

The Open Encyclopedia of Classical Sites:

Non-expressive Analysis of 20th Century Books



David Mimno
Princeton University
Computer Science

Funded by a **Google** Digital
Humanities Research grant

Thanks to the **Pleiades**
Project at NYU/ISAW

Outline

- Non-expressive use
- Building the database
- Do we need snippets?
- The Open Encyclopedia

Outline

- **Non-expressive use**
- Building the database
- Do we need snippets?
- The Open Encyclopedia

Realities of Copyright

- 1923 good. 1924 bad.
- Orphan works

Realities of Copyright

- 1923 good. 1924 bad.
- Orphan works



Realities of Copyright

- 1923 good. 1924 bad.
- Orphan works



Realities of Copyright

- 1923 good. 1924 bad.
- Orphan works



What is possible?

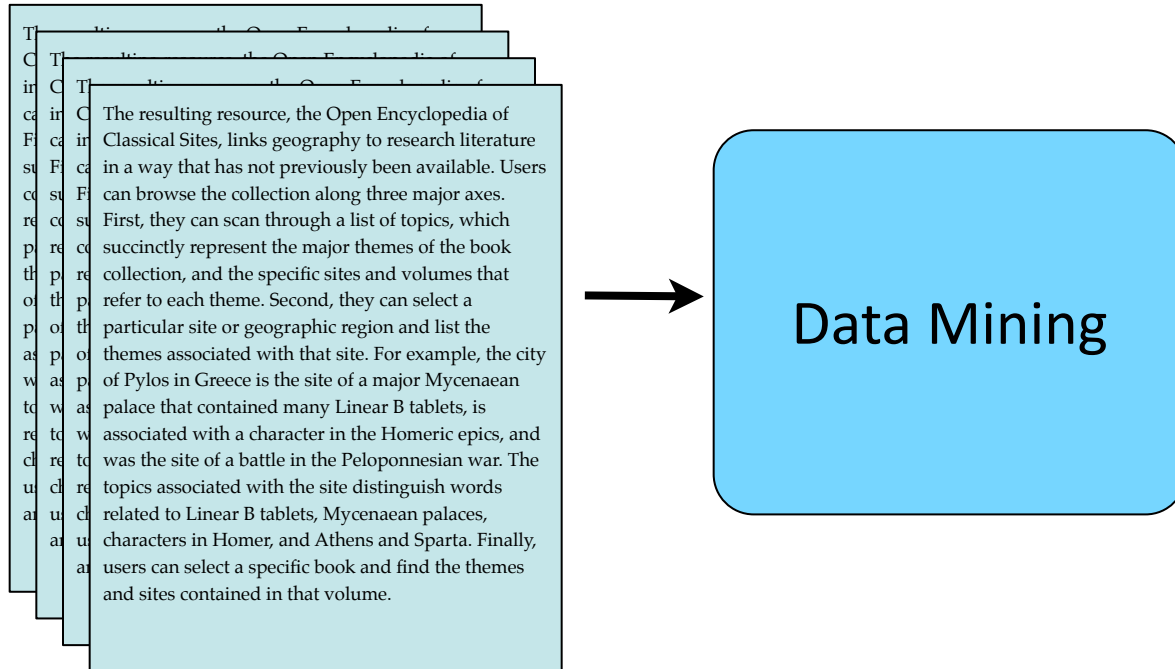
- *Non-consumptive* use
- *Non-expressive* use

Text Mining!

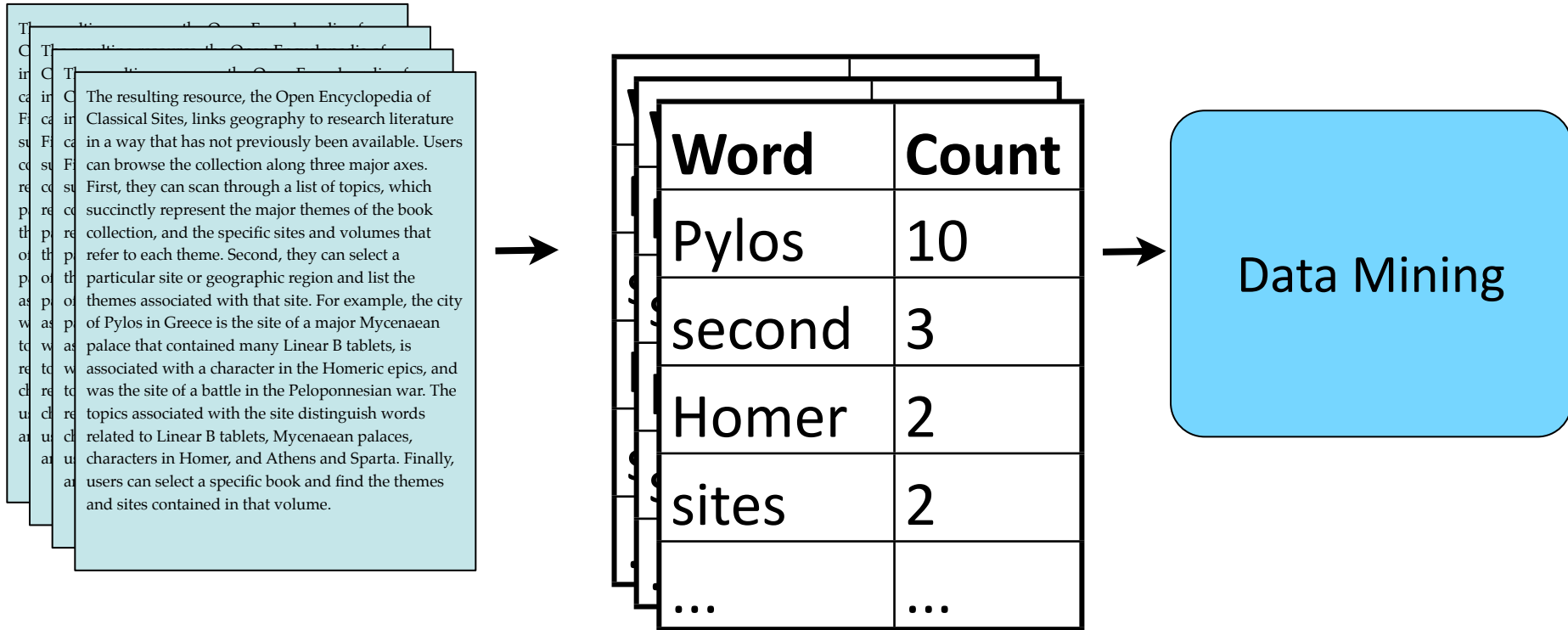
Pattern Recognition!

Machine Learning!

How can we get data?

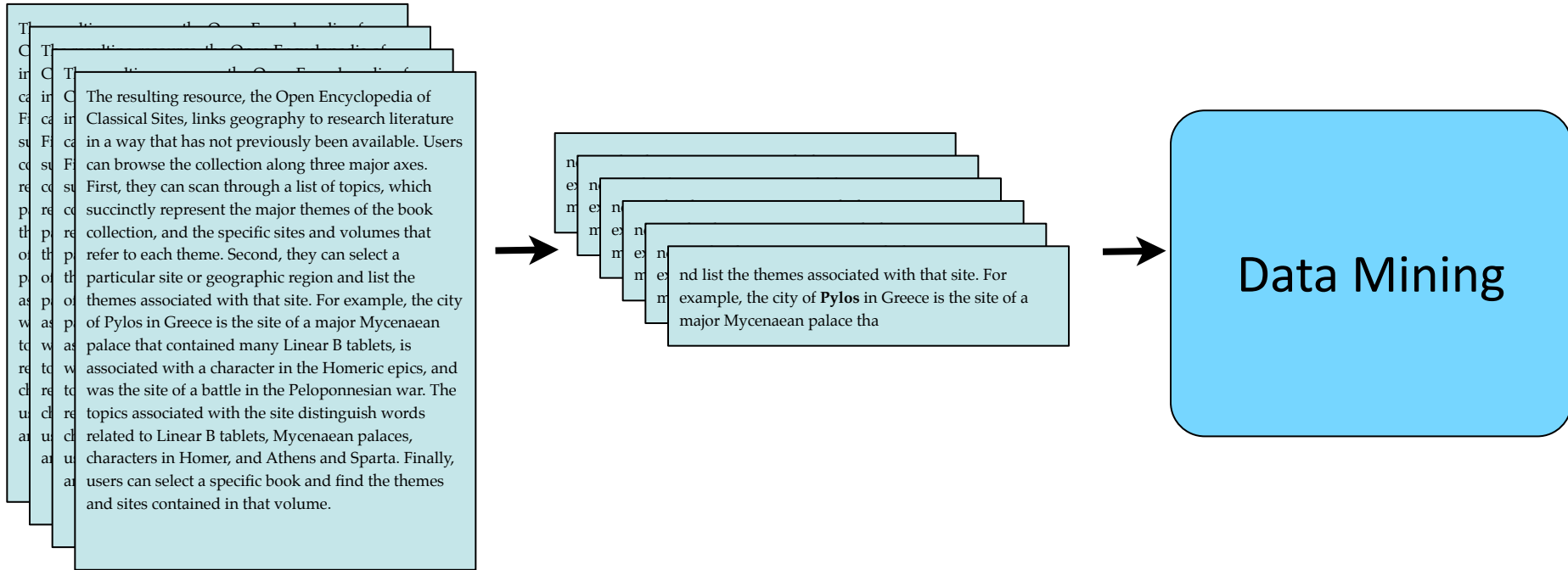


How can we get data?



Example: JStor Data For Research

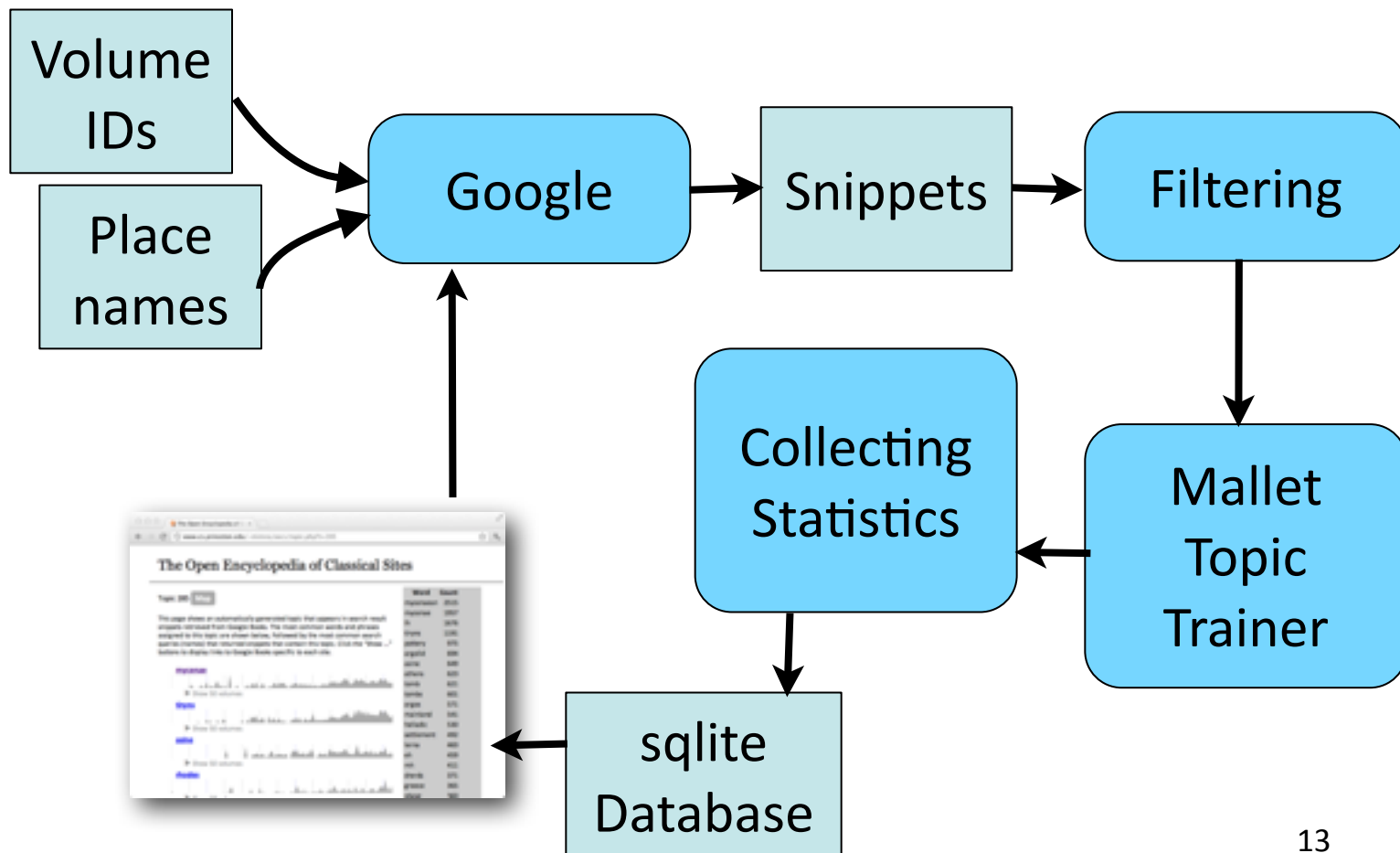
How can we get data?



Outline

- Non-expressive use
- **Building the database**
- Do we need snippets?
- The Open Encyclopedia

The process



Inputs and Outputs

- ~23,000 volumes
- ~5,000 site names
- 4.4 million English snippets

Filtering

- Overly ambiguous sites:
 - Split
 - Die
 - Philadelphia
- Indexes, tables of contents
- Syntactic words (stopwords)
 - the, and, of

Outline

- Non-expressive use
- Building the database
- **Do we need snippets?**
- The Open Encyclopedia

Example Data

- 9007199266296217 carthage th us, and last but not least is one of the most recent and most important international activities: the excavations of Carthage. You find in your Congress papers the first Bulletin of the Centre d'Etudes et de Documenta tion Archéologique de la
- 9007199266296217 carthage into the exchange. The Had rianic Baths at Lepcis Magna, completed in 127, and twenty years later the Antonine Baths at Carthage, were purely Roman in design, built in local African materials by African workmen; but a great deal of the
- 9007199266296217 carthage e picture of the issues of the mints of Citium, Paphos, and Salamis. Further west, the coinage of the mid-third century Carthage mint is the subject of an important new study by Bald us on the basis of the Bulla Regis hoard, while the silver coins

Example Data

- **9007199266296217** carthage th us, and last but not least is one of the most recent and most important international activities: the excavations of Carthage. You find in your Congress papers the first Bulletin of the Centre d'Etudes et de Documenta tion Archéologique de la
- **9007199266296217** carthage into the exchange. The Had rianic Baths at Lepcis Magna, completed in 127, and twenty years later the Antonine Baths at Carthage, were purely Roman in design, built in local African materials by African workmen; but a great deal of the
- **9007199266296217** carthage e picture of the issues of the mints of Citium, Paphos, and Salamis. Further west, the coinage of the mid-third century Carthage mint is the subject of an important new study by Bald us on the basis of the Bulla Regis hoard, while the silver coins

Example Data

- 9007199266296217 **carthage** th us, and last but not least is one of the most recent and most important international activities: the excavations of Carthage. You find in your Congress papers the first Bulletin of the Centre d'Etudes et de Documenta tion Archéologique de la
- 9007199266296217 **carthage** into the exchange. The Had rianic Baths at Lepcis Magna, completed in 127, and twenty years later the Antonine Baths at Carthage, were purely Roman in design, built in local African materials by African workmen; but a great deal of the
- 9007199266296217 **carthage** e picture of the issues of the mints of Citium, Paphos, and Salamis. Further west, the coinage of the mid-third century Carthage mint is the subject of an important new study by Bald us on the basis of the Bulla Regis hoard, while the silver coins

Example Data

- 9007199266296217 carthage **th us, and last but not least is one of the most recent and most important international activities: the excavations of Carthage. You find in your Congress papers the first Bulletin of the Centre d'Etudes et de Documenta tion Archéologique de la**
- 9007199266296217 carthage **into the exchange. The Had rianic Baths at Lepcis Magna, completed in 127, and twenty years later the Antonine Baths at Carthage, were purely Roman in design, built in local African materials by African workmen; but a great deal of the**
- 9007199266296217 carthage **e picture of the issues of the mints of Citium, Paphos, and Salamis. Further west, the coinage of the mid-third century Carthage mint is the subject of an important new study by Bald us on the basis of the Bulla Regis hoard, while the silver**

Example Data

- 9007199266296217 carthage th us, and last but not least is one of the most recent and most important international activities: the excavations of **Carthage**. You find in your Congress papers the first Bulletin of the Centre d'Etudes et de Documenta tion Archéologique de la
- 9007199266296217 carthage into the exchange. The Had rianic Baths at Lepcis Magna, completed in 127, and twenty years later the Antonine Baths at **Carthage**, were purely Roman in design, built in local African materials by African workmen; but a great deal of the
- 9007199266296217 carthage e picture of the issues of the mints of Citium, Paphos, and Salamis. Further west, the coinage of the mid-third century **Carthage** mint is the subject of an important new study by Bald us on the basis of the Bulla Regis hoard, while the silver coins

Example Data

- 9007199266296217 carthage thus, and last but not least is one of the most recent and most important international activities: the excavations of Carthage. You find in your Congress papers the first Bulletin of the Centre d'Etudes et de Documentation Archéologique de la
- 9007199266296217 carthage into the exchange. The Hadrianic Baths at Lepcis Magna, completed in 127, and twenty years later the Antonine Baths at Carthage, were purely Roman in design, built in local African materials by African workmen; but a great deal of the
- 9007199266296217 carthage e picture of the issues of the mints of Citium, Paphos, and Salamis. Further west, the coinage of the mid-third century Carthage mint is the subject of an important new study by Baldus on the basis of the Bulla Regis hoard, while the silver coins

Example Data

- 9007199266296217 carthage th us, and last but not least is one of the most recent and most important international activities: the excavations of Carthage. You find in your Congress papers the first Bulletin of the **Centre d'Etudes et de Documenta tion Archéologique de la**
- 9007199266296217 carthage into the exchange. The Had rianic Baths at Lepcis Magna, completed in 127, and twenty years later the Antonine Baths at Carthage, were purely Roman in design, built in local African materials by African workmen; but a great deal of the
- 9007199266296217 carthage e picture of the issues of the mints of Citium, Paphos, and Salamis. Further west, the coinage of the mid-third century Carthage mint is the subject of an important new study by Bald us on the basis of the Bulla Regis hoard, while the silver coins

Example Data

- 9007199266296217 carthage th us, and last but not least is one of the most recent and most important international activities: the excavations of Carthage. You find in your Congress papers the first Bulletin of the Centre d'Etudes et de **Documentation** Archéologique de la
- 9007199266296217 carthage into the exchange. The **Had rianic** Baths at Lepcis Magna, completed in 127, and twenty years later the Antonine Baths at Carthage, were purely Roman in design, built in local African materials by African workmen; but a great deal of the
- 9007199266296217 carthage e picture of the issues of the mints of Citium, Paphos, and Salamis. Further west, the coinage of the mid-third century Carthage mint is the subject of an important new study by **Bald us** on the basis of the Bulla Regis hoard, while the silver coins

Example Data

- 9007199266296217 carthage th us, and last but not least is one of the most recent and most important international activities: the excavations of Carthage. You find in your Congress papers the first Bulletin of the Centre d'Etudes et de Documenta tion Archéologique de la
- 9007199266296217 carthage into the exchange. The **Had rianic Baths** at **Lepcis Magna**, completed in 127, and twenty years later the **Antonine Baths** at Carthage, were purely Roman in design, built in local African materials by African workmen; but a great deal of the
- 9007199266296217 carthage e picture of the issues of the mints of Citium, Paphos, and Salamis. Further west, the coinage of the **mid-third century** Carthage mint is the subject of an important new study by Bald us on the basis of the **Bulla Regis** hoard, while the silver coins

Why not unigrams?

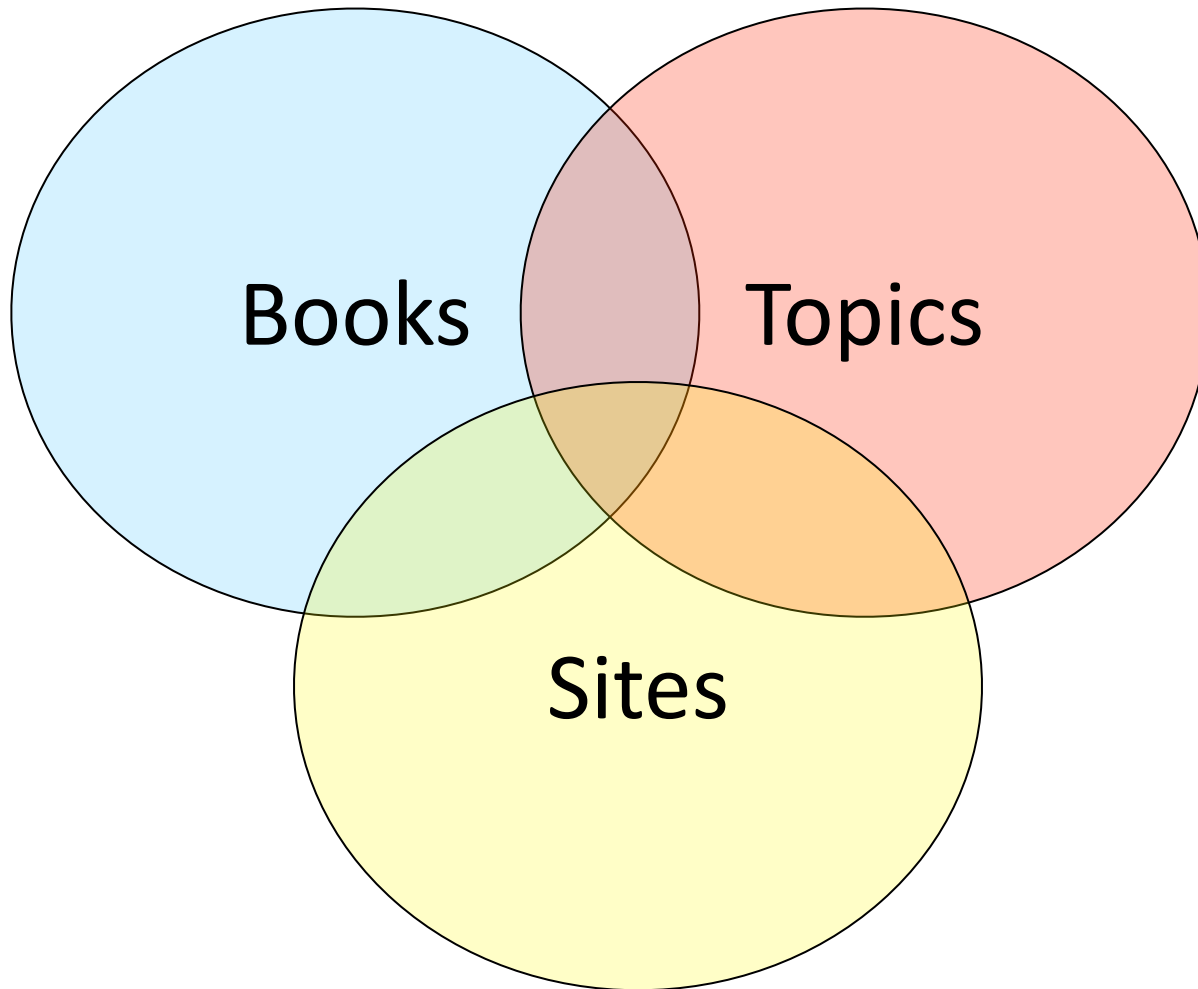
- Mixed languages
- Unmarked hyphens
- OCR errors
- Multi-word terms

- Targeted corpus creation

Outline

- Non-expressive use
- Building the database
- Do we need snippets?
- **The Open Encyclopedia**

Intersections of Three Entities



The Open Encyclopedia of Classical Sites

hadrian, emperor, trajan, rome, imperial, severus, caracalla, marcus, reign, coins, inscription, emperors, alexander, dio, arch, commodus, aurelius, pius, antoninus, domitian, vespasian, septimius, nero, syria, niger, claudius, inscriptions, empire, antioch, augustus, elagabalus, titus, coin, julia, severan, statue, visit, gordian, syrian

Books

Map

excavations, archaeology, excavation, survey, project, report, field, carried, antiquities, conducted, expedition, classical, results, excavated, reports, directed, team, direction, began, greek, season, archaeologists, preliminary, continued, worked, working, undertaken, exploration, athens, prehistoric, projects, finds, discoveries, archaeologist, ar, major, discovered, investigations, led

Books

Map

в, к, п, аз, és, вс, о, г, и, л, т, а, es, magyar, ш, sagalassos, í, я, waelkens, dl, acta, ez, nem, e, volt, у, ban, ben, budapest, el, ж, м, топ, excavation, оí, report, й, meg, р

Books

Map

The Open Encyclopedia of Classical Sites

Topic 285 [Map](#)

This page shows an automatically generated topic that appears in search result snippets retrieved from Google Books. The most common words and phrases assigned to this topic are shown below, followed by the most common search queries (names) that returned snippets that contain this topic. Click the "Show ..." buttons to display links to Google Books specific to each site.

[mycenae](#)

Show 50 volumes

[tiryns](#)

Show 50 volumes

[asine](#)

Show 50 volumes

[rhodes](#)

Word	Count
mycenaean	2515
mycenae	1957
lh	1676
tiryns	1191
pottery	975
argolid	694
asine	649
athens	623
tomb	621
tombs	601
argos	571
mainland	541
helladic	530
settlement	492
lerna	463
eh	459
mh	411
sherds	371
greece	365
nhase	360

buttons to display links to Google Books specific to each site.

mycena



▼ Hide 50 volumes



(605 words) Penelope A. Mountjoy, Deutsches Archäologisches Institut. *Regional Mycenaean decorated pottery.* (1999)



(453 words) Gisela Walberg. *Excavations on the Acropolis of Midea: results of the Greek-Swedish excavations under the direction of Katie Demakopoulou and Paul Åström.* Svenska Institutet i Athen. (1998)



(417 words) British School at Athens. *Lefkandi IV: the Bronze Age, the late Helladic III C settlement at Xeropolis.* British School of Athens. (2006)



(429 words) Gisela Walberg. *Excavations on the Acropolis of Midea: results of the Greek-Swedish excavations under the direction of Katie*

argolid	694
asine	649
athens	623
tomb	621
tombs	601
argos	571
mainland	541
helladic	530
settlement	492
lerna	463
eh	459
mh	411
sherds	371
greece	365
phase	360
aegean	354
excavations	340
grave	340
lefkandi	340
minoan	335
finds	326
wace	320
pylos	319
blegen	308
acropolis	305

The Open Encyclopedia of Classical Sites

lh iiic, lh iiib, mycenaean, mycenaean, lh, lh iiiia, tiryms, mycenaean tiryms, pottery, mycenaean, mycenaean pottery, argolid, asine, athens, tomb, tombs, middle helladic, argos, mainland, helladic, settlement, lerna, eh, tiryms mycenaean, mh, chamber tombs, sherds, greece, phase, aegean, excavations, grave, lefkandi, minoan, finds, shaft graves, wace, pylos, mh lh, stirrup jars, blegen, acropolis, lh iia, thebes, graves, tholos tomb, tholos tombs, citadel, frodin persson, palace

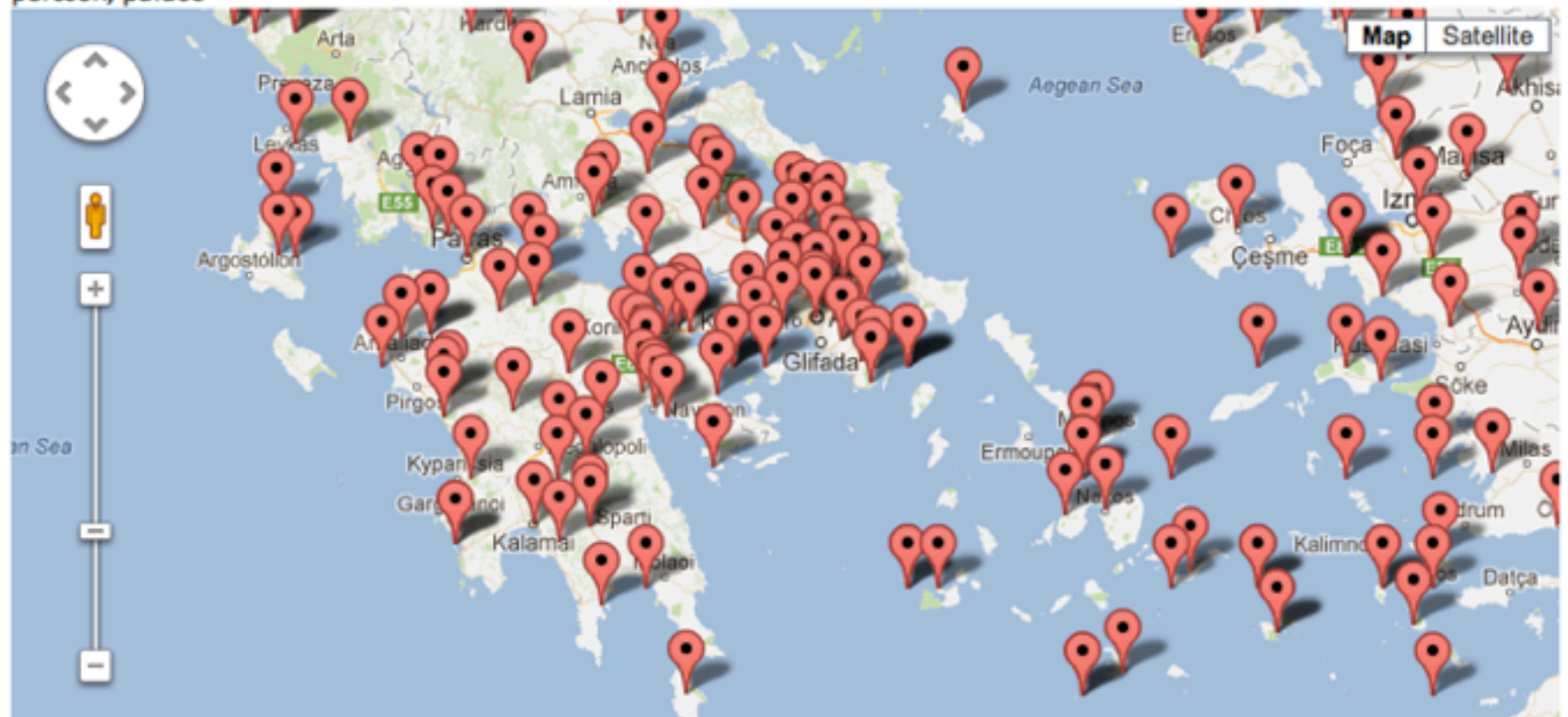
Books



The Open Encyclopedia of Classical Sites

lh iilc, lh iilib, mycenaean, mycenae, lh, lh iiii, tiryns, mycenae tiryns, pottery, myce naean, mycenaean pottery, argolid, asine, athens, tomb, tombs, middle helladic, argos, mainland, helladic, settlement, lerna, eh, tiryns mycenae, mh, chamber tombs, sherds, greece, phase, aegean, excavations, grave, lefkandi, minoan, finds, shaft graves, wace, pylos, mh lh, stirrup jars, blegen, acropolis, lh iia, thebes, graves, tholos tomb, tholos tombs, citadel, frodin persson, palace

Books



names, tablets, ra, personal names, ta, tablet, word, frithiof
rundgren, ri, pylos, texts, knossos, jo seph, ko, attested



Show 50 volumes

palace, linear tablets, tablets, mycenaean, crete, sir arthur evans,
knossos, minoan, mycenae, evans, pylos, linear, palace minos, cretan,
mainland



Show 50 volumes

lh iiic, lh iiib, mycenaean, mycenae, lh, lh iiia, tiryms, mycenae tiryms,
pottery, myce naean, mycenaean pottery, argolid, asine, athens,
tomb



Show 50 volumes

athens, war, sparta, athenian, peace, athenians, peloponnesian war,
thucydides, philip, spartan, king, alliance, spartans, sent, allies



Show 50 volumes

troy, homer, schliemann, trojan war, greek, heinrich schliemann,
homer, iliad, mycenae, trojan, iliad odyssey, homeric poems,
mycenaean, war, homer iliad



athens, war, sparta, athenian, peace, athenians, peloponnesian war, thucydides, philip, spartan, king, alliance, spartans, sent, allies

Hide 50 volumes



(198 words) Walter Robert Connor. *Thucydides*. Princeton Univ Pr. (1987)

pylos

Search



(197 words) Tim Rood. *Thucydides: narrative and explanation*. Oxford University Press, USA. (1998)

pylos

Search



(196 words) John Bagnell Bury, Stanley Arthur Cook, Frank E. Adcock, Martin Percival Charlesworth, Norman Hepburn Baynes, Charles Theodore Seltman. *The Cambridge Ancient History: Athens, 478-401 B. C.* (1954)

pylos

Search



(157 words) John Bagnell Bury, Stanley Arthur Cook, Frank Ezra Adcock, Martin Percival Charlesworth, Iorwerth Eiddon Stephen Edwards, John Boardman, Frank William Walbank. *The Cambridge Ancient History: The fifth century B.C., 2nd ed. 1992.* (1923)

pylos

Search



pylos



Books



Zu meiner Bibliothek hinzufügen

Rezension schreiben

Seite 197



1 Ergebnisse für pylos werden in diesem Buch angezeigt. - Sortieren nach: [Relevanz](#) | [Seiten](#)

[Suche löschen](#)

DRUCKVERSION DES BUCHES

KAUFEN

Kein E-Book verfügbar

[Princeton University Press](#)

[Amazon.de](#)

[Buch.de](#) - €27,80

[Buchkatalog.de](#)

[Libri.de](#)

[Weltbild.de](#)

[In einer Bibliothek suchen](#)

[Buchhandlungen in Ihrer Nähe suchen](#)

[Alle Händler »](#)



2+1 0



1 Rezension

Seite 197

The first of these, the comparison to **Pylos**, is eventually made explicit." But more important is the implicit comparison between the Athenian invasion of Sicily and the Persian invasion of Greece. The battle in the harbor inevitably evokes the ...

Für diese Seite ist keine Leseprobe verfügbar. [Dieses Buch kaufen.](#)

<http://www.cs.princeton.edu/~mimno/oecs>

mimno@cs.princeton.edu



David Mimno
Princeton University
Computer Science

Funded by a **Google** Digital
Humanities Research grant

Thanks to the **Pleiades**
Project at NYU/ISAW