Making CTS and its application with Greek, Portuguese, and English alignments on Ducat

June, 4th Sunoikisis Digital Classics Summer 2020 Session 9



Christopher W. Blackwell cwblackwell@gmail.com
Furman University



Anise d'Orange Ferreira anise.ferreira@unesp.br UNESP/CNPq









Context

- <u>CAPES/PrInT/UNESP Program</u>: Theme: Plural Societies > Projects: Linguagens e Produção do Conhecimento > Network: Línguas para aprender, incluir, inovar e preservar: explorando teorias, métodos e práticas > Open Projects in Digital Classics
- Academic mission: Learning more about CTS- URN applications
- At Furman: CITE-app and DUCAT creating CEX files for Portuguese texts



What Cite-app is

- Cite-app is a single web-page browser app based on the CITE architecture protocol to work with texts for exploring citable resources, written in the language ScalaJs
- What kind of texts? Those with citable resources: identifiable and retrievable via machine-actionable canonical citations in URN format. (CTS-URNs)
- CITE architecture is a protocol that allows the texts to be identified and retrieved via machine-actionable canonical citations in URN format
- What kind of work? Lexical items retrieval of one or more texts



CTS - URN: canonically cited passages

- CTS URNs have 4 components separated by a colon ":".
- URN:CTS:[NAMESPACE]:[WORK identifiers]:[PASSAGE] urn:cts: means that we have a URN under a CTS protocol.
- [NAMESPACE] represents the data set. Ex. greekLit
- [WORK identifiers separated by a dot: textgroup. work. edition | translation, exemplar Ex: tlg.0553.tlg001.1st1K-grc
- [PASSAGE] specifies the text part in the document. Ex. 1.1.1 and 5.1.1
 urn:cts:greekLit:tlg0553.tlg001.1st1K-grc1:1.1.1
 urn:cts:greekLit:tlg0553.tlg001.ferreira:5.1.1



Branch: master fuCite-jupyter / unesp / cex / artemidorus-grc.cex Find file Copy path Eumaeus Adding CEX files. b93ab47 on 29 Jan 1 contributor 590 lines (586 sloc) 800 KB History Raw Blame 1 #!cexversion 3.0 3 4 #!citelibrary name#CEX library urn#urn:cite2:cex:TEMPCOLL.TEMPVERSION:TEMP_ID license#CC 3.0 NC-BY 8 9 #!ctscatalog urn#citationScheme#groupName#workTitle#versionLabel#exemplarLabel#online#lang urn:cts:greekLit:tlg0553.tlg001.1st1K-grc1:#book/section/subsection#Artemidorus#Onirocriticon#Pack, Teubner, 1963, F1k 12 #!ctsdata 14 urn:cts:greekLit:tlg0553.tlg001.1st1K-grc1:1.prooimion.head#ΑΡΤΕΜΙΔΩΡΟΣ ΚΑΣΣΙΩ ΜΑΞΙΜΩ ΧΑΙΡΕΙΝ urn:cts:greekLit:tlq0553.tlq001.1st1K-grc1:1.prooimion.1#Πολλάκις προετράπην έπὶ τὴν πραγματείαν τὴν προκειμένην ὁρμή urn:cts:greekLit:tlg0553.tlg001.1st1K-grc1:1.prooimion.2#Ιτέον δὴ λοιπὸν ἐπὶ τὸ προκείμενον, ἵνα μὴ μεῖζον ἡμῖν ἦ τὸ π urn:cts:greekLit:tlg0553.tlg001.1st1K-grc1:1.1.1#Περὶ μὲν οὖν ἐνυπνίου καὶ ὀνείρου διαφορᾶς τῆς πρὸς ἄλληλα διαίρεσις 19 urn:cts:greekLit:tlg0553.tlg001.1st1K-grc1:1.1.2#Περὶ μὲν οὖν ἐνυπνίου τοσαῦτα εἰρήσθω τὸ δὲ ὄνομα αὐτὸ κύριον, οὐχ ἦ urn:cts:greekLit:tlq0553.tlq001.1st1K-grc1:1.2.1#Ετι τῶν ὀνείρων οἰ μέν εἰσι θεωρηματικοὶ οἰ δέ ἀλληγορικοί. καὶ θεωρη urn:cts:greekLit:tlg0553.tlg001.1st1K-grc1:1.2.2#Εὼ δ' οἶμαι δεῖν καὶ τὴν αἰτίαν αὐτῶν, ὡς μάλιστα δυνατὸν ἐμοί, καθ' 21 urn:cts:greekLit:tlg0553.tlg001.1st1K-grc1:1.2.3#Ονειρός έστι κίνησις ἢ πλάσις ψυχῆς πολυσχήμων σημαντικὴ τῶν ἐσομένων urn:cts:greekLit:tlq0553.tlq001.1st1K-grc1:1.2.4#Ετι τῶν ἀλληγορικῶν ὁρίζονταί τινες πέντε εἶναι εἴδη. τοὺς μὲν γὰρ ἰδ urn:cts:greekLit:tlg0553.tlg001.1st1K-grc1:1.3.1#Ετι δὲ λέγουσιν οἱ περὶ ταῦτα δεινοὶ ὅτι δεῖ κρίνειν αἵσια τὰ φύσει κ urn:cts:greekLit:tlg0553.tlg001.1st1K-grc1:1.3.2#Ταῦτα μὲν οὖν εἰς ἀναπλήρωσιν τῶν ὑπὸ τῶν παλαιῶν ένδεῶς εἰρημένων ί urn:cts:greekLit:tlq0553.tlq001.1st1K-grc1:1.4.1#Τῶν ὁνείρων οἱ μὲν πολλὰ διὰ πολλῶν προαγορεύουσιν, οἱ δὲ ὁλίγα δι' ἀ

How does it work?

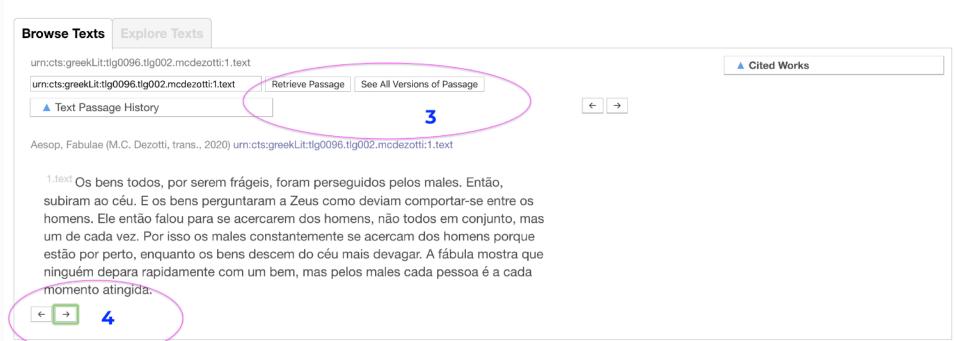
Repository: CEX library. Library URN: urn:cite2:cex:unesp_fu.v1:demo_all. License: CC 3.0 NC-BY **CITE Environment** version 1.14.0 [Online Help] Choose a local . cex file | Escolher arquivo | Nenh...nado **Browse Texts** urn:cts:greekLit:tlg0096.tlg002.First1K-grc1:1.head ▲ Cited Works urn:cts:greekLit:tlg0096.tlg002.First1K-grc1:1.head Retrieve Passage See All Versions of Passage \leftarrow \rightarrow ▲ Text Passage History \leftarrow \rightarrow

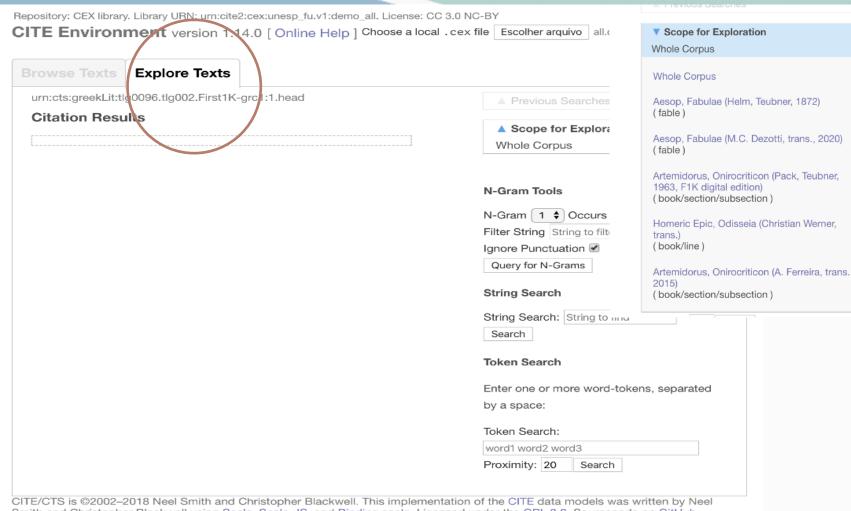
How does it work?

Repository: CEX library. Library URN: urn:cite2:cex:unesp_fu.v1:demo_all. License: CC 3.0 NC-BY

CITE Environment version 1.14.0 [Online Help]

Choose a local . cex file | Escolher arquivo | all.cex.txt





Smith and Christopher Blackwell using Scala, Scala-JS, and Binding.scala. Licensed under the GPL 3.0. Sourcecode on GitHub. Copyright and licensing information for the default library is available here. Report bugs by filing issues on GitHub.

Repository: CEX library. Library URN: urn:cite2:cex:unesp_fu.v1:demo_all. License: CC 3.0 NC-BY	
CITE Environment version 1.16.0 [Online Help]	Choose a local . cex file Escolher arquivo Nenhnado
Browse Texts Explore Texts	
urn:cts:greekLit:tlg0096.tlg002.mcdezotti:	▲ Previous Searches
N-Grams	▲ Scope for Exploration
NGram Query: n=3; threshold = 3; ignore-punctuation=true; filtered by ""; scope=urn:cts:greekLit:tlg0096.tlg002.mcdezotti:. Time: 0.038 seconds. Results: 3.	Aesop, Fabulae (M.C. Dezotti, trans., 2020)
(10) fábula mostra que (10) A fábula mostra (4) se pôs a Citation Results	N-Gram Tools N-Gram 3 • Occurs > 3 times Filter String String to filter results
Fetched 4 passages in 0.272 seconds: threshold = 3; ignore-punctuation = true; queried on 'urn:cts:greekLit:tlg0096.tlg002.mcdezotti:'.	Ignore Punctuation
2. Aesop, Fabulae (M.C. Dezotti, trans., 2020), 14 textetexto bem plausivel. E entao se pos a	
acusá-lo de que era um estorv" 3. Aesop, Fabulae (M.C. Dezotti, trans., 2020), 16.text "stado na frente do galinheiro **se pôs a**	Token Search
perguntar a elas como estavam"	Enter one or more word-tokens, separated by a space:
 Aesop, Fabulae (M.C. Dezotti, trans., 2020), 17.text "a e, tendo acertado o chifre, **se pôs a** pedir que ela não revelasse a" 	Token Search: word1 word2 word3

Proximity: 20

Search

Repository: CEX library. Library URN: urn:cite2:cex:unesp_fu.v1:demo_all. License: CC 3.0 NC-BY

CITE/CTS is ©2002–2018 Neel Smith and Christopher Blackwell. This implementation of the CITE data models was written by Neel Smith and Christopher Blackwell using Scala, Scala-JS, and Binding.scala. Licensed under the GPL 3.0. Sourcecode on GitHub. Copyright and licensing information for the default library is available here. Report bugs by filing issues on GitHub.

CITE Environment version 1.16.0 [Online Help]

Browse Texts Explore Texts

urn:cts:greekLit:tlg0096.tlg002.First1K-grc1:

N-Grams

NGram Query: n=3; threshold = 3; ignore-punctuation=true; filtered by ""; scope=urn:cts:greekLit:tlg0096.tlg002.First1K-grc1:. Time: 0.025 seconds. Results: 2.

(7) Ὁ λόγος δηλοῖ (4) λόγος δηλοῖ ὅτι

Citation Results

Fetched 7 passages in 0.262 seconds: threshold = 3; ignore-punctuation = true; queried on 'urn:cts:greekLit:tlg0096.tlg002.First1K-grc1:'.

- 1. Aesop, Fabulae (Helm, Teubner, 1872), 1.text "...ὰ βράδιον ἐξ οὐρανοῦ κάτεισι. **'Ο λόγος δηλοῖ**, ὅτι ἀγαθῷ μὲν οὐδεὶς ταχέως ..."
- Aesop, Fabulae (Helm, Teubner, 1872), 5.text "...ετοῦ πάντας αὐτοὺς κατέφαγεν. **'Ο λόγος δηλοῖ**, ὅτι οἱ φιλίαν παρασπονδοῦντε..."
 Aesop, Fabulae (Helm, Teubner, 1872), 7.text "...ι, τοὺς ἀετούς μὴ νεοττεύειν. **'Ο λόγος δηλοῖ**,
- υηδενὸς καταφρονεῖν, λογιζομ..."

 4 Δesop, Fabulae (Helm, Teubner, 1872), 9 text ", ἀ μηδέπω φαινόμενα διώκουμ " **!Ο λόνος δηλοῖ**
- 4. Aesop, Fabulae (Helm, Teubner, 1872), 9.text "...ὰ μηδέπω φαινόμενα διώκοιμι." **'Ο λόγος δηλοῖ**, ώς οὔτω καὶ τῶν ἀνθρώπων ἀλό..."
- 5. Aesop, Fabulae (Helm, Teubner, 1872), 11.text "...ὶ Διονυσίοις ἐρίφους τέξεται. **'Ο λόγος δηλοῖ**, ὅτι πολλοὶ διὰ τὸ ἴδιον κέρδ..."
- 6. Aesop, Fabulae (Helm, Teubner, 1872), 12.text "...ν, ἐκείνας ἡμῶν προκρινεῖς." **'Ο λόγος δηλοῖ** μὴ δεῖν τούτων προσεταιρίζεσθ..."
- 7. Aesop, Fabulae (Helm, Teubner, 1872), 15.text "...λαξ γένη, οὐ προσελευσόμεθα." **'Ο λόγος δηλοῖ**, ὅτι οἱ Φρόνιμοι τῶν ἀνθρώπων..."

▲ Previous Searches

▲ Scope for Exploration

Aesop, Fabulae (Helm, Teubner, 1872)

N-Gram Tools

N-Gram 3 ♦ Occurs > 3 times

Filter String String to filter results

Ignore Punctuation

Query for N-Grams

String Search

String Search: se pos a Search

Token Search

Enter one or more word-tokens, separated by a space:

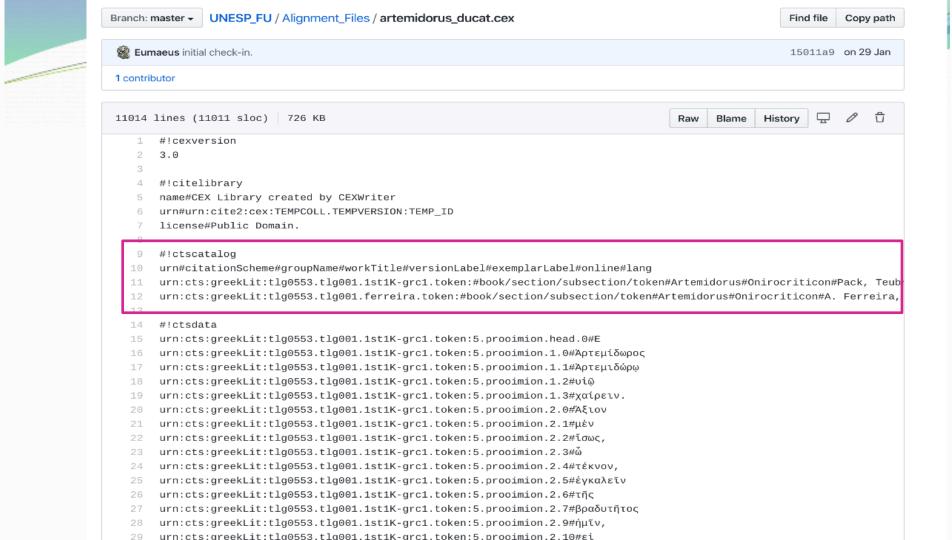
Token Search: word1 word2 word3

Proximity: 20 Search

What DUCAT is

- Daughter-of-Ugarit Citation Alignment Tool.
- An alignment tool in a single web-page app based on the CITE architecture protocol and based on Ugarit alignment tool (up to 3 texts).
- Texts and translations with citable resources identifiable and retrievable via machine-actional canonical citations in URN format. (CTS-URNs)
- It is able to load multiple translations
- The words also are identified

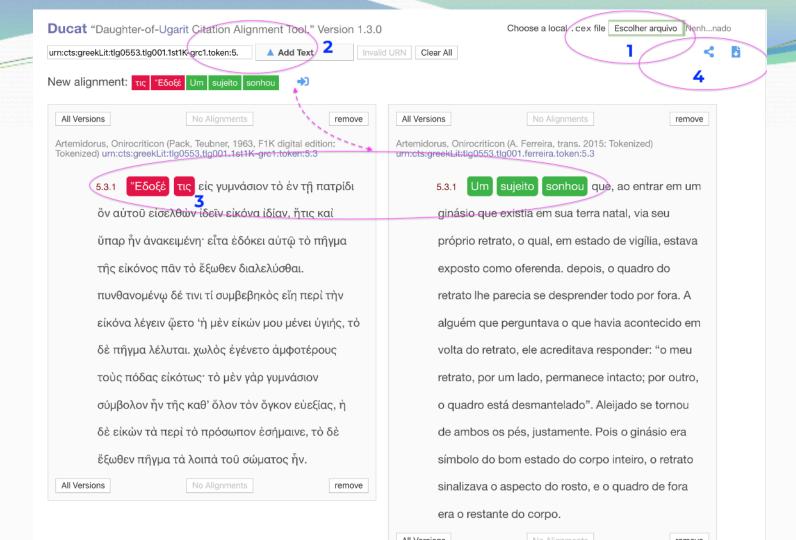




```
#!citelibrary
    name#CEX Library created by CEXWriter
    urn#urn:cite2:cex:TEMPCOLL.TEMPVERSION:TEMP_ID
    license#Public Domain.
    #!ctscatalog
    urn#citationScheme#groupName#workTitle#versionLabel#exemplarLabel#online#lang
    urn:cts:greekLit:tlg0553.tlg001.ferreira.token:#book/section/subsection/token#Artemidorus#Onirocriticon#A. Ferreira,
    urn:cts:greekLit:tlg0553.tlg001.1st1K-grc1.token:#book/section/subsection/token#Artemidorus#Onirocriticon#Pack, Teubl
14
    #!ctsdata
    urn:cts:greekLit:tlg0553.tlg001.ferreira.token:5.prooimion.head.0#5
16
    urn:cts:greekLit:tlg0553.tlg001.ferreira.token:5.prooimion.1.0#Artemidoro
17
    urn:cts:greekLit:tlg0553.tlg001.ferreira.token:5.prooimion.1.1#saúda
    urn:cts:greekLit:tlg0553.tlg001.ferreira.token:5.prooimion.1.2#o
    urn:cts:greekLit:tlg0553.tlg001.ferreira.token:5.prooimion.1.3#seu
    urn:cts:greekLit:tlg0553.tlg001.ferreira.token:5.prooimion.1.4#filho
    urn:cts:greekLit:tlq0553.tlq001.ferreira.token:5.prooimion.1.5#Artemidoro
    urn:cts:greekLit:tlg0553.tlg001.ferreira.token:5.prooimion.2.0#Talvez,
    urn:cts:greekLit:tlg0553.tlg001.ferreira.token:5.prooimion.2.1#filho,
    urn:cts:greekLit:tlg0553.tlg001.ferreira.token:5.prooimion.2.2#fosse
    urn:cts:greekLit:tlg0553.tlg001.ferreira.token:5.prooimion.2.3#justo
    urn:cts:greekLit:tlg0553.tlg001.ferreira.token:5.prooimion.2.4#nos
    urn:cts:greekLit:tlg0553.tlg001.ferreira.token:5.prooimion.2.5#acusar
    urn:cts:greekLit:tlg0553.tlg001.ferreira.token:5.prooimion.2.6#de
    urn:cts:greekLit:tlg0553.tlg001.ferreira.token:5.prooimion.2.7#morosidade,
```

#!cexversion

3.0



urn:cts:greekLit:tlg0553.tlg001.ferreira.token:5.1

Add Text

Load Passage from URN

Clear All



Saved Alignments: 0 1 2 3 4 5

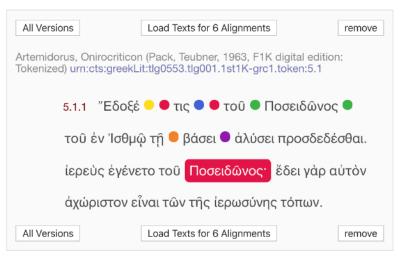






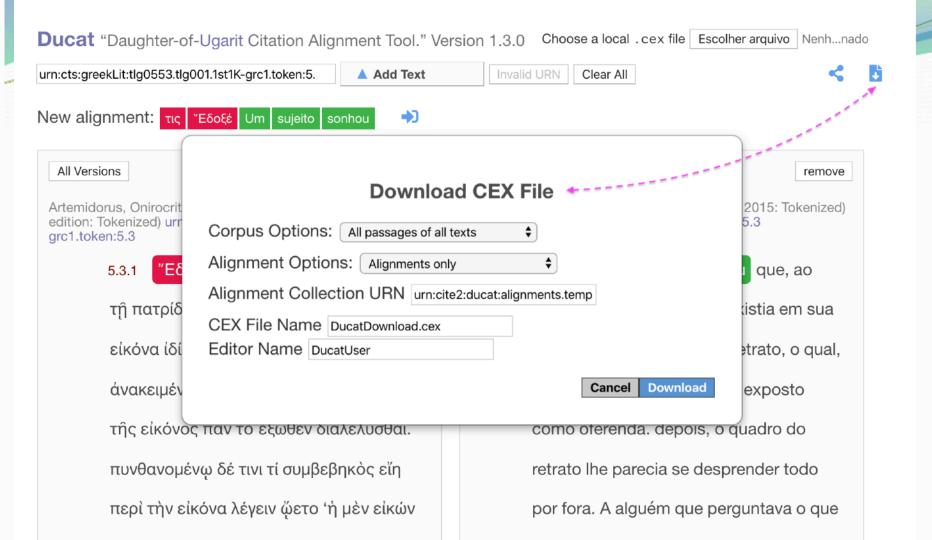


New alignment: Ποσειδώνος Poseidon:









FURMAN-UNESP page

https://furman-editions-in-progress.github.io/UNESP_FU/



A future plan for the DDGP (Dig.Dict.Grc-Por)

- Transform the DDGP corpus in a text with citable resources: identifiable and retrievable via machine-actional canonical citations in URN format
- 1) entries linked to other dictionaries
- 2) proper names (authors, works, places, person names)
- Exchange of visits, students, and training for our programmer

