

Citation Mining of Humanities Journals

A comparison of the projects
Cited Loci and *Linked Books*

Giovanni Colavizza & Matteo Romanello

TOC

1. Citation indexes
2. Citation mining
3. The humanities
4. Cited Loci and Linked Books
5. The citation mining framework
6. ...

Citation Indexes



Articles (include patents) Case law

Stand on the shoulders of giants

Citing publication



Cited
source



Dopo la ritirata di Pipino l'imperatore di Costantinopoli inviò ad Aquisgrana, presso Carlo Magno, il legato Arsafio, il quale nell'812 stipulò un trattato col sovrano franco. In esso fu stabilito il riconoscimento a Carlo della dignità imperiale (ma non di imperatore dei "romani") da parte del *basileus*, in cam-

3. GIOVANNI DIACONO, *Istoria Veneticorum*, a cura di L. A. Berto in *Cronache*, a cura di L. A. Berto e G. Fedalto, in *Corpus Scriptorum Ecclesiae Aquileiensis*, XII/2, Roma 2003, pp. 72-74.

4. Ivi, p. 74. G. OSTROGORSKY, *Storia dell'Impero bizantino*, Torino 1968, p. 177; A. CARILE, "La presenza bizantina dell'alto Adriatico tra il VII e il IX secolo" in *AAA*, XXVII (1983), pp. 107-129. Negli *Annales*, Eginardo attribuisce, invece, la vittoria ai franchi.

Grado originata dall'invasione longobarda (568), perché l'interrogativo posto da Massenzio riguardava direttamente la predicazione dell'evangelista Marco nella città di Aquileia⁵. Infatti, la risposta del rappresentante del patriarca di Grado Venerio fu dimessa e rassegnata: "*Negare non possumus et omnes sciunt Aquileiam civitatem primam et metropolim a beato Marco*

5. Sulla questione marciana cf. G. FEDALTO, *Aquileia. Una chiesa due patriarcati*, Roma 1999, pp. 44-54; ID., *San Marco tra Babilonia, Roma e Aquileia, nuove ipotesi e ricerche*, in *San Marco: aspetti storici e agiografici*, Atti a cura di A. Niero, Venezia 1996, pp. 35-50; G. BRASUTTI, *La tradizione marciana aquileiese*, Udine 1990; G. C. MENIS, *La XII lettera attribuita a sant'Ambrogio e la questione marciana aquileiese*, in *Rivista di storia della Chiesa in Italia*, XVIII (1964), pp. 243-253.

The Humanities

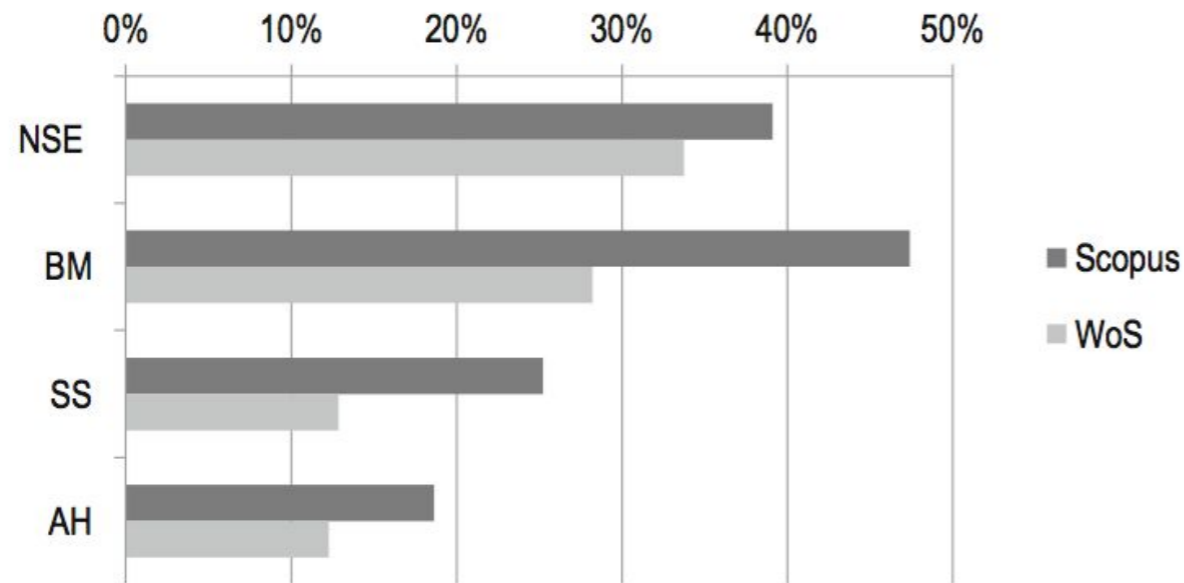


Fig. 1 Proportion of Ulrich academic journals indexed in Web of Science and Scopus

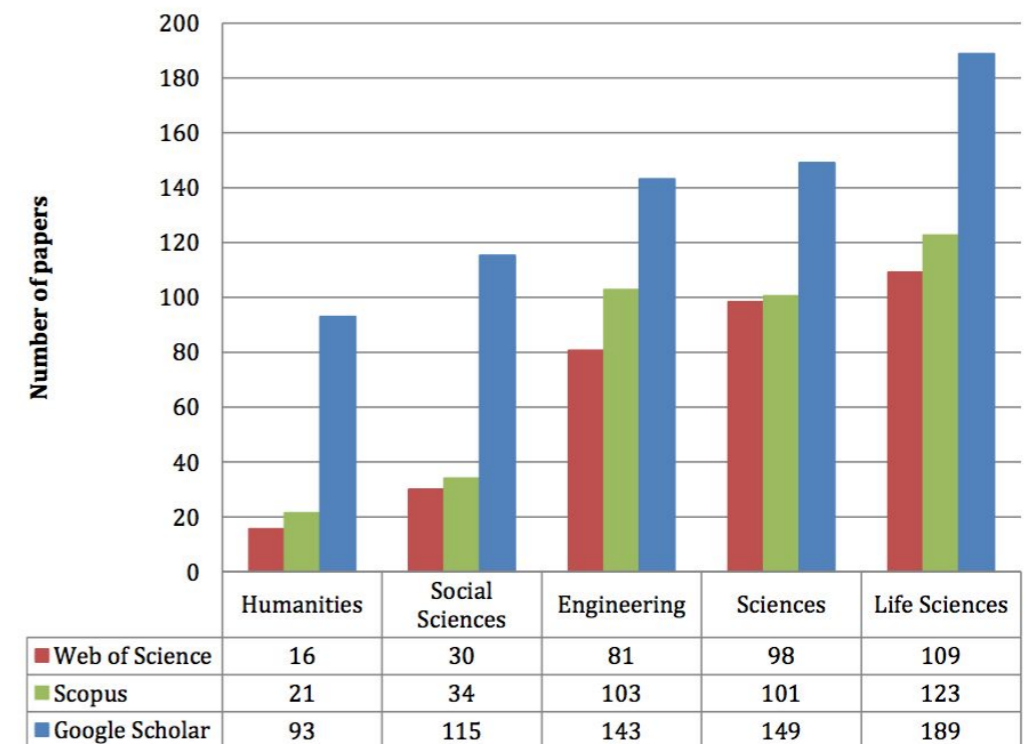


Fig. 3 Average number of papers per academic across five disciplines and three databases, July 2015

Harzing, Anne-Wil, and Satu Alakangas. 2016. "Google Scholar, Scopus and the Web of Science: A Longitudinal and Cross-Disciplinary Comparison." *Scientometrics* 106 (2): 787–804.

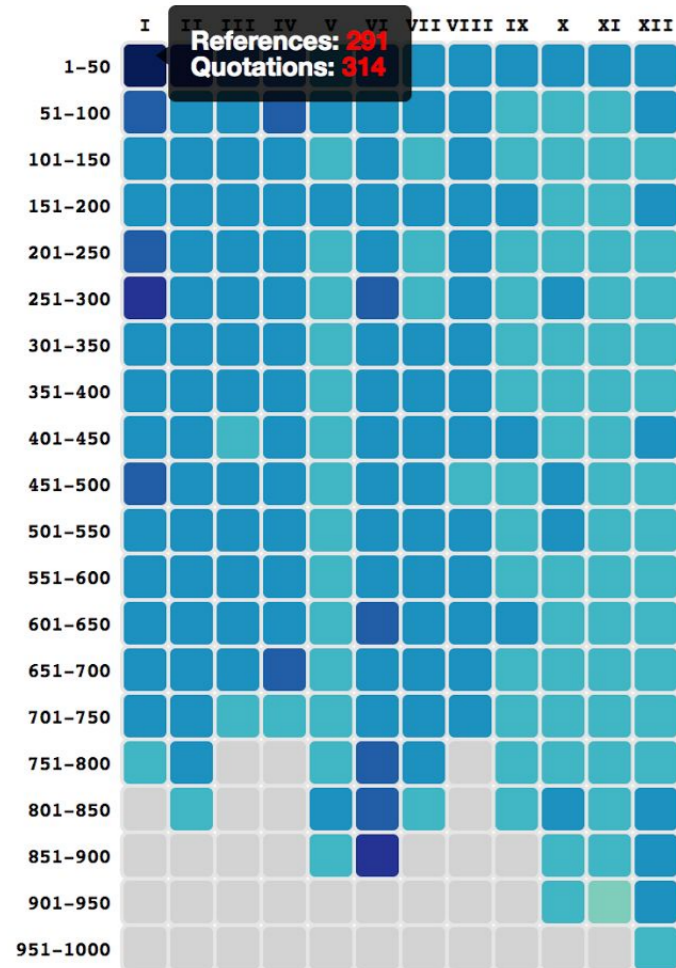
doi:10.1007/s11192-015-1798-9.

Mongeon, Philippe, and Adèle Paul-Hus. 2016. "The Journal Coverage of Web of Science and Scopus: A Comparative Analysis." *Scientometrics* 106 (1): 213–28.

doi:10.1007/s11192-015-1765-5.

Cited Loci

Display:



In Focus: Book 1, lines 1-50

Arma virumque cano, Troiae qui primus ab oris
Italiam, fato profugus, Laviniaque venit
litora, multum ille et terris iactatus et alto
vi superum saevae memorem Iunonis ob iram;
5 multa quoque et bello passus, dum conderet urbem,
inferretque deos Latio, genus unde Latinum,
Albanique patres, atque altae moenia Romae.
Musa, mihi causas memora, quo numine laeso,
quidve dolens, regina deum tot volvere casus
10 insignem pietate virum, tot adire labores
impulerit. Tantaene animis caelestibus irae?
Urbs antiqua fuit, Tyrii tenuere coloni,
Karthago, Italiam contra Tiberinaque longe
ostia, dives opum studiisque asperrima belli;
15 quam Iuno fertur terris magis omnibus unam
posthabita coluisse Samo; hic illius arma,
hic currus fuit; hoc regnum dea gentibus esse,
si qua fata sinant. iam tum tenditque fovetae.

Results: quotations references

Aen. 1.1

112 52

Menin Virumque: Translating Homer with Virgil in "Epigrammata Bobiensia" 46_47 and 64

SCOTT MCGILL
The Classical Journal 2006
DOI: [10.2307/30038019](https://doi.org/10.2307/30038019)

By rendering Iliad 1.1 with Aeneid 1.1 rather than translating Homer with an eye to word-for-word correspondence, the EB poets offer a striking perspective on that aspect of ancient education.

— (reference: 571a5798b634c1c68fef6343)

Silius Punica 6. 1-53

EDWARD L. BASSETT
Classical Philology 1959
DOI: [10.2307/266203](https://doi.org/10.2307/266203)

(The Silian phrase will help to endorse foeda over

Linked Books

Venice Scholar

Frederic Lane

giovanni Statistics About

6 Authors 32 Books 3 Articles 0 Contributions 1 Primary sources

Lane, Frederic Chaoin VIAF

Publication profile Details of cited by and citing

Books - 12 Articles - 1

Enable year filter

1676 2004

1676 2004

Language

Lane, Frederic Chaoin
Found occurrences in DB: 132
Citing: 113
Cited by: 23

Maitrejean, Frederic
Found occurrences in DB: 1
Citing: 1
Cited by: 0

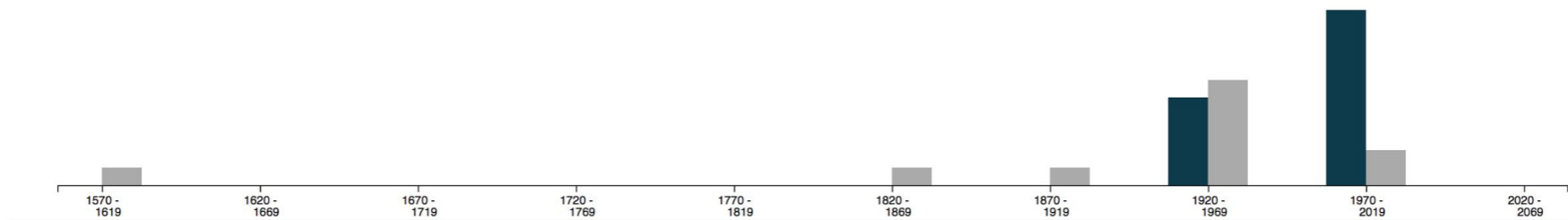
Venice Scholar

Frederic Lane

giovanni Statistics About

Lane, Frederic Chaoin VIAF

Publications (13) Citations (N° of references: 136)



Citing (N° of references: 113)

Cited by (N° of references: 23)

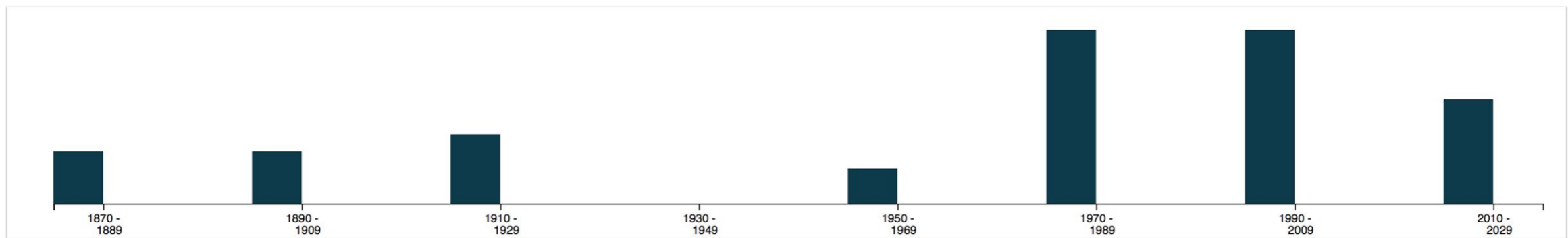
Books - 33

Title	Links	Author	Publisher	Year
Dalla galea alla nave di linea : le trasformazioni della marina veneziana (1572-1699)	Show reference (1) V Scholar Catalog	Candiani, Guido	Città del Silenzio,	2012
Venice's Mediterranean colonies : architecture and urbanism	Show references (2) V Scholar Catalog	Geörgopoulou, Maria 1961-	Cambridge university press,	2001
Glassmaking in Renaissance Venice : the fragile craft	Show reference (1) V Scholar Catalog	McCray, Patrick	Ashgate,	1999
Glassmaking in Renaissance Venice : the fragile craft	Show reference (1) V Scholar Catalog	McCray, Patrick	Ashgate,	1999
Money and banking in medieval and Renaissance Venice	Show reference (1) V Scholar Catalog	Mueller, Reinhold Christopher	Johns Hopkins University pres...	1997

Venice Scholar giovanni Statistics About

Savi all'eresia (Santo Ufficio), 1541 - 1797

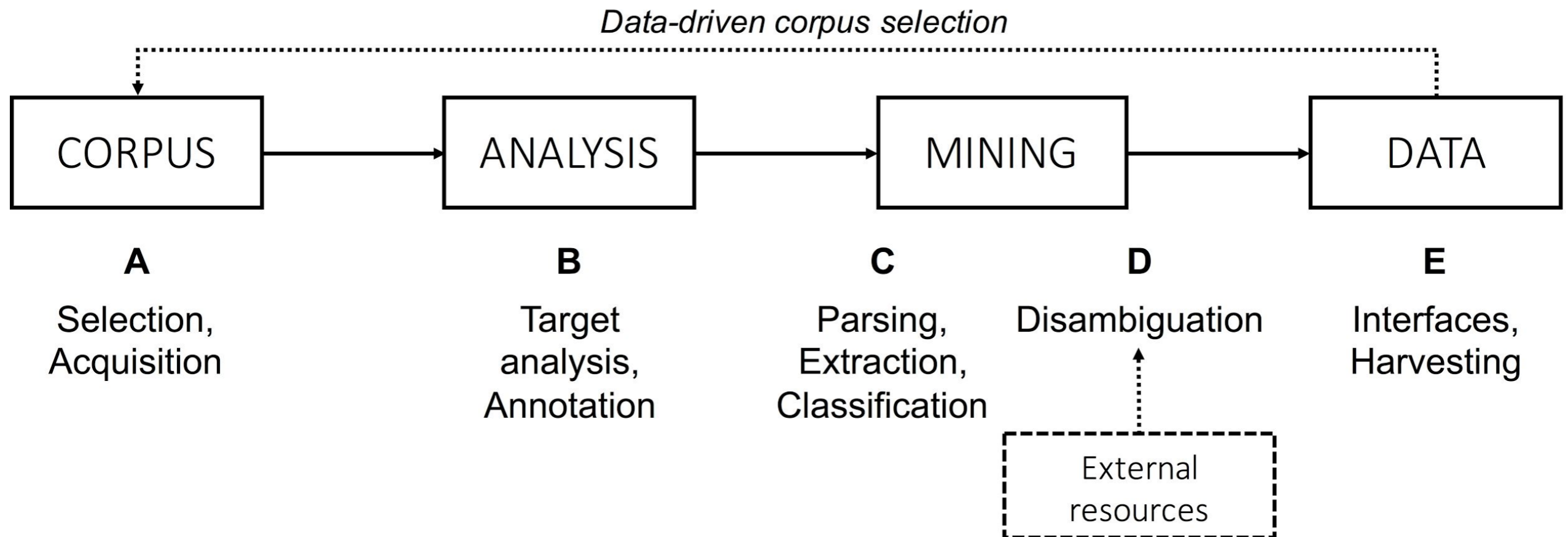
Citing
(N° of references:74)



Articles - 66

Title	Links	Journal	Issue n°	Year
Una guaritrice processata dal Tribunale del Sant'Ufficio di Venezia nel Seicento per eresia ed esorc...	Show references (3)	V Scholar Catalog	Studi veneziani	2013
Donne imprenditrici nella Terraferma Veneta della prima età moderna (secoli XV-XVI)	Show reference (1)	V Scholar Catalog	Archivio veneto	2012
Il terzo visitatore nella biblioteca di Marin Sanudo il Giovane e nelle sue camere	Show reference (1)	V Scholar Catalog	Studi veneziani	2011

The Citation Mining Framework



A. Corpus Selection & Acquisition

	Linked Books	Cited Loci
<i>Corpus' boundaries</i>	Open corpus	JSTOR = corpus
<i>Publication types</i>	Monographs, journal articles	Journal articles
<i>Corpus size</i>	~2k monographs ~5.5k articles	~130k articles
<i>Metadata source</i>	Opac SBN, later enriched (e.g. with article-level metadata)	JSTOR API
<i>Data</i>	Digital images (.tiff, .jpg) + OCR	Plain text
<i>Legal framework</i>	Holding libraries keep copyright on materials	License agreement with JSTOR, DfR scheme

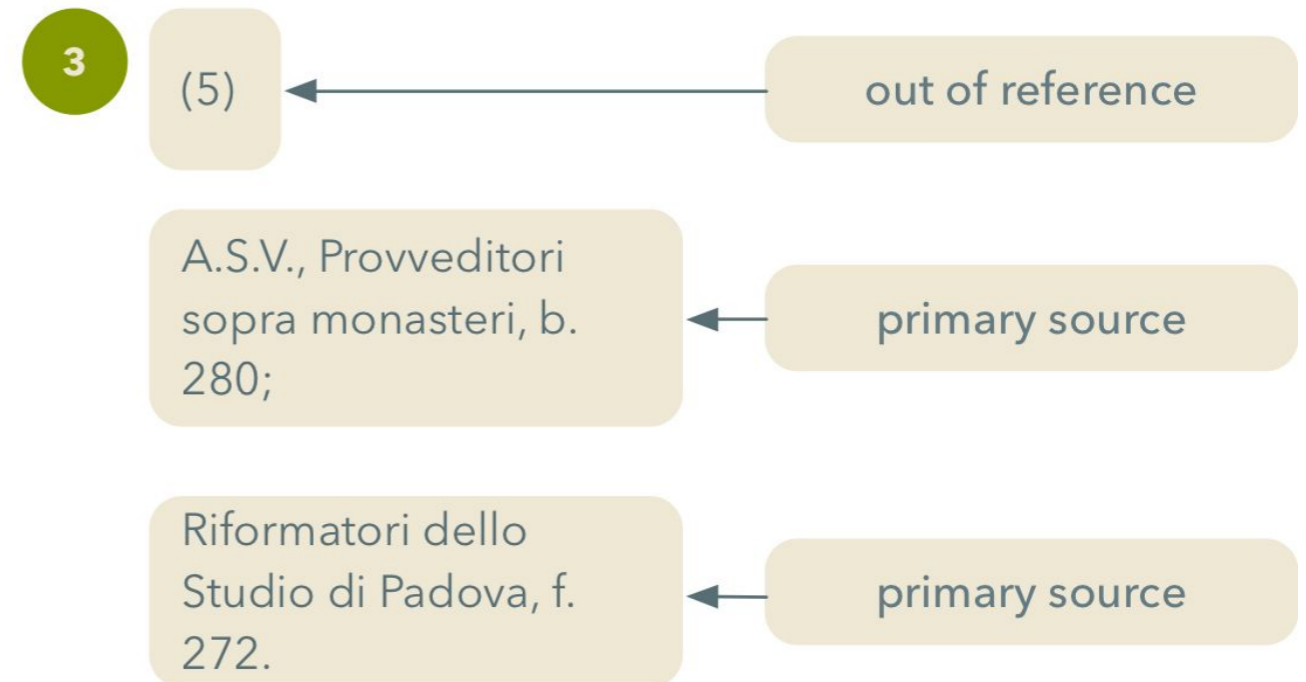
B. Target Analysis and Annotation

	Linked Books	Cited Loci
<i>Types of cited sources</i>	Primary, secondary (books, articles)	Primary (canonical texts)
<i>Ad-hoc annotation campaign</i>	Yes, 40k annotations	No
<i>Reference style classification</i>	Yes	No

- Archive of Venice, *Procuratori di San Marco, de citra, commissarie*, b. 1, c. 7
- ASVe, PSM, *de citra, commissarie*, b. 1, c. 7.
- Ivi, c. 8.
- canonical texts: e.g. Virgil, *Aen.* 12, 101-109, Hom. *Il.* 7.180
- fragmentary texts: e.g. FGrHist 688 F 13; Alc. fr. 34,1 Voigt
- inscriptions (e.g. CIL 3, 6174; AE 1991, 1405)
- papyri (e.g. PCair. inv. 10750)
- manuscripts (e.g. Vendôme, Bibl. mun. 31)
- museum objects (e.g. coins, vases, etc.)

C. Parsing, Extraction, Classification (LB)

1 (5) A.S.V., Provveditori sopra monasteri, b. 280; Riformatori dello Studio di Padova, f. 272.



C. Parsing, Extraction, Classification (CL)

1
 monks to bees occurs in Aldhelm of Malmesbury 's « De uirginitate ».
 s in their voluntary solidarity and obedience to leadership.
 vn from other Christian and pagan literature.

y influenced by REFAUWORK REFSCOPE REFSCOPE REFAUWORK REFSCOPE
 Pliny, nat. 11, 4, 11 and 11, 16, 46 and Vergil, georg. 4, 149-218.

2
 s to bees occurs in Aldhelm of Malmesbury 's « De uirginitate ».
 heir voluntary solidarity and obedience to leadership.
 om other Christian and pagan literature.

enced by REFAUWORK REFSCOPE REFSCOPE REFAUWORK REFSCOPE
 Pliny, nat. 11, 4, 11 and 11, 16, 46 and Vergil, georg. 4, 149-218.

Scope → Scope

3
 nks to bees occurs in Aldhelm
 n their voluntary solidarity and
 from other Christian and paga

REFAUWORK → Scope → REFSCOPE ID:R1
 "Pliny, nat." → "11, 4, 11"
 Note: urn:cts:latinLit:phi0978.phi001:11.4.11

fluenced by REFAUWORK REFSCOPE REFSCOPE REFAUWORK REFSCOPE
 Pliny, nat. 11, 4, 11 and 11, 16, 46 and Vergil, georg. 4, 149-218.

Scope → Scope



D. Disambiguation of References (LB)

Author mentions:

1. Lookup on internal DB
2. Search in VIAF (API)
3. Otherwise, add new record to DB

Secondary references:

1. Lookup on internal DB
2. Lookup on OPAC ISBN
3. Otherwise, add new record to DB

Primary references:

1. Lookup on internal DB
2. SiASVe lookup
 - ML-based classifier
 - If low confidence, roll back to Rule Based classifier

D. Disambiguation of References (CL)

1 Citation Extractor

Lorem ipsum Verg. Aen. 12.10 f. sit amet, consectetur adipiscing elit. Duis facilisis lacinia ipsum id pharetra.

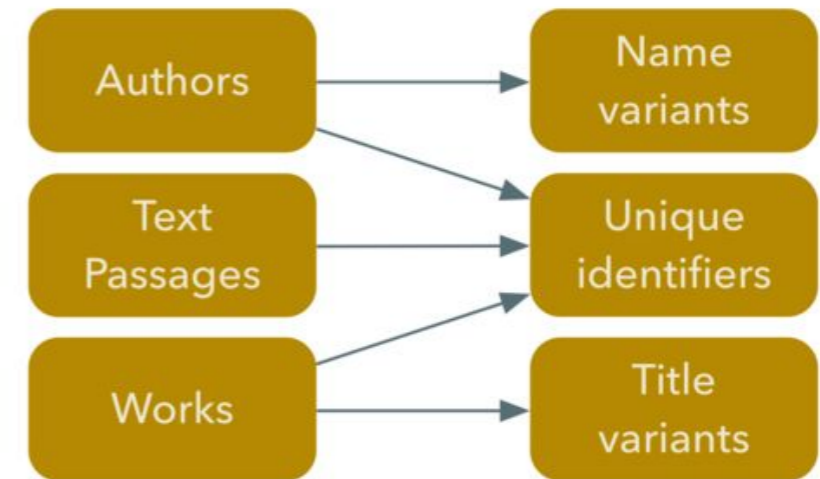
Verg. Aen. 12.10 f.

Verg. Aen

+

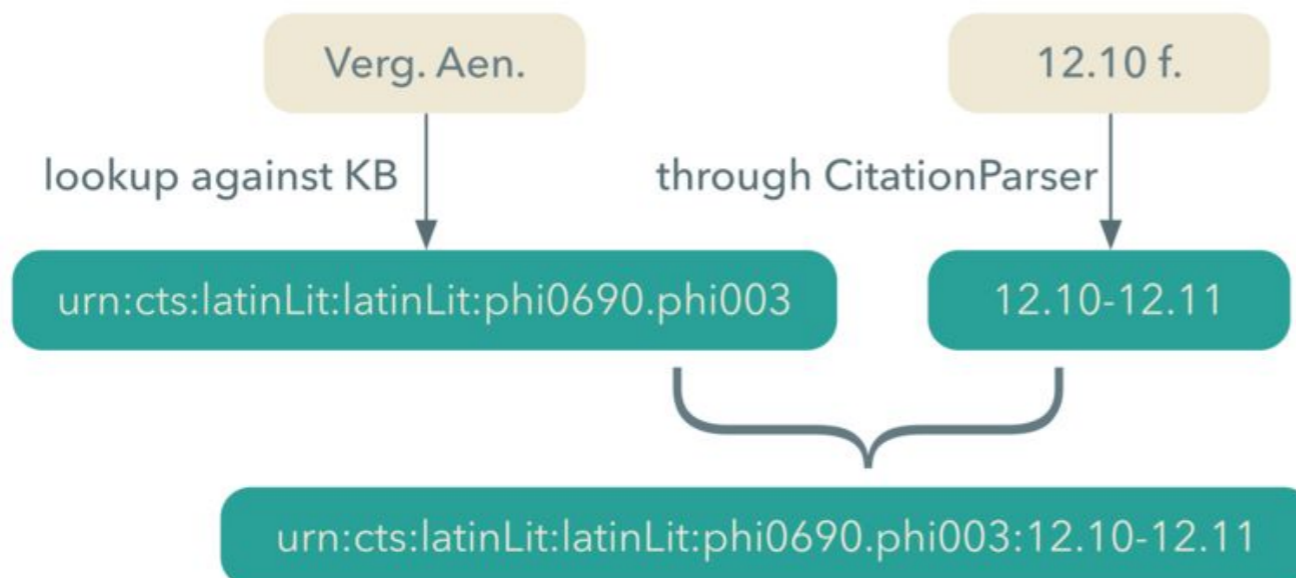
12.10 f.

3 Knowledge Base

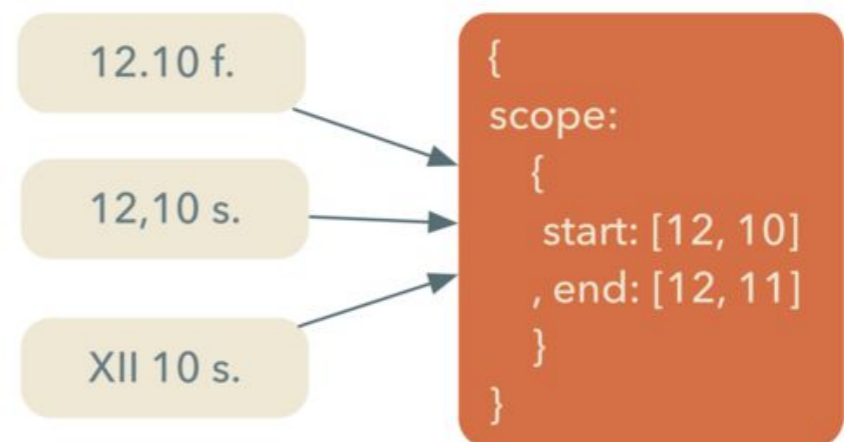


(1,500 authors; 5,500 works)

2 Citation Matcher



4 Citation Parser



(Context-Free Grammar)

E. Publication & Project Outputs

	Linked Books	Cited Loci
<i>Code</i>	Yes (partly)	Yes
<i>Annotation Data</i>	Yes	No
<i>Citation dataset</i>	Yes (in progress)	Yes (in progress)
<i>API</i>	Yes	Yes (in progress)

E. Publication & Project Outputs (LB)

Linked Books

131 documents

0 Books 0 Contributions 0 Journals 0 Articles

Page 1 / 11 >

Language
Provenance

Article Rassegna di letteratura italiana: Rinascimento PETRONIO, GIUSEPPE 1934 Volume : 117 Ateneo veneto	Article Alle origini del teatro veneto del Rinascimento: l'es... Zorzi, Ludovico 1964 Volume : 2 Ateneo veneto	Article Un grande artista del Rinascimento giudicato da al... Meneghin, Vittorino 1970 Volume : 8 Ateneo veneto
Article Il museo privato nel Rinascimento: da Albergo dell... Franzoni, Lanfranco 1984 Volume : 22 Ateneo veneto	Article Il trionfo della voluttà fra Umanesimo e Rinascime... Marchini, Nelli-Elena Vanzan 1994 Volume : 32 Ateneo veneto	Article La storia della città. Venezia e il Rinascimento neg... Calabi, Donatella 2013 Volume : 12 Ateneo veneto
Article Agricoltura e alchimia nel Rinascimento. Un'introd... Rugolo, Ruggero 1994 Volume : 27 Studi veneziani	Article Cultura e percorsi di Silvio Belli, «ingegnere» del ... Miraglia, Alessandra 2001 Volume : 42 Studi veneziani	Article Cultura, scuole, università in Venezia tra Medioev... Ortalli, Gherardo 2006 Volume : 52 Studi veneziani

Linked Books Catalog

E. Publication & Project Outputs (LB)

Venice Scholar

Publication year

Enable year filter

Language

Provenance

matteo

Statistics

About

5 Authors

23 Books

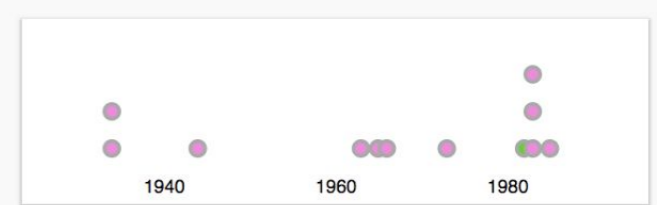
3 Articles

0 Contributions

1 Primary sources

Lane, Frederic Chaoin VIAF

Publication profile Details of cited by and citing



Articles - 1

Wages and Recruitment of Venet..., Studi veneziani, 1982

Books - 11

Venice : a maritime republic, Johns Hopkins Unive..., 1973

Venetian ship and shipbuilders..., The Johns Hopkins p..., 1934

Venise : une république mariti..., Flammarion,, 1985

Le navi di Venezia : fra i sec..., Einaudi,, 1983

Venetian ships and shipbuilder..., The Johns Hopkins p..., 1934

Navires et constructeurs a Ven..., S.E.V.P.E.N.,, 1965

Andrea Barbarigo : merchant of..., The Johns Hopkins p..., 1944

Venice and history : the colle..., Johns Hopkins Press..., 1966

Le navi di Venezia fra i secol..., G. Einaudi,, 1983

Le navi di Venezia fra i secol..., Einaudi,, 1983

Il debito pubblico della Repub..., Ist. editoriale cis..., 1963

Lane, Frederic Chaoin

Found occurrences in DB: 65
Citing: 51
Cited by: 12

Lane, Frederic Chaoin - Venice and history : the collected papers of Frederic C. Lane

Baltimore : Johns Hopkins Press,, 1966
Citing: 2
Cited by: 0

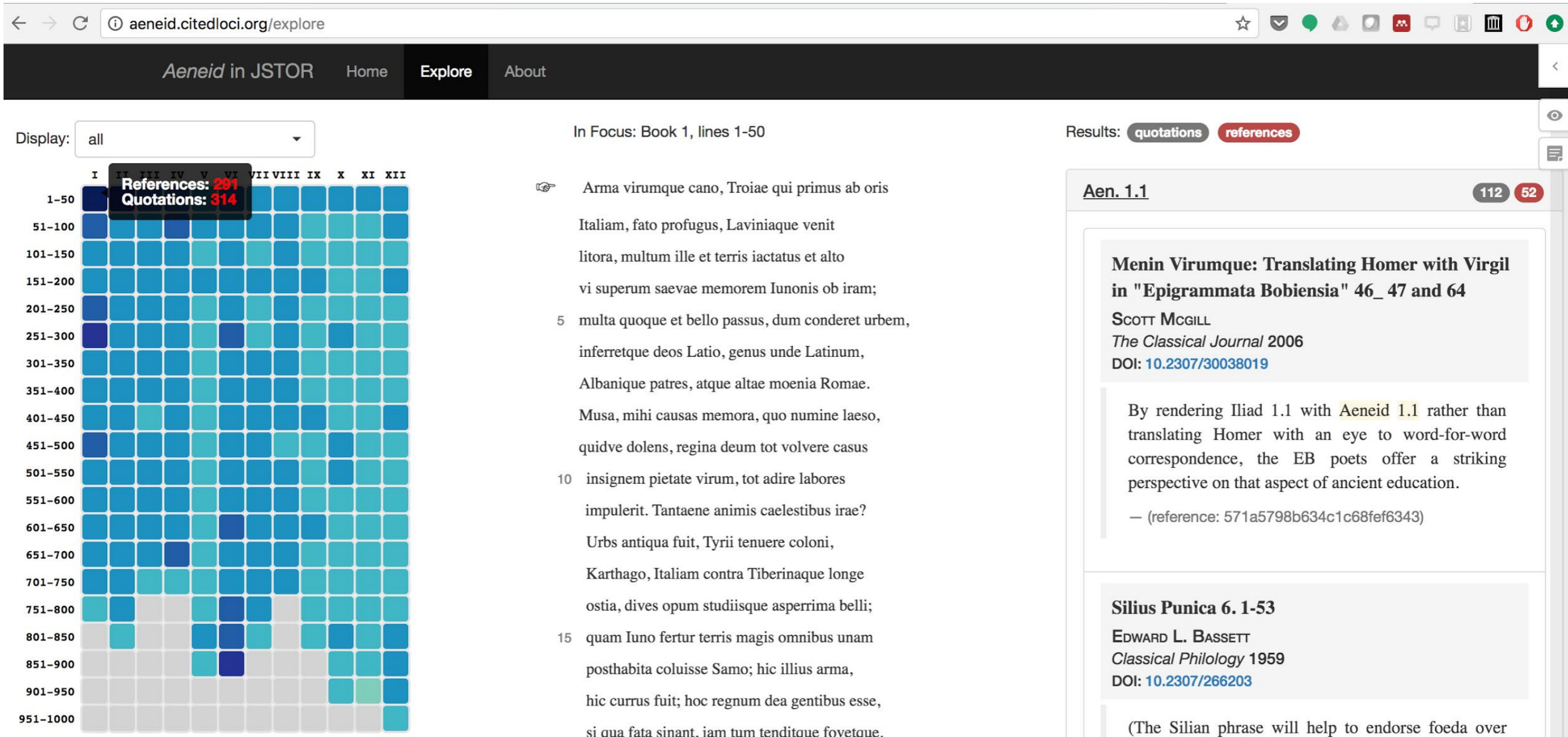
Maitrejean, Frederic

Found occurrences in DB: 1
Citing: 1
Cited by: 0

Church, Frederic Corss, 1885-1966

Found occurrences in DB: 3
Citing: 3
Cited by: 0

E. Publication & Project Outputs (CL)



The screenshot displays the website aeneid.citedloci.org/explore. The interface includes a navigation bar with "Aeneid in JSTOR", "Home", "Explore", and "About". A "Display:" dropdown is set to "all". On the left, a heatmap shows the distribution of references and quotations across the 12 books of the Aeneid. A tooltip indicates "References: 291" and "Quotations: 314". The "In Focus" section shows Book 1, lines 1-50, with the Latin text: "Arma virumque cano, Troiae qui primus ab oris Italiam, fato profugus, Laviniaque venit litora, multum ille et terris iactatus et alto vi superum saevae memorem Iunonis ob iram; multa quoque et bello passus, dum conderet urbem, inferretque deos Latio, genus unde Latinum, Albanique patres, atque altae moenia Romae. Musa, mihi causas memora, quo numine laeso, quidve dolens, regina deum tot volvere casus insignem pietate virum, tot adire labores impulerit. Tantaene animis caelestibus irae? Urbs antiqua fuit, Tyrii tenuere coloni, Karthago, Italiam contra Tiberinaque longe ostia, dives opum studiisque asperrima belli; quam Iuno fertur terris magis omnibus unam posthabita coluisse Samo; hic illius arma, hic currus fuit; hoc regnum dea gentibus esse, si qua fata sinant. iam tum tenditque fovetae."

The "Results" section shows two search results for "Aen. 1.1":

- Menin Virumque: Translating Homer with Virgil in "Epigrammata Bobiensia" 46_47 and 64**
SCOTT MCGILL
The Classical Journal 2006
DOI: [10.2307/30038019](https://doi.org/10.2307/30038019)
- Silius Punica 6. 1-53**
EDWARD L. BASSETT
Classical Philology 1959
DOI: [10.2307/266203](https://doi.org/10.2307/266203)

The second result includes a snippet: "(The Silian phrase will help to endorse foeda over

<http://aeneid.citedloci.org/>

1. Issues intrinsic to the Humanities:

- Fragmentation and variety of literature
- Lack of incentives
- Lack of central actors either private or public

2. Infrastructural issues:

- Availability of digital corpora
- Accessibility of catalog metadata
- Little adoption of standards to expose metadata uniformly
- Limited availability of APIs for harvesting (or data dumps)
- Legal frameworks for the use and publication of data
 - issues related to copyright on digitised materials
 - No copyright exception for text mining on contemporary scholarly publications

Thank you!

<https://www.researchgate.net/project/Linked-Books>

<https://www.researchgate.net/project/Cited-Loci>

giovanni.colavizza
matteo.romanello
{@epfl.ch}

Giovanni Colavizza & Matteo Romanello