1sg. present active
Root:	*käm-*
Word form:	*kumsanträ*
Form analysis:	Basic paradigm: 3pl. present medium
Root:	*käm-*
Word form:	*kumseñc*
Form analysis:	Basic paradigm: 3pl. present active
Root:	*käm-*
German translation:	**kommen**

No.	Word form	Alternate form	Location	Word no.
1	kumsam	()	Tocharian (A) Corpus: TochA, THT, 727, 94b, 3	(38213)

Fugure 17


TITUS Tocharian Verb Data Base Query Result - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: Tocharian ... TITUS: INDEX TITUS Text Retriev... TITUS: INDEX: FR... TITUS TEXT DAT... TITUS Tocharian V... TITUS TEXT DAT... TITUS Text Retriev... TITUS TEXT DAT... TITUS Text Retriev... TITUS: INDEX: FR...
Google
Meistbesuchte Seiten Erste Schritte Aktuelle Nachrichten SAXAREBA PROJECT KIT-Bibliothek | Katalo...

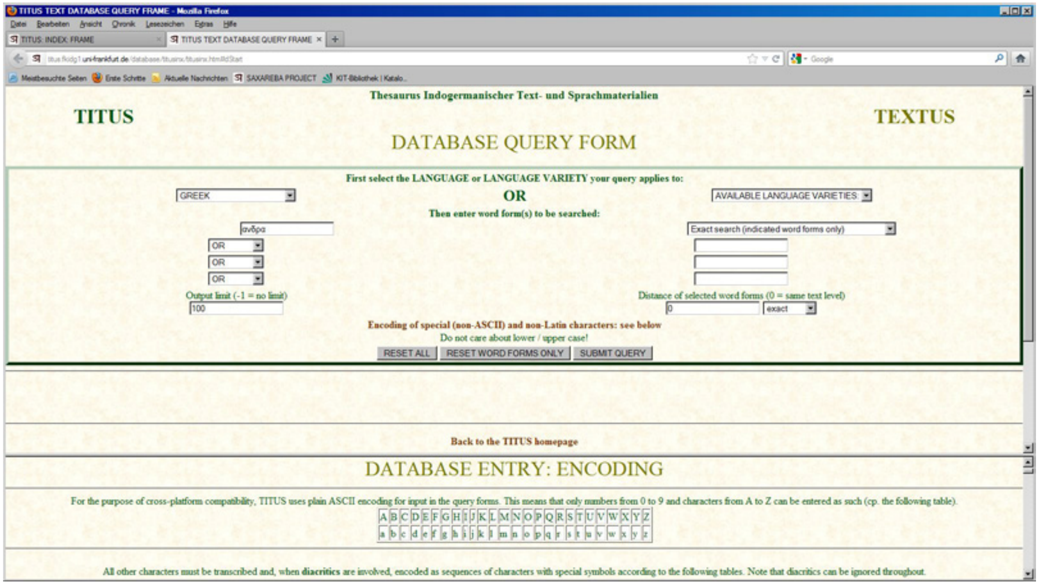
Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

Tocharian verbal forms

Root: *käm-*					
Basic paradigm ("Grundverb")					
Meaning:	kommen	kommen	Sanskrit equivalent:		
Present indicative				(Class A:)	(Class B:)
1st sg act	*kumsam*		1st sg med		
2nd sg act			2nd sg med		
3rd sg act	*kumnäṣ*	*känmaṣṣäṃ*	3rd sg med	*kumnäṣtär*	*känmasträ*
1st pl act			1st pl med		
2nd pl act			2nd pl med		
3rd pl act	*kumseñc*	*känmaskeṃ*	3rd pl med	*kumsanträ*	
Infinite forms of the present					
Present active participle			Present medium participle		
Infinitive	*kumnässi*		Verbal noun		
Gerund			Abstract of the gerund		
Verbal adjective			Agent noun		
Imperfect					
1st sg act			1st sg med		
2nd sg act			2nd sg med		
3rd sg act	*kumṣä/kumpä*		3rd sg med		
1st pl act			1st pl med		
2nd pl act			2nd pl med		

Figure 18


TITUS TEXT DATABASE QUERY FRAME - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: INDEX: FRAME
TITUS TEXT DATABASE QUERY FRAME
titus.fkidg1.uni-frankfurt.de/database/titusinx/titusinx.htm#dStart
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

DATABASE QUERY FORM

First select the LANGUAGE or LANGUAGE VARIETY your query applies to:

GREEK OR AVAILABLE LANGUAGE VARIETIES:

Then enter word form(s) to be searched:

ανδρα

Exact search (indicated word forms only)

OR

OR

OR

Output limit (-1 = no limit): 100

Distance of selected word forms (0 = same text level): 0 exact

Encoding of special (non-ASCII) and non-Latin characters: see below

Do not care about lower / upper case!

RESET ALL | RESET WORD FORMS ONLY | SUBMIT QUERY

Back to the TITUS homepage

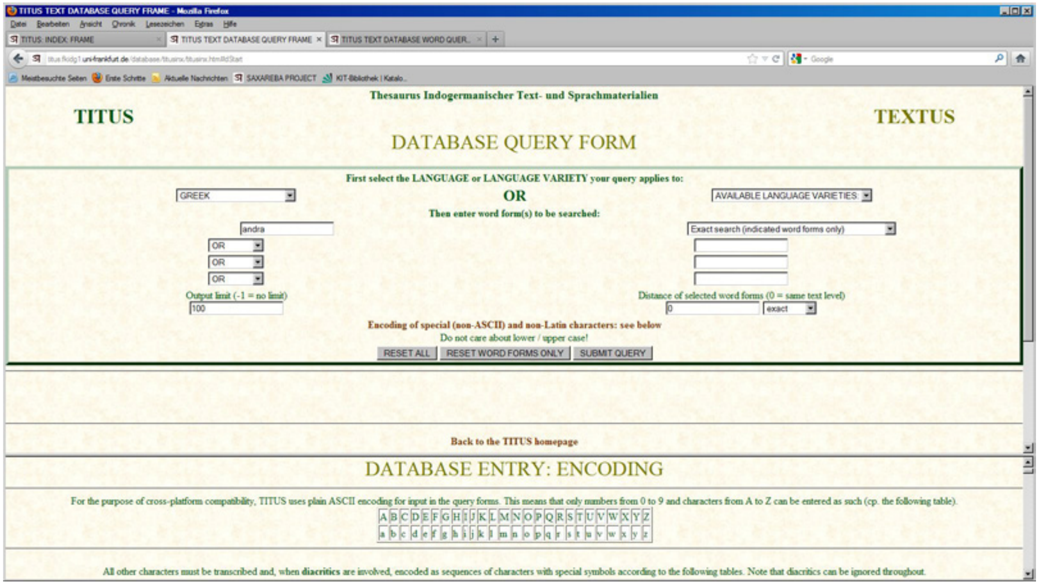
DATABASE ENTRY: ENCODING

For the purpose of cross-platform compatibility, TITUS uses plain ASCII encoding for input in the query forms. This means that only numbers from 0 to 9 and characters from A to Z can be entered as such (cp. the following table).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z

All other characters must be transcribed and, when **diacritics** are involved, encoded as sequences of characters with special symbols according to the following tables. Note that diacritics can be ignored throughout.

Figure 19


TITUS TEXT DATABASE QUERY FRAME - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: INDEX: FRAME
TITUS TEXT DATABASE QUERY FRAME
TITUS TEXT DATABASE WORD QUER...
titus.fkidg1.uni-frankfurt.de/database/titusinx/titusinx.htm#dStart
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

DATABASE QUERY FORM

First select the LANGUAGE or LANGUAGE VARIETY your query applies to:

GREEK OR AVAILABLE LANGUAGE VARIETIES:

Then enter word form(s) to be searched:

andra

Exact search (indicated word forms only)

OR

OR

OR

Output limit (-1 = no limit): 100

Distance of selected word forms (0 = same text level): 0 exact

Encoding of special (non-ASCII) and non-Latin characters: see below

Do not care about lower / upper case!

RESET ALL | RESET WORD FORMS ONLY | SUBMIT QUERY

Back to the TITUS homepage

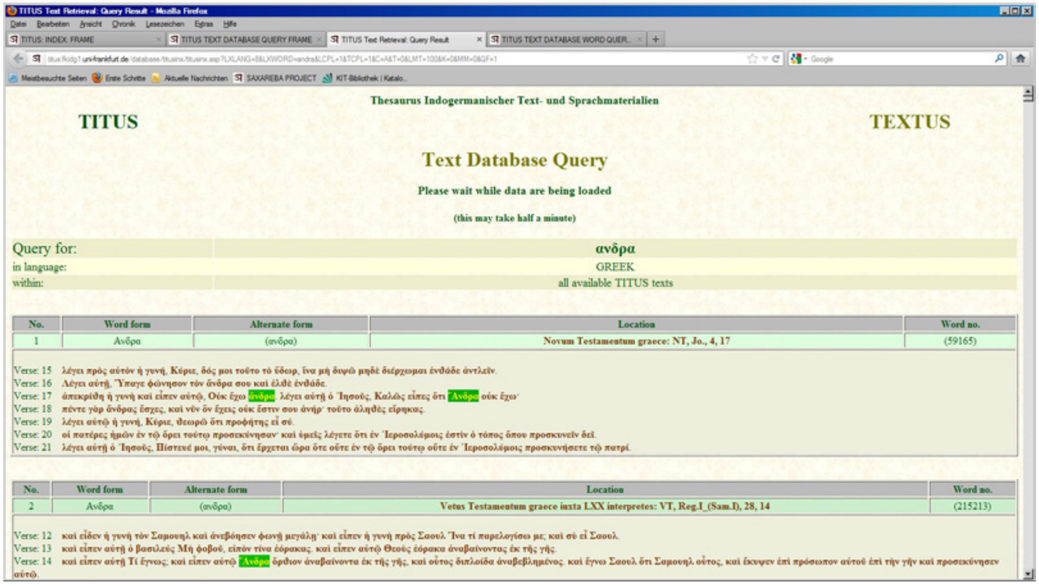
DATABASE ENTRY: ENCODING

For the purpose of cross-platform compatibility, TITUS uses plain ASCII encoding for input in the query forms. This means that only numbers from 0 to 9 and characters from A to Z can be entered as such (cp. the following table).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z

All other characters must be transcribed and, when **diacritics** are involved, encoded as sequences of characters with special symbols according to the following tables. Note that diacritics can be ignored throughout.

Figure 20


TITUS Text Retrieval: Query Result - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: INDEX: FRAME
TITUS TEXT DATABASE QUERY FRAME
TITUS Text Retrieval: Query Result
TITUS TEXT DATABASE WORD QUER...
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

Text Database Query

Please wait while data are being loaded

(this may take half a minute)

Query for:	ανδρα
in language:	GREEK
within:	all available TITUS texts

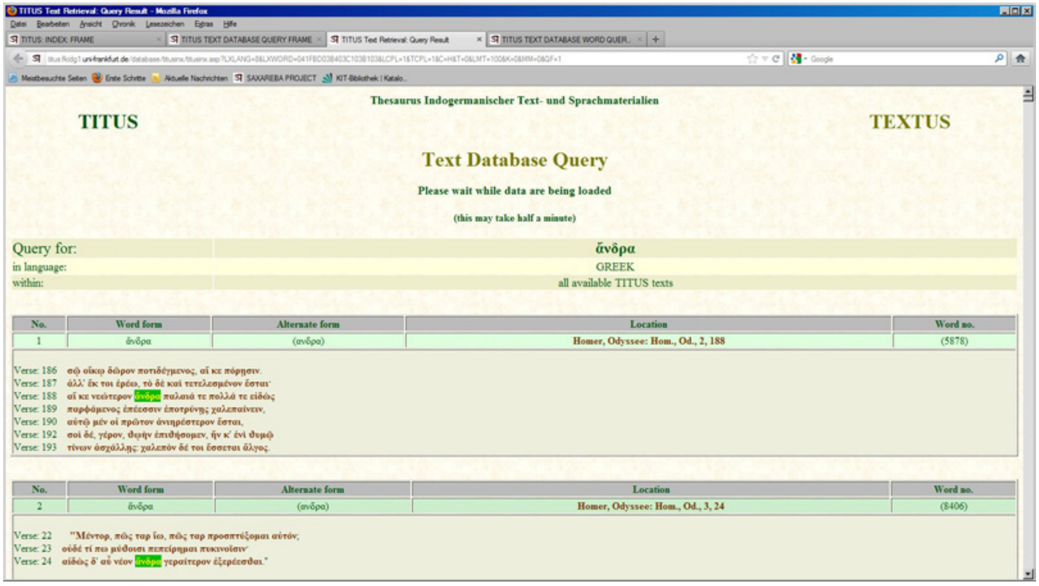
No.	Word form	Alternate form	Location	Word no.
1	Ανδρα	(ανδρα)	Novum Testamentum graece: NT, Jo., 4, 17	(59165)

Verse: 15 λέγει πρὸς αὐτὸν ἡ γυνή, Κύριε, δός μοι τοῦτο τὸ ὕδωρ, ἵνα μὴ διψῶ μηδὲ διέρχωμαι ἐνθάδε ἀντλεῖν.
Verse: 16 Λέγει αὐτῇ, Ὕπαγε φώνησον τὸν ἄνδρα σου καὶ ἐλθὲ ἐνθάδε.
Verse: 17 ἀπεκρίθη ἡ γυνὴ καὶ εἶπεν αὐτῷ, Οὐκ ἔχω ἄνδρα. λέγει αὐτῇ ὁ Ἰησοῦς, Καλῶς εἶπας ὅτι Ἄνδρα οὐκ ἔχω·
Verse: 18 πέντε γὰρ ἄνδρας ἔσχες, καὶ νῦν ὃν ἔχεις οὐκ ἔστιν σου ἀνήρ· τοῦτο ἀληθὲς εἴρηκας.
Verse: 19 λέγει αὐτῷ ἡ γυνή, Κύριε, θεωρῶ ὅτι προφήτης εἶ σύ.
Verse: 20 οἱ πατέρες ἡμῶν ἐν τῷ ὄρει τούτῳ προσεκύνησαν· καὶ ὑμεῖς λέγετε ὅτι ἐν Ἱεροσολύμοις ἐστὶν ὁ τόπος ὅπου προσκυνεῖν δεῖ.
Verse: 21 λέγει αὐτῇ ὁ Ἰησοῦς, Πίστευέ μοι, γύναι, ὅτι ἔρχεται ὥρα ὅτε οὔτε ἐν τῷ ὄρει τούτῳ οὔτε ἐν Ἱεροσολύμοις προσκυνήσετε τῷ πατρί.

No.	Word form	Alternate form	Location	Word no.
2	Ανδρα	(ανδρα)	Vetus Testamentum graece iuxta LXX interpretes: VT, Reg.I_(Sam.I), 28, 14	(215213)

Verse: 12 καὶ εἶδεν ἡ γυνὴ τὸν Σαμουηλ καὶ ἀνεβόησεν φωνῇ μεγάλῃ· καὶ εἶπεν ἡ γυνὴ πρὸς Σαουλ Ἵνα τί παρελογίσω με; καὶ σὺ εἶ Σαουλ.
Verse: 13 καὶ εἶπεν αὐτῇ ὁ βασιλεύς Μὴ φοβοῦ, εἰπὸν τίνα ἑόρακας. καὶ εἶπεν αὐτῷ Θεοὺς ἑόρακα ἀναβαίνοντας ἐκ τῆς γῆς.
Verse: 14 καὶ εἶπεν αὐτῇ Τί ἔγνως; καὶ εἶπεν αὐτῷ Ἄνδρα ὄρθιον ἀναβαίνοντα ἐκ τῆς γῆς, καὶ οὗτος διπλοΐδα ἀναβεβλημένος. καὶ ἔγνω Σαουλ ὅτι Σαμουηλ οὗτος, καὶ ἔκυψεν ἐπὶ πρόσωπον αὐτοῦ ἐπὶ τὴν γῆν καὶ προσεκύνησεν αὐτῷ.

Figure 21


TITUS Text Retrieval: Query Result - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: INDEX: FRAME
TITUS TEXT DATABASE QUERY FRAME
TITUS Text Retrieval: Query Result
TITUS TEXT DATABASE WORD QUER...
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

Text Database Query

Please wait while data are being loaded

(this may take half a minute)

Query for:	ἄνδρα
in language:	GREEK
within:	all available TITUS texts

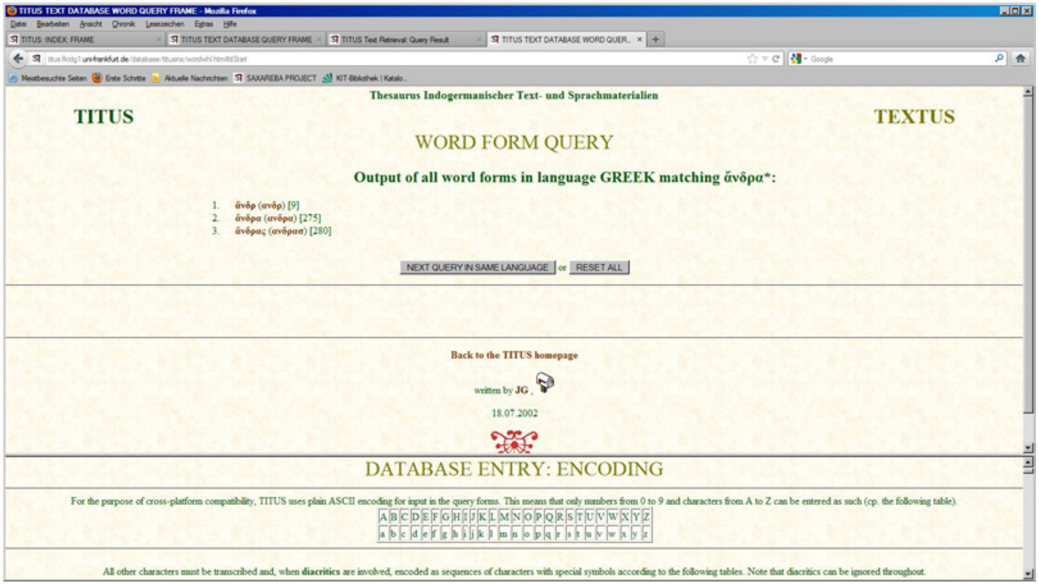
No.	Word form	Alternate form	Location	Word no.
1	ἄνδρα	(ανδρα)	Homer, Odyssee: Hom., Od., 2, 188	(5878)

Verse: 186 σῷ οἴκῳ δῶρον ποτιδέγμενος, αἴ κε πόρῃσιν.
Verse: 187 ἀλλ' ἔκ τοι ἐρέω, τὸ δὲ καὶ τετελεσμένον ἔσται·
Verse: 188 αἴ κε νεώτερον ἄνδρα παλαιά τε πολλά τε εἰδὼς
Verse: 189 παρφάμενος ἐπέεσσιν ἐποτρύνῃς χαλεπαίνειν,
Verse: 190 αὐτῷ μὲν οἱ πρῶτον ἀνιηρέστερον ἔσται,
Verse: 192 σοὶ δέ, γέρον, θωὴν ἐπιθήσομεν, ἥν κ' ἐνὶ θυμῷ
Verse: 193 τίνων ἀσχάλλῃς· χαλεπὸν δέ τοι ἔσσεται ἄλγος.

No.	Word form	Alternate form	Location	Word no.
2	ἄνδρα	(ανδρα)	Homer, Odyssee: Hom., Od., 3, 24	(8406)

Verse: 22 "Μέντορ, πῶς τ' ἄρ ἴω, πῶς τ' ἄρ προσπτύξομαι αὐτόν;
Verse: 23 οὐδέ τί πω μύθοισι πεπείρημαι πυκινοῖσιν·
Verse: 24 αἰδὼς δ' αὖ νέον ἄνδρα γεραίτερον ἐξερέεσθαι."

Figure 22


TITUS TEXT DATABASE WORD QUERY FRAME - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: INDEX: FRAME
TITUS TEXT DATABASE QUERY FRAME
TITUS Text Retrieval: Query Result
TITUS TEXT DATABASE WORD QUER...
titus.fkidg1.uni-frankfurt.de/database/titusinx/wordwhl.htm#dStart
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

WORD FORM QUERY

Output of all word forms in language GREEK matching ἄνδρα*:

1. ἄνδρ (ανδρ) [9]
2. ἄνδρα (ανδρα) [275]
3. ἄνδρας (ανδρασ) [280]

NEXT QUERY IN SAME LANGUAGE or RESET ALL

Back to the TITUS homepage

written by **JG** ,

18.07.2002

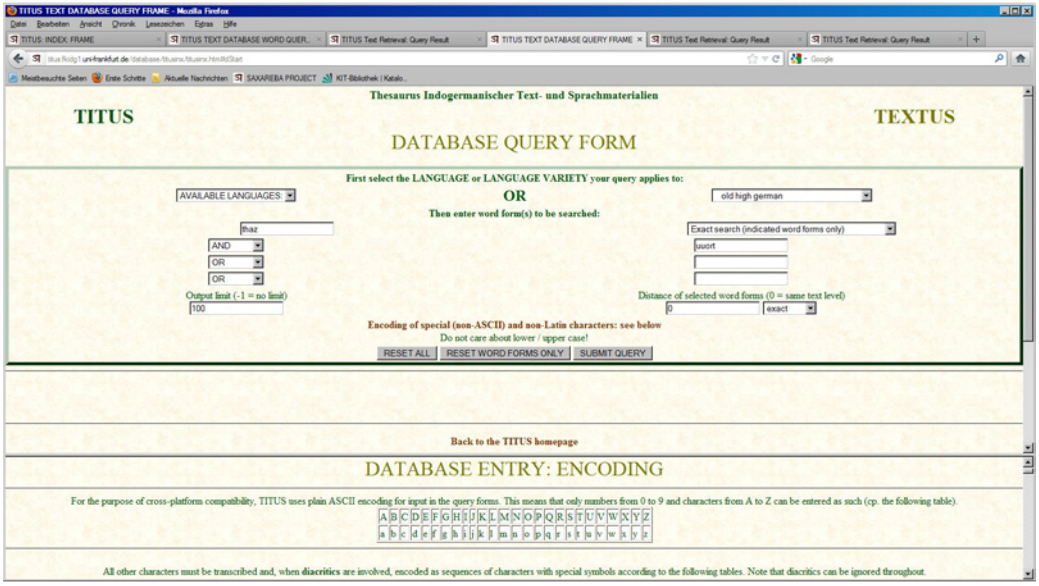
DATABASE ENTRY: ENCODING

For the purpose of cross-platform compatibility, TITUS uses plain ASCII encoding for input in the query forms. This means that only numbers from 0 to 9 and characters from A to Z can be entered as such (cp. the following table).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z

All other characters must be transcribed and, when **diacritics** are involved, encoded as sequences of characters with special symbols according to the following tables. Note that diacritics can be ignored throughout.

Figure 23


TITUS TEXT DATABASE QUERY FRAME - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: INDEX: FRAME
TITUS TEXT DATABASE WORD QUER...
TITUS Text Retrieval: Query Result
TITUS TEXT DATABASE QUERY FRAME
TITUS Text Retrieval: Query Result
TITUS Text Retrieval: Query Result
titus.fkidg1.uni-frankfurt.de/database/titusinx/titusinx.htm#dStart
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

DATABASE QUERY FORM

First select the LANGUAGE or LANGUAGE VARIETY your query applies to:

AVAILABLE LANGUAGES: **OR** old high german

Then enter word form(s) to be searched:

thaz — Exact search (indicated word forms only)

AND — uuort

OR

OR

Output limit (-1 = no limit): 100

Distance of selected word forms (0 = same text level): 0 exact

Encoding of special (non-ASCII) and non-Latin characters: see below

Do not care about lower / upper case!

RESET ALL | RESET WORD FORMS ONLY | SUBMIT QUERY

Back to the TITUS homepage

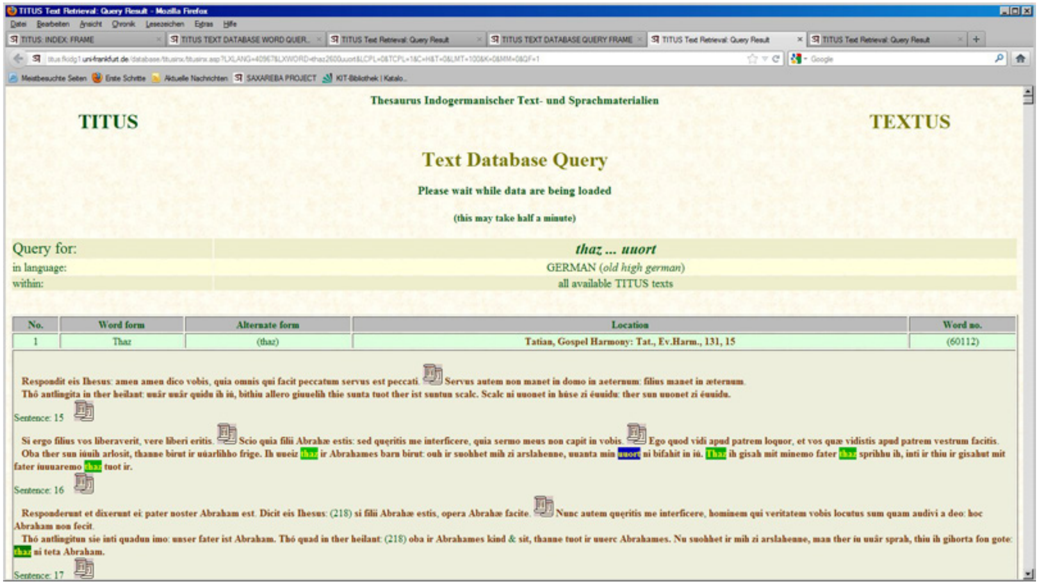
DATABASE ENTRY: ENCODING

For the purpose of cross-platform compatibility, TITUS uses plain ASCII encoding for input in the query forms. This means that only numbers from 0 to 9 and characters from A to Z can be entered as such (cp. the following table).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z

All other characters must be transcribed and, when **diacritics** are involved, encoded as sequences of characters with special symbols according to the following tables. Note that diacritics can be ignored throughout.

Figure 24


TITUS Text Retrieval: Query Result - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: INDEX FRAME
TITUS TEXT DATABASE WORD QUER...
TITUS Text Retrieval: Query Result
TITUS TEXT DATABASE QUERY FRAME
TITUS Text Retrieval: Query Result
TITUS Text Retrieval: Query Result
titus.fkidg1.uni-frankfurt.de/database/titusinx/titusinx.asp?LXLANG=40967&LXWORD=thaz2600uuort&LCPL=0&TCPL=1&C=H&T=0&LMT=100&K=0&MM=0&QF=1
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

Text Database Query

Please wait while data are being loaded

(this may take half a minute)

Query for:	*thaz ... uuort*
in language:	GERMAN (*old high german*)
within:	all available TITUS texts

No.	Word form	Alternate form	Location	Word no.
1	Thaz	(thaz)	Tatian, Gospel Harmony: Tat., Ev.Harm., 131, 15	(60112)

Respondit eis Ihesus: amen amen dico vobis, quia omnis qui facit peccatum servus est peccati. Servus autem non manet in domo in aeternum: filius manet in æternum.
Thô antlingita in ther heilant: uuâr uuâr quidu ih iû, bithiu allero giuuelih thie sunta tuot ther ist suntun scalc. Scalc ni uuonet in hûse zi êuuidu: ther sun uuonet zi êuuidu.

Sentence: 15

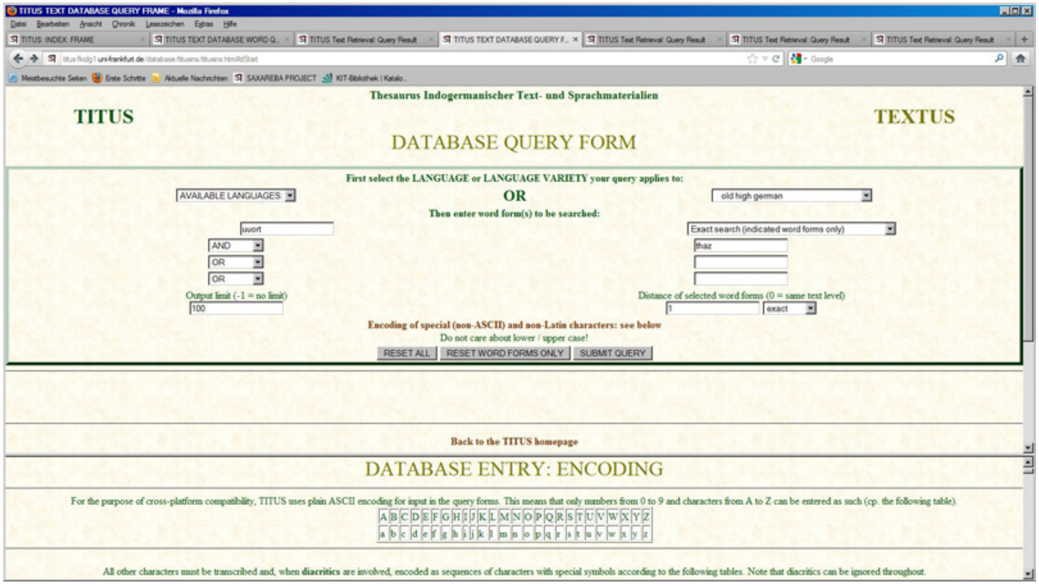
Si ergo filius vos liberaverit, vere liberi eritis. Scio quia filii Abrahæ estis: sed queritis me interficere, quia sermo meus non capit in vobis. Ego quod vidi apud patrem loquor, et vos quæ vidistis apud patrem vestrum facitis.
Oba ther sun iûuih arlosit, thanne birut ir uûarlihho frige. Ih uueiz thaz ir Abrahames barn birut: ouh ir suohhet mih zi arslahenne, uuanta min uuort ni bifahit in iû. Thaz ih gisah mit minemo fater thaz sprihhu ih, inti ir thiu ir gisahut mit fater iuuuaremo thaz tuot ir.

Sentence: 16

Responderunt et dixerunt ei: pater noster Abraham est. Dicit eis Ihesus: (218) si filii Abrahæ estis, opera Abrahæ facite. Nunc autem queritis me interficere, hominem qui veritatem vobis locutus sum quam audivi a deo: hoc Abraham non fecit.
Thô antlingitun sie inti quadun imo: unser fater ist Abraham. Thô quad in ther heilant: (218) oba ir Abrahames kind & sit, thanne tuot ir uuerc Abrahames. Nu suohhet ir mih zi arslahenne, man ther iu uuâr sprah, thiu ih gihorta fon gote: thaz ni teta Abraham.

Sentence: 17

Figure 25


TITUS TEXT DATABASE QUERY FRAME - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: INDEX FRAME
TITUS TEXT DATABASE WORD Q...
TITUS Text Retrieval: Query Result
TITUS TEXT DATABASE QUERY F...
TITUS Text Retrieval: Query Result
TITUS Text Retrieval: Query Result
TITUS Text Retrieval: Query Result
titus.fkidg1.uni-frankfurt.de/database/titusinx/titusinx.htm#Start
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...
RESET ALL
RESET WORD FORMS ONLY
SUBMIT QUERY

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

DATABASE QUERY FORM

First select the LANGUAGE or LANGUAGE VARIETY your query applies to:

AVAILABLE LANGUAGES: OR old high german

Then enter word form(s) to be searched:

uuort — Exact search (indicated word forms only)
AND — thaz
OR
OR

Output limit (-1 = no limit): 100

Distance of selected word forms (0 = same text level): 1 exact

Encoding of special (non-ASCII) and non-Latin characters: see below

Do not care about lower / upper case!

Back to the TITUS homepage

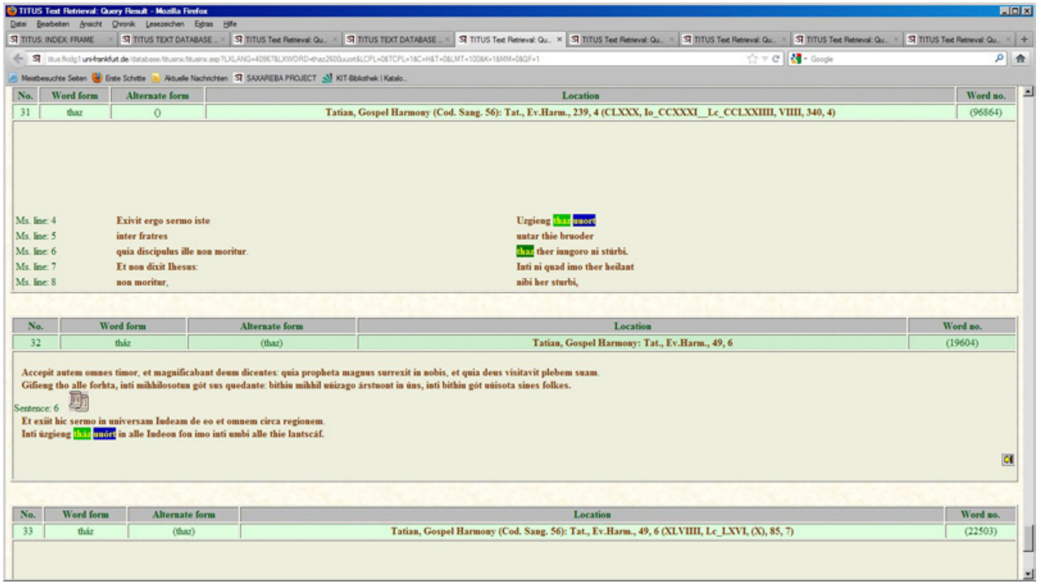
DATABASE ENTRY: ENCODING

For the purpose of cross-platform compatibility, TITUS uses plain ASCII encoding for input in the query forms. This means that only numbers from 0 to 9 and characters from A to Z can be entered as such (cp. the following table).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z

All other characters must be transcribed and, when **diacritics** are involved, encoded as sequences of characters with special symbols according to the following tables. Note that diacritics can be ignored throughout.

Figure 26



TITUS Text Retrieval: Query Result - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: INDEX FRAME
TITUS TEXT DATABASE ...
TITUS Text Retrieval: Qu...
TITUS TEXT DATABASE ...
TITUS Text Retrieval: Qu...
TITUS Text Retrieval: Qu...
TITUS Text Retrieval: Qu...
TITUS Text Retrieval: Qu...
TITUS Text Retrieval: Qu...
titus.fkidg1.uni-frankfurt.de/database/titusinx/titusinx.asp?LXLANG=4096&LXWORD=thaz2600uuort&LCPL=0&TCPL=1&C=H&T=0&LMT=100&K=1&MM=0&QF=1
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

No.	Word form	Alternate form	Location	Word no.
31	thaz	()	Tatian, Gospel Harmony (Cod. Sang. 56): Tat., Ev.Harm., 239, 4 (CLXXX, Io_CCXXXI__Lc_CCLXXIIII, VIIII, 340, 4)	(96864)

Ms. line: 4	Exivit ergo sermo iste	Uzgieng thaz uuort
Ms. line: 5	inter fratres	untar thie bruoder
Ms. line: 6	quia discipulus ille non moritur.	thaz ther iungoro ni stúrbi.
Ms. line: 7	Et non dixit Ihesus:	Inti ni quad imo ther heilant
Ms. line: 8	non moritur,	nibi her sturbi,

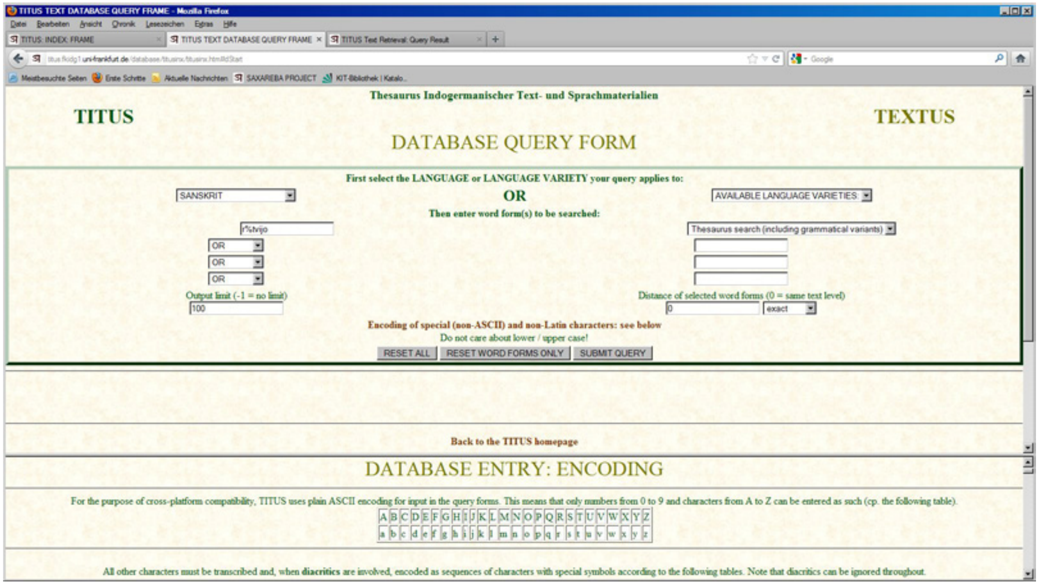
No.	Word form	Alternate form	Location	Word no.
32	tház	(thaz)	Tatian, Gospel Harmony: Tat., Ev.Harm., 49, 6	(19604)

Accepit autem omnes timor, et magnificabant deum dicentes: quia propheta magnus surrexit in nobis, et quia deus visitavit plebem suam.
Gifieng tho alle forhta, inti mihhilosotun gót sus quedante: bithiu mihhil uúizago árstuont in úns, inti bithiu gót uúisota sines folkes.

Sentence: 6
Et exiit hic sermo in universam Iudeam de eo et omnem circa regionem.
Inti úzgieng tház uuórt in alle Iudeon fon imo inti umbi alle thie lantscáf.

No.	Word form	Alternate form	Location	Word no.
33	tház	(thaz)	Tatian, Gospel Harmony (Cod. Sang. 56): Tat., Ev.Harm., 49, 6 (XLVIIII, Lc_LXVI, (X), 85, 7)	(22503)

Figure 27



TITUS TEXT DATABASE QUERY FRAME - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: INDEX FRAME
TITUS TEXT DATABASE QUERY FRAME
TITUS Text Retrieval: Query Result
titus.fkidg1.uni-frankfurt.de/database/titusinx/titusinx.htm#dStart
Google
Meistbesuchte Seiten
Erste Schritte
Aktuelle Nachrichten
SAXAREBA PROJECT
KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

DATABASE QUERY FORM

First select the LANGUAGE or LANGUAGE VARIETY your query applies to:

SANSKRIT ▾ OR AVAILABLE LANGUAGE VARIETIES: ▾

Then enter word form(s) to be searched:

r%tvijo — Thesaurus search (including grammatical variants) ▾

OR ▾

OR ▾

OR ▾

Output limit (-1 = no limit): 100

Distance of selected word forms (0 = same text level): 0 exact ▾

Encoding of special (non-ASCII) and non-Latin characters: see below

Do not care about lower / upper case!

RESET ALL | RESET WORD FORMS ONLY | SUBMIT QUERY

Back to the TITUS homepage

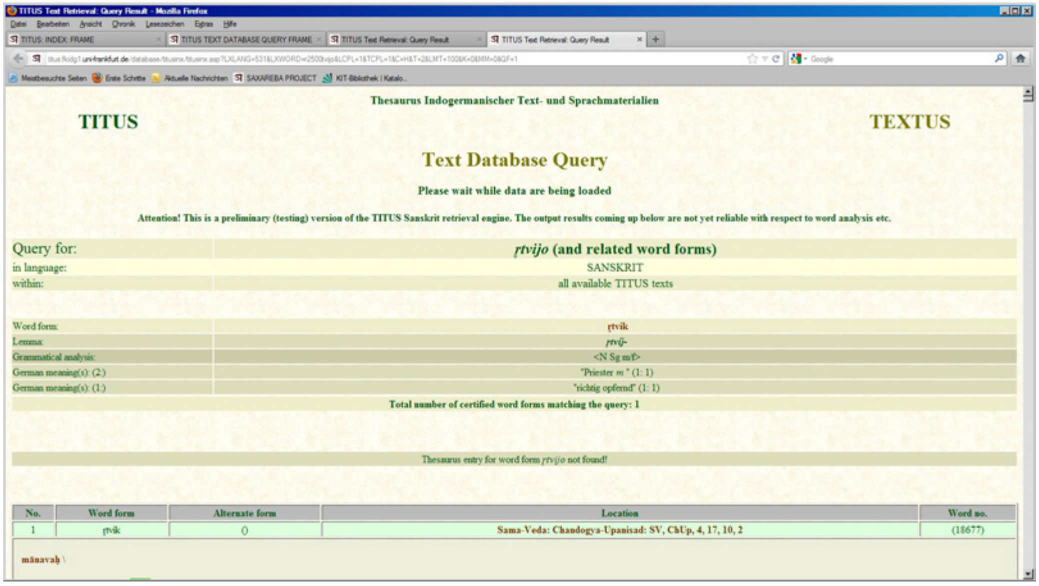
DATABASE ENTRY: ENCODING

For the purpose of cross-platform compatibility, TITUS uses plain ASCII encoding for input in the query forms. This means that only numbers from 0 to 9 and characters from A to Z can be entered as such (cp. the following table).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z

All other characters must be transcribed and, when **diacritics** are involved, encoded as sequences of characters with special symbols according to the following tables. Note that diacritics can be ignored throughout.

Figure 28


TITUS Text Retrieval: Query Result - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
TITUS: INDEX: FRAME
TITUS TEXT DATABASE QUERY FRAME
TITUS Text Retrieval: Query Result
TITUS Text Retrieval: Query Result
Meistbesuchte Seiten Erste Schritte Aktuelle Nachrichten SAXAREBA PROJECT KIT-Bibliothek | Katalo...

Thesaurus Indogermanischer Text- und Sprachmaterialien

TITUS TEXTUS

Text Database Query

Please wait while data are being loaded

Attention! This is a preliminary (testing) version of the TITUS Sanskrit retrieval engine. The output results coming up below are not yet reliable with respect to word analysis etc.

Query for:	*ṛtvijo* (and related word forms)
in language:	SANSKRIT
within:	all available TITUS texts

Word form:	ṛtvik
Lemma:	*ṛtvij-*
Grammatical analysis:	<N Sg m/f>
German meaning(s): (2:)	"Priester *m* " (1: 1)
German meaning(s): (1:)	"richtig opfernd" (1: 1)

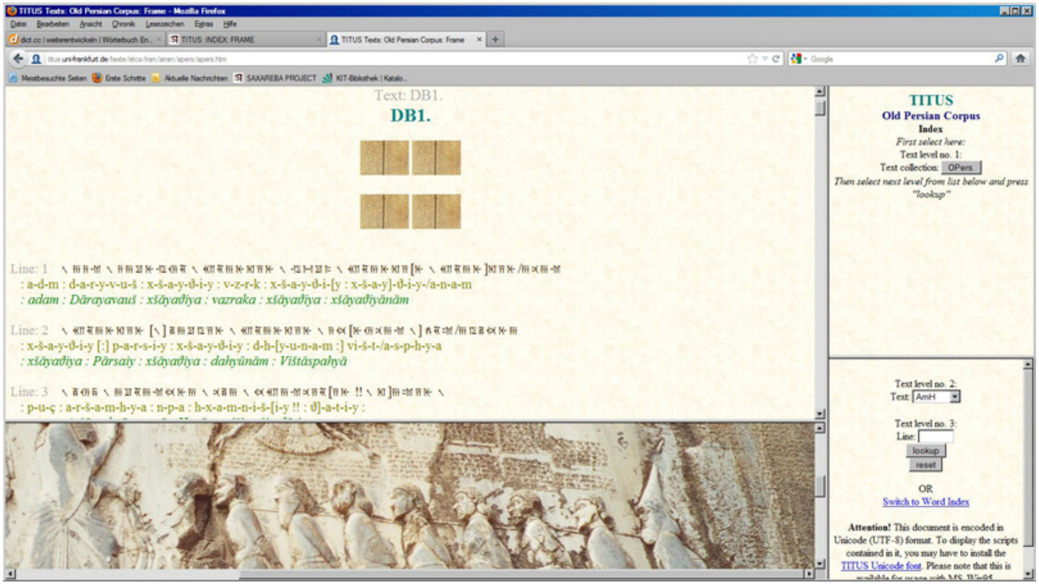
Total number of certified word forms matching the query: 1

Thesaurus entry for word form *ṛtvijo* not found!

No.	Word form	Alternate form	Location	Word no.
1	ṛtvik	()	Sama-Veda: Chandogya-Upanisad: SV, ChUp, 4, 17, 10, 2	(18677)

mānavaḥ \

Figure 29


TITUS Texts: Old Persian Corpus: Frame - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
dict.cc | weiterentwickeln | Wörterbuch En...
TITUS: INDEX: FRAME
TITUS Texts: Old Persian Corpus: Frame
Meistbesuchte Seiten Erste Schritte Aktuelle Nachrichten SAXAREBA PROJECT KIT-Bibliothek | Katalo...

Text: DB1.

DB1.

Line: 1
: a-d-m : d-a-r-y-v-u-š : x-š-a-y-ϑ-i-y : v-z-r-k : x-š-a-y-ϑ-i-[y : x-š-a-y]-ϑ-i-y-/a-n-a-m
: adam : Dārayavauš : xšāyaϑiya : vazraka : xšāyaϑiya : xšāyaϑiyānām

Line: 2
: x-š-a-y-ϑ-i-y [:] p-a-r-s-i-y : x-š-a-y-ϑ-i-y : d-h-[y-u-n-a-m :] vi-š-t-/a-s-p-h-y-a
: xšāyaϑiya : Pārsaiy : xšāyaϑiya : dahyūnām : Vištāspahyā

Line: 3
: p-u-ç : a-r-š-a-m-h-y-a : n-p-a : h-x-a-m-n-i-š-[i-y !! : ϑ]-a-t-i-y :

TITUS
Old Persian Corpus
Index
First select here:
Text level no. 1:
Text collection: OPers.
Then select next level from list below and press "lookup"

Text level no. 2:
Text: AmH
Text level no. 3:
Line:
lookup
reset
OR
Switch to Word Index

Attention! This document is encoded in Unicode (UTF-8) format. To display the scripts contained in it, you may have to install the TITUS Unicode font. Please note that this is

Figure 30