Practical Approaches to Gazetteer Alignment

Ryan Baumann Duke Collaboratory for Classics Computing

Search results - 107 items matching your search terms

Ids from this batch: toggle visibility

Castellum Medianum Matidium Alexandrianum Tillinvense (of Vicus Augusti*Castellum Medianum Matidium andrianum Tillinvense?) 44 by E.W.B. Fentress - last modified Feb 14, 2012 01:05 PM - Relevance: 100%

Col. Nova Traiana Alexandriana (of stra/Col. Novs Traiana Alexandriana)

by J.P. Brown — last modified Jan 11, 2014 06:46 AM — Relevance: 97% (b) Alexandriana (of Arbela/Arba-ik.) A place name from the TWO Index (Nol. 1, p. 72) by B. Siewert-Mayer - last modified

Jan 11, 2014 05:57 AM - Relevance: 87% Vicus Augusti/*Castellum Medianum Matidium Alexandrianum Tillivense? An ancient place, cited: BAtlas 31 C4 Vicus Augusti/*Castellum Medianum Matidium Alexandrianum Tilinense? by E.W.B. Fentress - last modified Oct 20. 2012 12:02 PM - filed under: dare:ancient=1, dare:feature=setti dare:major+0 - Relevance: 77%

Colonia Nea Traiana Alex. B. (of Bostra/Col. Nova Traiana Alexandriana) A place name from the TAVO Index (Vol. 1, p. 352) by B. Siewert-Mayer - last modified Jan 11, 2014 06:48 AM -Relevance: 77%

Alexandria (of Alexandria) by S.E. Kroll - last modified Feb 14, 2012 06:05 PM - Relevance: 76%

() Alexandria (of Alexandria) by M.U. Erdosy - last modified Feb 14,

Next 20 items - [1] 2 3 4 5 6



🗉 📍 spot, building, farm 🗌

E 📍 forest, heath,... 🖽 🕈 undersea 🗆

What?

Ç	μ, F	Research		
	G S	me + Tools + Getty Thesaurus of Geographic Names + Search etty Thesaurus of Geographic Name earch Results	Results es® Online	
3.8	en Search	4 Previous Plage		E Hel
Find	l Name	: alexandria	Vernacular Display Eng	lish Display
Plac	e Type			
Nət	ion:			80 result
Clic Che	k the a	I con to view the hierarchy. es to view multiple records at once.	Page: 1 2 3 4	ante Laste
1.	- A	Alessandria (inhabited place) (World, Europe, Italy, Piedmont, Alessandria prov Alexandria a Palea Alexandria Stalicellorum Alexandria Statiellorum	rince) [7005125]	
2.	۰ ۸	Alexandra Channel (channel) (World, Asia, India, Andaman and Nicobar, Andar Alexandria Channel	nan Islands) [7544786]	
3.	 ▲ 	Alexándreia (inhabited place) (World, Europe, Greece, Macedonia, Imathía regio Alexándria	onal division) [7233761]	
4.	ି 🗚	Alexandria (inhabited place) (World, Africa, Egypt, Al-Iskandariyah governorat	te) [7001188]	

Alexándria

Alexandria Ad Aegyptum

ş

GeoNames Home | Postal Codes | Download / Webservice | About search alexandria login Map center : N 39º 56' 27" E 26º 35' 35" google earth tagzania mapquesi 4.52 Ann. Ukraine France Goog Map data 02016 Terms of Use 登ら日期日 € € Rinch Sea Italy Georgia searching for "alexandria" Azerb GeoNames Wikipedia Greece Turkmenis features 图 🖗 city, village,.... 🖽 📍 mountain,hill,rock,.... 🗌 🗄 🕈 stream, lake, 🗵 🕈 country, state, region,... 🗌 Im 🖩 📍 parks, area, 🗌 🗄 📍 road, railroad 🗌

Libya

Algeria

Low Map data @2016 Google Imagery @2016 TerraMetrics

	Name	country	feature	km to center
18	Alexandria 🗐	Egypt	seat of a first-order administrative division	1016.46 km
28	Alexandria 🗊	Romania	seat of a first-order administrative division	461.37 km
38	Alessandria	Italy	seat of a second-order administrative division	1571.84 km
48	Iskenderun 🤍	Turkey	seat of a second-order administrative division	915.1 km
58	Abū Qir 🗇	Egypt	populated place	1009.21 km
68	Al 'Ämiriyah	Egypt	populated place	1034.35 km
78	Borg El Arab 🗐	Egypt	populated place	1047.12 km
88	Bitash	Egypt	populated place	1023 km

Why?

- Transitive property: find out information about a resource from or through another context
- What does X have to say about this place?
- Pull information you want *back* into your project
- Discover weird things at the boundaries of your and other people's data
- Adding actual *links* between Linked Open Data



How?

- Use heuristics based on resource properties in common between the datasets you're trying to align
- Commonly: name, location
- My approach: find all name matches, then use location to filter
- Would also be possible to find all location matches, then use name to filter!

What's a name match?

- Pleiades: Alexandreia
- GeoNames: Alexandreia
- That was easy

What's a name match?

- Pleiades: Alexandreia
- GeoNames: Αλεξάνδρεια
- Not as easy

"Normalization"

- Perform a normalization step on name strings before matching
- Could write some manual rule or transliteration process that transforms "Αλεξάνδρεια" to "Alexandreia"
- But then what about للإسكندرية or الإسكندرية or
 אלכסנדריה?
- Seems like a hard problem other smart people would have already dealt with

International Components for Unicode

- We're already using Unicode, might as well benefit from it
- ICU developed for all sorts of hard internationalization problems
- Also established standards for transliteration between scripts

ICU Transliteration

Any-Latin; Lower; NFD; [:Nonspacing Mark:] Remove; [:Punctuation:] Remove; NFC

- Any-Latin: convert any character to Latin script
- Lower: convert to lowercase
- NFD: convert to Normalized Form D ("decompose" accents)
- [:Nonspacing Mark:] Remove; [:Punctuation:] Remove: remove accents and punctuation
- NFC: convert to Normalized Form C

ICU Transliteration

- transliterator = Transliterator.getInstance("Any-Latin; Lower; NFD; [:Nonspacing Mark:] Remove; [:Punctuation:] Remove; NFC")
- transliterator.transliterate("Alexandreia") = "alexandreia"
- transliterator.transliterate("Αλεξάνδρεια") = "alexandreia"

What's a name match?

- String equality: easy
- Substring match: easy, but sloppy
- Edit/Levenshtein distance: minimum number of edits to transform one string into another
- Cosine similarity: measure distance of strings in vector space - robust to rearrangement of words
- Phonetic algorithms: e.g. Metaphone

What's a location match?

- What's a "location"?
 - Point, collection of points, bounding box, line, arbitrary n-dimensional polygon...
- What's a "match" between any two kinds of these?
 - Seems like a hard problem other smart people would have already dealt with

Spatial Relationships/ DE-9IM

- Equals all points equal
- Disjoint no points in common
- Within, Contains one geometry is completely within another
- Overlaps some points in common, but not all
- Intersects at least one point in common
- Touches at least one boundary point in common, but no interior points in common
- Crosses the geometries actually overlap edges, e.g. a line crossing into a polygon
- DE-9IM matrix represents all 512 possible combinations of relations

What's a location match? Pleiades/GeoNames

- Pleiades: point(s) or bounding boxes
- GeoNames: points
- Point/point relationships or bounding box/point relationships

What's a location match? Pleiades/GeoNames

- Pleiades is a point: match if GeoNames point is within a distance threshold
- Pleiades is a bounding box: match if GeoNames point is inside that box
- Pleiades is a collection of points: match if GeoNames point is within distance threshold of centroid

Wait, what's a "place" match?

- Related to LOD vocabulary alignment problem
- Any meaningful alignment between an ancient place and a modern administrative boundary?

You are here: Home → Ancient Places → (Nea) Paphos

(Nea) Paphos

Creators: D. Rupp Contributors: DARMC, R. Talbert, Sean Gillies, Johan Åhlfeldt, Jeffrey Becker, Tom Elliott, Anne User Copyright © The Creators. Sharing and remixing permitted under terms of the Creative Commons Attribution 3.0 License (cc-by). Last modified Feb 01, 2014 01:58 PM — History

tags: dare:ancient=1, dare:major=1, dare:feature=major settlement, UWHS

The legendary birthplace of the goddess Aphrodite, (Nea) Paphos has been a center of activity since the Final Bronze Age when Mycenaeans honored the goddess there. The site was listed on the UNESCO World Heritage List in 1980.

Canonical URI for this page:



w place in AWMC's Antiquity A-la-carte, Google Earth, or Pelagios' Peripieo. Show area in GeoNames, Google Maps, or OpenStreetMap.

Search Site Search

Advanced Search...

Photos



Odeon at Paphos (I) by isawnyu.

46 other related photos...

Use this tag in Flickr to mark depictions of this place's site(s):

pleiades:depicts=707586

or this one to mark objects found here:

pleiades:findspot=707586

Related Content from Pelagios

(Nea) Paphos

American Numismatic Society (124); Inventory of Greek Coin Hoards (22); Nomisma.org Partner Objects (10); PELLA (9); iDAI.objects (3); Epigraphic Database Heidelberg (2)

Pelagios Datasets

Editing Places

http://pleiades.stoa.org/places/707586











PLEIADES



Show place in AWMC's Antiquity À-la-carte, Google Earth, or Pelagios' Peripleo.

Citation Guide





















Download - API -

Help -

Paris, Mount Everest, New York

Q × +

anonymous -



https://github.com/PerseusDL/gazComp



Wait, what's a "place" match?

- Related to LOD vocabulary alignment problem
- Any meaningful alignment between an ancient place and a modern administrative boundary?
- May be many-to-many, not one-to-one

Accessibility Contact Help Join/Login Site Status Log In

PLEIADES



This place has no portrait photo. One from the list of related photos could be suggested for the Pleiades Places group on Flickr.

mark depictions of this

or this one to mark objects

pleiades:findspot-606349

Pielades Data Model

Beference: Attributes of

Technical Introduction to

Accessibility Contact Help Join/Login Site Status

Barrington Atlas Map Grids

- <u>1: Internum Mare</u>
- <u>2: Hibernia-Scandinavia</u>
- <u>3: Asia Occidentalis</u>
- <u>4: Arabia-Azania</u>
- <u>5: India</u>
- 6: Asia Orientalis
- <u>7: Aremorica</u>
- <u>8: Britannia Superior</u>
- <u>9: Britannia Inferior</u>
- <u>10: Rhenus-Albis</u>
- 11: Sequana-Rhenus
- <u>12: Mogontiacum-Reginum-Lauriacum</u>
- <u>13: Vindobona</u>
- <u>14: Caesarodunum-Burdigala</u>
- <u>15: Arelate-Massalia</u>
- <u>16: Colonia Forum Iulii-Albingaunum</u>
- <u>17: Lugdunum</u>
- <u>18: Augustonemetum-Vindonissa</u>
- <u>19: Raetia</u>
- <u>20: Pannonia-Dalmatia</u>
- <u>21: Dacia-Moesia</u>
- <u>22: Moesia Inferior</u>
- <u>23: Tomis-Olbia-Chersonesos</u>
- <u>24: Asturica-Conimbriga</u>
- <u>25: Hispania Tarraconensis</u>
- <u>26: Lusitania-Baetica</u>
- 27: Hispania Carthaginiensis
- <u>28: Mauretania Tingitana</u>
- <u>29: Mauretania Caesariensis</u>
- <u>30: Iol Caesarea</u>
- <u>31: Cirta</u>
- <u>32: Carthago</u>
- <u>33: Theveste-Hadrumetum</u>
- <u>34: Numidia</u>
- <u>35: Tripolitana</u>
- <u>36: Garama</u>
- <u>37: Syrtica</u>
- <u>38: Cyrene</u>
- <u>39: Mediolanum</u>
- 40: Patavium
- <u>41: Pisae</u>
- 42: Arretium-Asculum
- 43: Latium Vetus

http://atlantides.org/capgrids/61

Grid of Barrington Atlas Map 61: Ephesus

Cartographic Scale 1:500,000 Map Bounds: Left, Bottom, Top, Right (degrees) (25.0, 36.0, 29.0, 38.5) Grid Columns A-H Grid Rows 1-5

See also: the Classical Atlas Project's Layout and Map Order document.

Alternate representations

- <u>RDF+XML</u>
- <u>Turtle</u>

GeoNames

About - Browse -

Download - API -

Help -

Paris, Mount Everest, New York



anonymous -





+



PLEIADES

Home Browse Places Download Data Blog Credits Documentation

You are here: Home --- Ancient Places --- Termessos

Termessos

Creators: C. Foss, S. Mitchell Contributors: DARMC, R. Talbert, Sean Gilles, Johan Ändekt, Jeffrey Becker, Tom Elliott Copyright © The Contributors: Sharing and remoting permitted under terms of the Creative Commons Attribution 3:0 License (co-by). Last modified Fub 26, 2013 05:06 FM — History

tags: dare ancient=1, dare major=1, dare feature=major settlement

An ancient place, cited: BAtlas 65 D4 Termessos

Canonical URI for this page:



DARMC location 6306 (550 BC - AD 640)

DARMC location 1455 (550 BC - AD 640)

DARE Location (550 BC - AD 640)

Termessos (550 BC - AD 640)

Place type anternary

Makes a connection with: Solution M.

Has a connection with:

References See Further: BAdas 65 D4 Termessos

 See Further: RE Termessos 2 Initial Provenance:

Barrington Atlas: BAtlas 65 D4 Termessos

Details The Barrington Atlas Directory notes: GUIUk Dağ

Alternate representations: Atom, JSON, KML, KML (Neighborhood), RDF+XML, Turtle

Suggested citation

Foss, C., S. Mitchell. "Places: 639139 (Termessos)". Pleiades. http://pleiades.stoa.org/places/639139> (Accessed: June 8, 2016 7:44 am]

Cite this resource in Wikipedia:

{{cite web !url=http://pleiades.stog.org/places/639139 !title=Places: 639139 (Termessas)
!author=Foss, C., S. Mitchell !accessdate=June 8, 2016 7:44 am !publisher=Pleiades}}

Log in to add comments

Copyright © Ancient World Mapping Center and Institute for the Study of the Anc cient World. Sharing and re

Accessibility Contact Help Join/Login Site Status

Research 11 M

Research Home + Tools + Thesaurus of Geographic Names + Full Record Display Getty Thesaurus of Geographic Names® Online

Full Record Display

4 Previous Page G New Learch

12 Help

Vernacular Display | English Display

Click the 1 icon to view the hierarchy.

Semantic View (JSON, JSONLD, RDF, NVTurtle, N-Triples)

ID: 5004400

A Termessus (deserted settlement)

Coordinates:

Use this tag in Flickr to mark depictions of this place's site(s):

pleiades:depicts-639139

or this one to mark objects found here:

pleisdes: Findspot-639139

Related Content from Pelagios

Termessos

ibility Contact Help Join/Login Site Status

Vocabs

Search

Photos

Advanced Search ...

This place has no portrait

photo. One from the list of related photos could be

suggested for the Pleiades

Places group on Flickr.

10 other related photos.

Log Ir

Search

Inventory of Greek Coin toards (5); PELLA (2); Nomisma.org Partner Objects (2): American Numismatic Society (1) Pelagios Datasets

Editing Places

Ctation Guide

Pielades Contributor Agreement

Pielades Data Model

Reference: Attributes of

Places

Technical Introduction to Places

Note:

English..... [VP]

G Hay Search

A Beck to tap

E Printer-driendly version

Record Type: administrative

Lat: 36 58 57 N degrees minutes Lat: 36.9820 decimal degrees Long: 030 27 53 E degrees minutes Long: 30.4640 decimal degrees

Elevation: 3280.840 feet 1000.0000 meters

Note: Was a Pisidian city built at an altitude of more than 1000 meters at the south-west side of the mountain Güllük Dağın; located 25 km NW of Antalya.

Names:

Termessus (preferred,C,V,S) Termessos (C,V) ancient Τερμησσός (C,V,Greek,U)

Hierarchical Position:

World (facet) Asia (continent) (P) A Turkey (nation) (P) Antalya (province) (P)

Place Types:

deserted settlement (preferred, C) was ancient city archaeological site (C) inhabited place (H)

Related geographic places:

near/adjacent to Antalya (inhabited place) (World, Asia, Turkey, Antalya) [7002378] near/adjacent to Güllük Dağı (mountain) (World, Asia, Turkey, Antalya) [7594003]

Sources and Contributors:

Termessos...... [GRLPSC, VP Preferred] GRI Photo Archive, Authority File (1998) 8543 Princeton Encyclopedia of Classical Sites (1979) 896-897 Termessus..... [VP] Atlas of Greek & Roman World (1981) 52 Τερμησσός..... [VP] ... Archaeology Magazine [online] (2000-) Subject: [GRLPSC, VP] ARLIS/NA: Ancient Site Names (1995) Atlas of Greek & Roman World (1981) 52 GRI Photo Archive, Authority File (1998) 8543

.......... Princeton Encyclopedia of Classical Sites (1979)

The J. Paul Getty Trust

@ 2004 J. Paul Getty Trust Terms of Use / Privacy Policy / Contact Us

Current Alignments

- Pleiades<->GeoNames: <u>https://github.com/ryanfb/pleiades-plus</u>
- Pleiades<->TGN (Getty Thesaurus of Geographic Names): <u>https://github.com/ryanfb/pleiades-tgn</u>
- CHGIS/Regnum Francorum<->GeoNames: Merrick Lex Berman & Johan Åhlfeldt, Historical Gazetteer System Integration

Future Alignments?

- Pleiades<->OpenStreetMap/OpenHistoricMap
- Pleiades<->Wikipedia/Wikidata
- Pleiades<->VIAF
- Syriac Gazetteer<->*
- Al-Thurayya<->*

Future Integration

Surface quality links back in source projects themselves

Bonus Round: GeoRefine

- <u>https://github.com/ryanfb/georefine</u>
- Adds GeoSpatial functions to OpenRefine: <u>http://openrefine.org/</u>

GREL functions

- · point : creates a point from latitude and longitude in the WGS84 coordinate reference system
- · distance : computes the distance between two points in meters
- parseGeoJson : parses GeoJSON string
- parseWkt : parses WKT string
- All geometry relationship functions described here:
 - equals
 - contains
 - within
 - overlaps
 - crosses
 - touches
 - intersects
 - disjoint
 - relate

Bonus Bonus Round? Book Aligner

http://ryanfb.github.io/book-aligner/





Book Aligner

Find matching books for a given HathiTrust, Internet Archive, or Google Books URL.

https://archive.org/details/opvscvlaacademi03heyngoog						Submit		
						Search:		
	Identifier 🔶	Title		Year	Volume 🔶	Pages		Score 🕌
盦	opvscvlaacademi03heyngoog	Opvscvla academica collecta et animadversionibvs locvpletata		1785	1	448	9772746	100
G	n_VOqOVGGcYC	Opvscvla academica collecta et animadversionibvs locvpletata		1785		458		100
Ş?	mdp.39015058525919	Opvscvla academica collecta et animadversionibvs locvpletata.		1785	v.1		9772746	5
P	uiug.30112098696773	Opvscvla academica collecta et animadversionibvs locvpletata.		1785	v.1-2		9772746	5
Ê	opvscvlaacademi01heyngoog	Opvscvla academica collecta et animadversionibvs locvpletata		1812	6	494	9772746	0
G	BxZLh5ZZcA0C	Opvscvla academica collecta et animadversionibvs locvpletata		1812		518		0

Showing 1 to 6 of 6 entries

Thank you!!!