University of Kentucky UKnowledge

Library Presentations

University of Kentucky Libraries

4-25-2024

Preserving the Uncrawlable: Serving the Server

Andrew McDonnell University of Kentucky, mcdonnell@uky.edu

Follow this and additional works at: https://uknowledge.uky.edu/libraries_present

Part of the Archival Science Commons, Byzantine and Modern Greek Commons, and the Digital Humanities Commons

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Repository Citation

McDonnell, Andrew, "Preserving the Uncrawlable: Serving the Server" (2024). *Library Presentations*. 273. https://uknowledge.uky.edu/libraries_present/273

This Presentation is brought to you for free and open access by the University of Kentucky Libraries at UKnowledge. It has been accepted for inclusion in Library Presentations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu, rs_kbnotifs-acl@uky.edu.

Preserving the Uncrawlable: Serving the Server

tiny.cc/ukweb2024

 $\bullet \bullet \bullet$

Andrew McDonnell, Digital Archivist University of Kentucky Libraries Special Collections Research Center April 25, 2024



University of Kentucky Libraries Web Archiving

Established in 2018 <u>https://archive-it.org/organizations/915</u>

ARCHIVE-IT	EARN MORE CONTACT US	The leading web archiving service for collecting and accessing cultural heritage on the web <i>Built at the Internet Archive</i>		
Explore >> University of Kentucky Libraries	University of Kentucky Libraries Archive-II Partner Since: Jun, 2018 Organization Type: <u>Colleges & Universities</u> Organization IRI: <u>http://http://http://kentucky.uk/</u> Description: The University of Kentucky (UK) Libraries selects, preserves, and facilitates access to archive copies of websites and social media that are official and permanent records of the University of Kentucky that are relevant to Special Collections Research Center (SCR) collection priorities; and that support university faculty and student research and teaching. You may come across content in these archived onl resources that you find harmful or offensive. For a full statement about harmful content and to contact us about it, Jenes see our Web archiving guide. The guide also includes information about the UK Libraries web archiving program, including how to search our web archiver, Guide: https://libguides.aky.edu/web_archiving/intro If you have a Kentucky-related or University of Kentucky or UK-related website or social media account that you believe has historical value and should be preserved			
Narrow Your Results	UK-related website or social media accountell us about it using this form: https://formside.com/sites and collections from this organization			
Narrow Your Results Subject Sort By: Count (A-Z)	UK-related website or social media accountell us about it using this form: https://formside.com/sites and collections from this organization	ms.gle/dtpVAqH2XfY6/iucA n are listed below. Narrow your results at left, or enter a search q		

Founded in 1998

Still live: www.cs.uky.edu/~raphael/sol/sol-html



Publis

David V Techni Rapha

Search Abbrevi Σορός Σοσσυ

Σού Σουβο Σούκιν Σοῦ γ Σούπε Σουπη Σούσια Σουφεί Σοφάκ Suda On Line: Byzantine Lexicography This site has moved to a temporary new home.

5

Register About SOL Features Editors Rights Hel

Published by the Stoa Consortium	Entire list of entries					
Senior Editor David Whitehead	Login ID: Password: Login					
Technical Director Raphael Finkel	The Suda on Line					
Managing Editors William Hutton Catharine Roth Patrick Rourke Elizabeth Vandiver David Whitehead Founders Raphael Finkel William Hutton Patrick Rourke Ross Scaffe † Elizabeth Vandiver Programmers Emeriti Huar En Ng Mukund Chandak Shahid Saleen Mohammed Kamal Shah Help Search Tios Dasay. Octons About Relators Hadrond List Search Tios Search Thos Search Thos Creeks Abbreviations	For Attic phrase in Plato let them seek, I poach in Suidas for unlicens'd Greek. — Alexander Pope, <i>The Dunciad</i> <u>4.227-8</u> Pope's 'Suidas' is not a man but a work. The <i>Suda</i> (or Stronghold): a massive 10 th century Byzantine Greek historical encyclopedia of the ancient Mediterranean world, covering the whole of Greek and Roman antiquity and also including Biblical and Christian material. Preserved in several medieval manuscripts, it has been edited and published several times since the end of the 14th century in traditional hard-copy scholarly editions, most recently that of Ada Adler (Teubner, 5 volumes: 1928-1938, reprinted 1971). The Suda On Line (SOL) project, begun in 1998 as part of the Stoa Consortium, opens up this stronghold of information by means of a freely accessible, keyword-searchable database, with English translations, notes, bibliography, and links to other electronic resources. With contributions (as Translators and/or Editors) from more than two hundred people worldwide, the SOL reached the landmark of all entries being translated and "vetted" (edited) to a usable standard on July 21, 2014. But more can, and will, be done. A fuller history of the project may be found <u>here</u> .					
Σορός Σοσίανος Σού Σούβαρμάχιος Σούγαίακα Σούγτίακα Σούγτίακα Σούγτιρός Σούαιοι Σουποριανός Σούαιοι Σουφείρ Σουφάκας	Users of our website have three options. 1. Without registration you can gain immediate access to the <u>entire list of entries</u> . It may be sufficient if you already know which entry or entries (by Adler number) you want to consult. You may also undertake simple searching here: Either Choose a subject area and click to browse					

Founded in 1998

https://wayback.archive-it.org/11384/*/https: //www.cs.ukv.edu/~raphael/sol/sol-html/

Looks Great!

You are viewing an archived web page collected at the request of <u>University of Kentucky Libraries</u> using <u>Archive-It</u> . This page was captured on 21:47:56 Sep 21, 2019, and is part of the <u>University of Kentucky Faculty, Staff, and</u> <u>Student Work Web Archives</u> collection. The information on this web page may be out of date. See <u>All versions</u> of this archived page. Found 0 archived media items out of 0 total on this page. <u>Metadata</u> <u>Enable QA</u>							
AN A		uda On Line: Byzantíne Lexicography nis site has moved to a temporary new home.					
	<u>Regi</u>	ster About SOL Features Editors Rights Help					
Published by the Stoa Consortium Senior Editor David Whitehead		Entire list of entries Login ID: Password: Login					
Technical Director Raphael Finkel		The Suda on Line					
Managing Editors William Hutton Catharine Roth Patrick Rourke Elizabeth Vandiver David Whitehead		For Attic phrase in Plato let them seek, I poach in Suidas for unlicens'd Greek. — Alexander Pope, <i>The Dunciad</i> <u>4,227-8</u> Pope's 'Suidas' is not a man but a work, The Suda (or Stronghold): a massive 10 th century					
Founders Raphael Finkel		Byzantine Greek historical encyclopedia of the ancient Mediterranean world, covering the whole of Greek and Roman antiquity and also including Biblical and Christian material.					

Preserved in several medieval manuscripts, it has been edited and published several times since the end of the 14th century in traditional hard-copy scholarly editions, most recently that of Ada Adler (Teubner, 5 volumes: 1928-1938, reprinted 1971). The Suda On Line (SOL) project, begun in 1998 as part of the Stoa Consortium, opens up this stronghold of information by means of a freely accessible, keyword-searchable database, with English translations, notes, bibliography, Shahid Saleem Mohamme and links to other electronic resources. With contributions (as Translators and/or Editors) from more than two hundred people worldwide the SOL reached the landmark of all entries being

House of lies.

William Hutton Patrick Rourke

Elizabeth Vandiver

Mukund Chandak

Programmers Emeriti

Ross Scaife †

Huar En Ng

Kamal Shah

Founded in 1998

- What and Why?(Appraisal)
- How? (Capture/Ingest)
- Larger philosophical questions (Context/Respect des fonds)

1499 edition housed by the Bavarian State Library and shared online by the Munich DigitiZation Center (MDC)

https://daten.digitale-sammlungen.de/bsb0005075 <u>6/image_16</u>

שיותו, אמלט פל משיטי שי ביואף מול "אירפסי. פי זו לעו אב ארפי אידו מג, אנאיזו לע עברא נוסה", צידו ע זו ל קלול פולבת ווש אוקטיב, שידם אי כ די לא עו עו לווח מב, מי שי לא אפליי, שע די פעוע אבוד ליטי ישי די לי אשרדה אוש שי שלא האש על שלא האשר לעו לעו איז איש לעל איל איל לעו אין איש אין איש אין איש אין איש אין איש אין אי o Ta The and pos o Ta The o To F was down we rised by Fyrenew, The but Ta rieds of we A Bartoc 250 Hubs. A paris on usac on its yoc oursed 406 & yearwood - & we rou us our הא עלואים ול - ב זמ אווי לאפצ זה אלגעושל. באמל מאווים ליום זהי אמל מאשו לאיי ב לדם אומיסב מקוצוא Gic บัสบริตออร์อเบ นี้ แต่ for หนะ อู้ อน อน เช่นติส เร็ญ dda. อีน รอบ อ แปลอ ภอรอบ แมือออ อ เรอ c าลิ สมาอ אין א לידי וליא ל לפר אין ביש איי שהקאספי מושלי ילל א לימוש איא א איי געיט ב. לי לע, ב רביים Proc otrora ve co Baic for migar & an Ador Contapia upin ulu anni D. pan gon round adia ו אל נוד לא נוסד ע שלא חמוי דוור, לי ובי ב די וא המ אמו של נקצ לע טוי לי אי או שי נ לר שני חול ב מני דועל , ב מ ל מפוי בני שה לניני שרינה אודי ע קיוא לשע אין שיות לי לעיד לאעשומלש. טדו ליע למלפור טואשא Japon Zued Tos an Buy & Oan 1000. Or oi 2 poles & To 1 2 2 rana oup 1000 . Wardranges us-אינו עיף לאישף פואקטע לפר אל לי שתפעוגמסו דוו מאלי לי לאנט אוא לשרא ואי דעד לומ ים לבעימצע העוב יד טוו בנויד ט און קצע שלו בן איט אלי ל ארי שלישא לי שאו גלטב מי מ ל מרטי ט שלי לע אין אב איידט אוביי שטי הש שמו שריי " אספמי איוג הב שמו ל P מישי אי גם כולום אי לשי דשי אל גמטים עלעים P למיעוי . שאמרום 20 cons on paror. Co to anor bor du abori eup or roy on run i Dodor u Disa usen unan i sa vio. Low Cononpa puent rang on the most antie our is po work whire a paris a paleidos AS. Abac oo Diac isolina on pur pate Erige prochadarine. Apararisoc an En The statiunas cadoninasogo elendo deno poparans of sever sinas riles . or adon malan o Hunori jehond st air sand on to moiners & a paran so and a be obing. A paran a + Wing avant ofor Squar wed pa on an give no cual one a Fai wat to va be bunar . A Gar is dreing Sau codo gapar sarelube. Apartan riber peans in abachdu popoe. A 651 runte la più pidison Ta d'in di Tig 651 rugdela E duo jegado fe. Jaguartega di à אוצער מוריעי אסר עעפעולט אי אפא עלשי (אוא אי א אי אי איין איין לאזי A Cone ad אימטע ב טיים את א אנטרים שדאו דאר, ג א איצו דאר אלו של אופת ב אט לאפת לו איז ד. זה לע לעופת לי, ב בע קפופת יועל To marod. Had So TELPA Dir of Edipolotis. A Cons. Spond Ruleion. Applanocod salis E Line la Jamioa. A Gibara la pipan pa youla i ond di no ona pipara di i riaro la piona pidi וֹהָם . המוידה המ הועם. כי הי הוא היום וועוווו וועו וועו וועו היה האי באטיו מ הולה איוי הולה פרולי ב Bibanasonabagoc. Aban yocada u. She Boroc Edinaso vaneye. E min mobar . This Eturi Bou menuzazar Corzen c, av auger rau Obayateis con ra dor A pa range in . aver u de nap ושי דעדומיעוד כעולם ל אונש אבופטיער אוע מער גביעש באועשוני יי אל אבא לע ברעדולטיב ליני אמלו השיי ש ואשלוסי אומון מי אי אמשלט למשלט אמלו בא. שלט על לוע ל ע אמלוטו מע זה הצימופוריוע שהאי 24 . 5 Sind in Suaripsic auto a res pur la me Sk gern pale to se flad med ano boc E'a mp na לא השאה על כל נוגוים עולא עולי הוומני חפמרי אוויוא נוע מי ל אמאחשי אל דיור של א לא איז איז איז איז איז איז איז Air Mon un la ano obivia your uala me avegue & to . and la uter us par maice ab mea. A BEATERONANNE ONCOOCE api ATE. A BY TEROC. avoids. douvebe. orope AT un unionen ม์ ch สิสปี คลิยคล่ mpoc รี่ได สรีส ปี paye ps Ame py mp no o for poc 3 แล้ ก็ร้าง สางอาร์ท אליותי אבי יי דער עמו מיוש עושי מיום מיט אלב ב ל אלצור עבי געיט דעלב , אטע משלפט לי אלצו אינ יי ב שור בא מ pupla Einstenen Tup a mea puora Englo Barra nu liga mor, inci 20 no be usa se or bi dame mear עצעים בישע הגאדוף שמורד אדו מ אי אדופטי. אנושטה של מארא דופו ל מאראדוני מ איצמטטאיואיי אטיא ב אועפת אל אוףטבי האלי טביט אי אלא אנה שאפא אאל כים לוא דופוצ דוטאיווטט מיוצ א מיי פי

Founded in 1998

- Translation of 10th century encyclopedia written in Greek
- Crowd-sourced project, pre-dating Wikipedia
- Product of the Stoa
 Consortium for Electronic
 Publication in the Humanities
- Only English translation of the Suda, in any format

Founded in 1998

Still live: www.cs.uky.edu/~raphael/sol/sol-html



Publis

David V Techni Rapha

Search Abbrevi Σορός Σοσσυ

Σού Σουβο Σούκιν Σοῦ γ Σούπε Σουπη Σούσια Σουφεί Σοφάκ Suda On Line: Byzantine Lexicography This site has moved to a temporary new home.

5

Register About SOL Features Editors Rights Hel

Published by the Stoa Consortium	Entire list of entries					
Senior Editor David Whitehead	Login ID: Password: Login					
Technical Director Raphael Finkel	The Suda on Line					
Managing Editors William Hutton Catharine Roth Patrick Rourke Elizabeth Vandiver David Whitehead Founders Raphael Finkel William Hutton Patrick Rourke Ross Scaffe † Elizabeth Vandiver Programmers Emeriti Huar En Ng Mukund Chandak Shahid Saleen Mohammed Kamal Shah Help Search Tios Dasay. Octons About Relators Hadrond List Search Tios Search Thos Search Thos Creeks Abbreviations	For Attic phrase in Plato let them seek, I poach in Suidas for unlicens'd Greek. — Alexander Pope, <i>The Dunciad</i> <u>4.227-8</u> Pope's 'Suidas' is not a man but a work. The <i>Suda</i> (or Stronghold): a massive 10 th century Byzantine Greek historical encyclopedia of the ancient Mediterranean world, covering the whole of Greek and Roman antiquity and also including Biblical and Christian material. Preserved in several medieval manuscripts, it has been edited and published several times since the end of the 14th century in traditional hard-copy scholarly editions, most recently that of Ada Adler (Teubner, 5 volumes: 1928-1938, reprinted 1971). The Suda On Line (SOL) project, begun in 1998 as part of the Stoa Consortium, opens up this stronghold of information by means of a freely accessible, keyword-searchable database, with English translations, notes, bibliography, and links to other electronic resources. With contributions (as Translators and/or Editors) from more than two hundred people worldwide, the SOL reached the landmark of all entries being translated and "vetted" (edited) to a usable standard on July 21, 2014. But more can, and will, be done. A fuller history of the project may be found <u>here</u> .					
Σορός Σοσίανος Σού Σούβαρμάχιος Σούγαίακα Σούγτίακα Σούγτίακα Σούγτιρός Σούαιοι Σουποριανός Σούαιοι Σουφείρ Σουφάκας	Users of our website have three options. 1. Without registration you can gain immediate access to the <u>entire list of entries</u> . It may be sufficient if you already know which entry or entries (by Adler number) you want to consult. You may also undertake simple searching here: Either Choose a subject area and click to browse					

Why Won't you Just Crawl?

 $\bullet \bullet \bullet$

Server-Side Operations

Using this website

Users of our website have three options.

1. Without registration you can gain immediate access to the <u>entire list of entries</u>. It may be sufficinumber) you want to consult. You may also undertake simple searching here:

			1 0
Eithe	athletics	\sim	and click to browse
	Choose a subject area aetiology	-	
or find	agriculture		100 per page ∽ and click to search
	architecture		
2. Registration as a C		l	vanced search features, including the Greek t
these but would r	athletics biography		ne login ID guest and the password sol in the
<u>a guest</u> .	botany		
Registration as an	' children		opriate qualifications, if approved by the Ma
See the instruction	Christianity chronology		<u>egister as an editor</u> .
	clothing		
Citing the Suda On Li	comedy		
Detailed instructions for o	constitution i daily life definition		blished work or in CVs, may be found <u>here</u> .
All our entries, including	dreams		e covered by a Creative Commons 'Attribution
features of how we apply			<u>ite</u> .
_	epic	-	
Source identifications	in the Suda on Line		

You are viewing an archived web page collected at the request of <u>University of Kentucky Libr</u>. <u>University of Kentucky Faculty, Staff, and Student Work Web Archives</u> collection. The inform archived media items of

Suda On Line

The search string is empty Search file invoked

Preserving function, not just form.

Idea 1: Preserve the Whole Server

The Importance of the Donor

Passive vs. Active Web Archiving

A bird in hand.

OVF/OVA file

-Open Virtualization Format/Appliance

-Preserve the full disk image as a virtual machine

-Document content and instructions

-Easy!

A bird in hand.

OVF/OVA file

-Open Virtualization Format

-Preserve the full disk image as a virtual machine, import in VirtualBox

-Document content and instructions

-Easy!

Possible?

A bird in hand.

OVF/OVA file

-Open Virtualization Format

-Preserve the full disk image as a virtual machine

-Document content and instructions

-Easy!

Possible?

Possible!

The .OVA It's something.

Pros

-The full server is preserved

-It works

Cons

-Offline storage limits access

-Long-term sustainability?

-Requires full access to server

-Requires all website components share one server (no external resources/API integration)

Idea 2: ReproZip-Web

They're presenting here tomorrow!

Capture server-side actions

Also capture external website elements

Package & reproduce sites far more complex than Suda On Line

Pros: -Potentially ideal for efficiently capturing DH projects

-Smaller output than server image

-Lower barrier to serving access than VM

Challenges: -Technical complexity -Requires access/permissions to servers -Widespread acceptance of .rpz standard/sustainability

That being said: Oui. .RPZ

A Larger Question

What is a website when it's no longer online?

The Offline Suda On Line

To what extent is SOL a tool?

To what extent a text?

The Offline Suda On Line

Aboutness & Context.

The Future of the Suda On Line

The conversation continues...

References

Boss, Katherine, Ilya Kreymer, Vicky Rampin, and Rémi Rampin. "Server-Side Web Archiving Using Reprozip-Web". *iPRES 2023 Proposal Submission*. <u>https://drive.google.com/file/d/1D9h_qc7jd4KP58DxlsFhzNxr7ckiPtPy/view</u>. Accessed 2024 April 14.

Mahoney, Anne. "Tachypaedia Byzantina: The Suda On Line as Collaborative Encyclopedia". *digital humanities quarterly*. 3:1, 2009, https://www.digitalhumanities.org/dhq/vol/003/1/000025/000025.html. Accessed 2024 April 14.

Suda On Line: Byzantine Lexicography. https://www.cs.uky.edu/~raphael/sol/sol-html/index.html . Accessed 2024 April 14.

Suda, 1499, National Library of Munich, <u>https://daten.digitale-sammlungen.de/0005/bsb00050756/images/index.html?fip=193.174.98.30&id=00050756&seite=1</u>. Accessed 2024 April 14.

Theimer, Kate. "Archives in Context and as Context." Journal of Digital Humanities. June 2012, <u>https://journalofdigitalhumanities.org/1-2/archives-in-context-and-as-context-by-kate-theimer/</u>. Accessed 2024 April 14.

University of Kentucky Web Archives, <u>https://archive-it.org/organizations/915</u>. Accessed 2024 April 14.

Thank you! Questions?

Andrew McDonnell Digital Archivist University of Kentucky Libraries Special Collections Research Center mcdonnell@uky.edu

> Presentation Link: tiny.cc/ukweb2024