

Shawn M. Jones

825 Calle Mejia Apt. 224
Santa Fe, New Mexico
USA, 87501

✉ jones.shawn.m@gmail.com
☎ +1 (757) 560-7909
🌐 <https://www.shawnmjones.org>

Summary

I am a self-motivated researcher and software engineer looking for a research position where I can advance the state of the art of computer science. My interests include automatic summarization, machine learning, natural language processing, information retrieval, image analysis, computer vision, digital preservation, digital libraries, and social media.

Education

Ph. D. (Computer Science) 2021, Old Dominion University, Norfolk, VA, USA
M. Sc. (Computer Science) 2015, Old Dominion University, Norfolk, VA, USA
B. Sc. (Computer Science) 1999, Old Dominion University, Norfolk, VA, USA

Experience

Postdoc October 2021 – present, Information Sciences Division,
Los Alamos National Laboratory, Los Alamos, New Mexico, USA

- explored the intersection between web resources and computer vision

Experience

Graduate Research Assistant October 2019 – October 2021, Research Library Prototyping Team,
Los Alamos National Laboratory, Los Alamos, New Mexico, USA

- developed automatic summarization tools for scholarly articles and web archive collections
- conducted an image analysis study that produced a solution for predicting the striking image that best summarizes a news article or scholarly publication
- continued to lead the Dark and Stormy Archives (DSA) Project that conducts research and develops tools for summarizing web archive collections and presented project at the Web Archiving and Digital Libraries workshop and Coalition for Network Information (CNI) meeting
- helped secure grant funding for the DSA Project
- developed and documented the Robust Links Project with a journal article
- implemented the Robust Links Zotero extension with JavaScript
- released the tool Hypercane, allowing researchers and archivists to intelligently sample web archive collections as part of corpus summarization
- co-mentored students on engineering and research topics
- integrated web archives and news articles to produce automated daily news summaries in the SHARI process, presented at WADL 2020

Experience (continued)

Graduate Research Assistant July 2013 – October 2019, Department of Computer Science, Old Dominion University, Norfolk, Virginia, USA

- restarted and led the DSA Project
- explored the relationships between wiki revisions and web archive document captures
- collaborated on a mathematical model of the Memento Protocol in an effort to better understand its effects on temporally sensitive information
- developed a MediaWiki extension in PHP for Memento
- gathered data via the REST API interfaces of various web services as part of multiple efforts
- quantified how well some US government web sites had been preserved in public web archives, helping the archiving community capture as much of the content as possible before the White House shuttered the sites
- researched and developed summarization algorithms for documents and corpora
- developed Raintale and MementoEmbed for summarizing archived web pages and then applied them as a part of a crowdsourced user study to determine which summarization visualization techniques work best for understanding web archive collections
- helped draft reports for grant agencies on project progress
- applied topic modeling, textual similarity measures, and metrics to identify off-topic documents in web archive collections
- leveraged Weka's machine learning to predict collection types with 10-fold cross-validation
- quantified how well Google+ was preserved so that teams could archive the rest of it before Google decommissioned it

Graduate Research Assistant July 2015 – May 2017, Research Library Prototyping Team, Los Alamos National Laboratory, Los Alamos, New Mexico, USA

- explored the intersection of scholarly communications, social networking, and web archiving
- developed the Memento client Python library for easily accessing archived web pages
- leveraged text similarity measures such as cosine similarity, Jaccard Index, Sørensen-Dice distance, Spansum, and Simhash as part of a study quantifying the degree of reference rot in scholarly publications
- categorized publisher websites for potential linked data as part of the Signposting project
- conducted a study on the issues surrounding the application of web archives to temporally sensitive information, using television show spoilers as a case study
- analyzed the poor archivability of scholarly portals with tools such as Heritrix and Webrecorder
- categorized the different potential crawling patterns for various web portals as part of the Scholarly Orphans project
- explored the use of blockchains for timestamping archived web pages
- analyzed how to disambiguate scholars on online portals
- collaborated on a standard for accessing un-augmented (raw) archived web content

Experience (continued)

Computer Scientist May 1997 – July 2015, Space and Naval Warfare Systems Command (SPAWAR), Systems Center Norfolk, Norfolk, Virginia, USA

- led several software projects, working with project managers and technical leads to support Agile software engineering methods
- mentored newer employees, helping them improve career prospects
- served as the Build Manager for a 50+ developer software project, leading a team of engineers to create solutions for automated deployment and software dependency analysis
- created software release and deployment systems by integrating Python, Git, BASH, Maven, Make, Ant, and Java
- implemented automated testing strategies for integrating software components
- developed Maven plugins in Java to manage dependency information
- created installation packages for a variety of systems using tools such as RPM, PIP, and InstallShield
- architected build systems for continuous integration using SVN, Hudson/Jenkins, and Nexus for a 50+ developer project
- ported current applications to older operating systems like SCO Unixware and HP/UX
- programmed security exploits in C and C++ to help test US Navy software
- maintained snort as a Network Intrusion Detection System
- employed the eEye Retina and Tenable Nessus vulnerability scanners to report vulnerabilities in software applications
- performed forensics on user workstations to assist the Naval Criminal Investigative Services (NCIS)

Publications, Presentations, and Other Writings

Journal Articles

- 1 **Jones, Shawn M.**, Martin Klein and Herbert Van de Sompel (2021). ‘Robustifying Links To Combat Reference Rot’. In: *Code4Lib Journal* 50.2021-02-10. URL: <https://journal.code4lib.org/articles/15509>.
- 2 **Jones, Shawn M.**, Michael L. Nelson and Herbert Van de Sompel (2018). ‘Avoiding spoilers: wiki time travel with Sheldon Cooper’. In: *International Journal on Digital Libraries*, pp. 1–17. URL: <http://dx.doi.org/10.1007/s00799-016-0200-8>.
- 3 **Jones, Shawn M.**, Herbert Van de Sompel, Harihar Shankar, Martin Klein, Richard Tobin and Claire Grover (2016). ‘Scholarly Context Adrift: Three out of Four URI References Lead to Changed Content’. In: *PLOS ONE* 11.12, pp. 1–32. URL: <http://dx.doi.org/10.1371/journal.pone.0167475>.

Book Chapters

- 1 **Jones, Shawn M.**, Martin Klein, Herbert Van de Sompel, Michael L. Nelson and Michele C. Weigle (2021). ‘Interoperability for Accessing Versions of Web Resources with the Memento Protocol’. In: *The Past Web: Exploring Web Archives*. Ed. by Gomes, Daniel and Demidova, Elena and Winters, Jane and Risse, Thomas. New York: Springer International Publishing, 101–126. ISBN: 978-3-030-63290-8. URL: https://doi.org/10.1007/978-3-030-63291-5_9.

Conference Proceedings

- 1 **Jones, Shawn M.**, Valentina Neblitt-Jones, Michele C. Weigle, Martin Klein and Michael L. Nelson (2021). 'It's All About The Cards: Sharing on Social Media Probably Encouraged HTML Metadata Growth'. In: *Proceedings of the 2021 ACM/IEEE Joint Conference on Digital Libraries*. [To be published in 2021]. URL: <https://arxiv.org/abs/2104.04116>.
- 2 **Jones, Shawn M.**, Michele C. Weigle, Martin Klein and Michael L. Nelson (2021a). 'Automatically Selecting Striking Images for Social Cards'. In: *Proceedings of the 13th ACM Web Science Conference*. Southampton, England, UK. URL: <https://doi.org/10.1145/3447535.3462505>.
- 3 **Jones, Shawn M.**, Michele C. Weigle, Martin Klein and Michael L. Nelson (2021b). 'Hypercanes: Intelligent Sampling for Web Archive Collections'. In: *Proceedings of the 2021 ACM/IEEE Joint Conference on Digital Libraries*. [To be published in 2021].
- 4 **Jones, Shawn M.**, Michele C. Weigle and Michael L. Nelson (2019). 'Social Cards Probably Provide For Better Understanding Of Web Archive Collections'. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. Beijing, China, pp. 2023–2032. URL: <https://doi.org/10.1145/3357384.3358039>.
- 5 **Jones, Shawn M.**, Michael L. Nelson and Michele C. Weigle (2018). 'The Many Shapes of Archive-It'. In: *Proceedings of the 15th International Conference on Digital Preservation*. Boston, Massachusetts. URL: <http://dx.doi.org/10.17605/OSF.IO/EV42P>.
- 6 **Jones, Shawn M.**, Michael Nelson and Michele Weigle (2018). 'The Off-Topic Memento Toolkit'. In: *Proceedings of the 15th International Conference on Digital Preservation*. Boston, Massachusetts. URL: <http://dx.doi.org/10.17605/OSF.IO/UBW87>.
- 7 Van de Sompel, Herbert, Martin Klein and **Shawn M. Jones** (2016). 'Persistent URIs Must Be Used To Be Persistent'. In: *Proceedings of the 25th International Conference Companion on World Wide Web*. Montréal, Québec, Canada, pp. 119–120. URL: <http://dx.doi.org/10.1145/2872518.2889352>.

Masters Thesis

- 1 **Jones, Shawn M.** (2015). 'Avoiding Spoilers on MediaWiki Fan Sites Using Memento'. MSc. Norfolk, VA, United States of America: Old Dominion University. URL: <https://doi.org/10.25777/d8hw-b984>.

Technical Reports and Preprints

- 1 **Jones, Shawn M.**, Martin Klein, Michele C. Weigle and Michael L. Nelson (2020). *MementoEmbed and Raintale for Web Archive Storytelling*. Tech. rep. arXiv:2008.00137. Los Alamos National Laboratory, pp. 1–54. URL: <https://arxiv.org/abs/2008.00137>.
- 2 **Jones, Shawn M.**, Alexander C. Nwala, Martin Klein, Michele C. Weigle and Michael L. Nelson (2020). *SHARI – An Integration of Tools to Visualize the Story of the Day*. Tech. rep. arXiv:2008.00139. Los Alamos National Laboratory, pp. 1–19. URL: <https://arxiv.org/abs/2008.00139>.
- 3 **Jones, Shawn M.**, Michael L. Nelson and Michele C. Weigle (2019). *Social Cards Probably Provide For Better Understanding Of Web Archive Collections*. Tech. rep. arXiv:1905.11342. [preprint]. Old Dominion University, pp. 1–58. URL: <https://arxiv.org/abs/1905.11342>.
- 4 **Jones, Shawn M.** and H. Shankar (2016). *Rules of Acquisition for Mementos and Their Content*. Tech. rep. arXiv:1602.06223. Los Alamos National Laboratory, pp. 1–16. URL: <http://arxiv.org/abs/1602.06223>.

- 5 **Jones, Shawn M.** and Michael L. Nelson (2015). *Avoiding Spoilers in Fan Wikis of Episodic Fiction*. Tech. rep. arXiv:1506.06279. [preprint]. Old Dominion University, pp. 1–18. URL: <http://arxiv.org/abs/1506.06279>.
- 6 **Jones, Shawn M.**, Michael L. Nelson, Harihar Shankar and Herbert Van de Sompel (2014). *Bringing Web Time Travel to MediaWiki: An Assessment of the Memento MediaWiki Extension*. Tech. rep. arXiv:1406.3876. Old Dominion University, pp. 1–23. URL: <http://arxiv.org/abs/1406.3876>.

Other

- 1 **Jones, Shawn M.** (2021). ‘Dark and Stormy Archives Tutorials’. Continuing Education to Advance Web Archiving (CEDWARC). URL: <https://www.youtube.com/playlist?list=PLNDhuTLyz9EPPhan-U3TfSh3uMnXNWL0A>.
- 2 **Jones, Shawn M.**, Martin Klein, Michele C. Weigle and Michael L. Nelson (2021). ‘The Dark and Stormy Archives Project: Summarizing Web Archives Through Social Media Storytelling’. International Internet Presentation Consortium. URL: <https://netpreserveblog.wordpress.com/2021/02/03/the-dark-and-stormy-archives-project/>.
- 3 **Jones, Shawn M.** (2020a). ‘MementoEmbed and Raintale for Web Archive Storytelling’. Web Archiving and Digital Libraries (WADL) Workshop. URL: <https://youtu.be/7mvMEZWM8mA>.
- 4 **Jones, Shawn M.** (2020b). ‘SHARI - An Integration of Tools to Visualize the Story of the Day’. Web Archiving and Digital Libraries (WADL) Workshop. URL: <https://youtu.be/hnSTiePxQAA>.
- 5 **Jones, Shawn M.** (2020c). ‘Summarizing Web Archives through Storytelling with the Dark and Stormy Archives Project’. Coalition for Networked Information (CNI) Fall 2020 Meeting. URL: <https://vimeo.com/478666716>.
- 6 Klein, Martin, **Shawn M. Jones**, Herbert Van de Sompel and Michael L. Nelson (2020). ‘Robustify your links! A working solution to create persistently robust links’. International Internet Presentation Consortium. URL: <https://netpreserveblog.wordpress.com/2020/07/29/robustify-your-links-a-working-solution-to-create-persistently-robust-links/>.
- 7 **Jones, Shawn M.** (2018). ‘Improving Collection Understanding in Web Archives’. ACM/IEEE Joint Conference on Digital Libraries Doctoral Consortium. URL: <https://www.slideshare.net/shawnmjones/improving-collection-understanding-in-web-archives>.
- 8 Van de Sompel, Herbert, Michael L. Nelson, Lyudmila Balakireva, Martin Klein, **Shawn M. Jones** and Harihar Shankar (2017). ‘Uniform Access to Raw Mementos’. International Internet Preservation Consortium Web Archiving Conference 2017. URL: <https://www.slideshare.net/martinklein0815/uniform-access-to-raw-mementos>.
- 9 **Jones, Shawn M.** (2014a). ‘Reconstructing the past with MediaWiki’. WikiConference USA. URL: <http://www.slideshare.net/shawnmjones/reconstructing-the-past-with-media-wiki-35333484>.
- 10 **Jones, Shawn M.** (2014b). ‘Using the Memento MediaWiki Extension to Avoid Spoilers’. WikiConference USA. URL: <http://www.slideshare.net/shawnmjones/using-the-memento-mediawiki-extension-to-avoid-spoilers-35333526>.
- 11 **Jones, Shawn M.** (Apr. 2013a). ‘A Brief Introduction to Test-Driven Development’. Old Dominion University Chapter of the Association for Computing Machinery. URL: <http://www.slideshare.net/shawnmjones/tdd-36889681>.

- 12 Jones, Shawn M. (2013b). 'Continuous Integration: Finding Problems Soonest'. Defense Acquisition University. URL:
<http://www.slideshare.net/shawnmjones/continuous-integration-36889803>.

Grant Project Participation

- 2021-2021 Improving the Dark and Stormy Archives Framework by Summarizing the Collections of the National Library of Australia
International Internet Preservation Consortium
<https://netpreserve.org/projects/dark-and-stormy-archives/>
- 2018-2020 Continuing Education to Advance Web Archiving
Institute of Museum and Library Services Grant # RE-70-18-0005-18
- 2017-2018 Combining Social Media Storytelling With Web Archives
Institute of Museum and Library Services Grant # LG-71-15-0077-15
- 2016-2017 Preservation of Scholarly Record on the Web
Andrew W. Mellon Foundation Grant # 11600663
- 2013-2015 Enhanced Memento Support for MediaWiki
Andrew W. Mellon Foundation Grant # 31300111

Research Project Participation

- 2020-present Robust Links (<https://robustlinks.mementoweb.org>)
- 2019-present Scholarly Orphans (<https://myresearch.institute/about/>)
- 2017-present The Dark and Stormy Archives Project (<https://oduwsdl.github.io/dsa/>)
- 2015-2016 Hiberlink (<http://hiberlink.org>)
- 2013-present Memento (<https://mementoweb.org>)

Software Project Participation

- 2020-present Robust Links (<https://robustlinks.mementoweb.org/>)
- 2019-present Hypercane (<https://oduwsdl.github.io/hypercane/>)
- 2019-present Raintale (<https://oduwsdl.github.io/raintale/>)
- 2018-present MementoEmbed (<https://oduwsdl.github.io/dsa/>)
- 2017-present The Off-Topic Memento Toolkit (<https://pypi.org/project/otmt/>)
- 2017-present Archive-It Utilities (<https://pypi.org/project/aiu/>)
- 2015-2016 Python Memento Client (https://pypi.org/project/memento_client/)
- 2013-2016 Memento MediaWiki Extension
(<https://www.mediawiki.org/wiki/Extension:Memento>)
- 2010-2013 Naval Tactical Command Support System Open Architecture (NTCSS-OA)
- 2009-2010 BarCode Supply (BCS)
- 2008-2015 Naval Aviation Logistics Command Management. Information System (NALCOMIS)
- 2007-2007 Expeditionary Pack Up Kit (EPUK)
- 1998-1999 Intranet Web Services (IWS)
- 1997-2015 Naval Tactical Command Support System (NTCSS)

Service To The Profession

Reviewer - International Journal on Digital Libraries 2020

Reviewer - The Web Conference 2020

Program Committee - Web Archiving and Digital Libraries 2020

Professional Memberships

Association for Computing Machinery (ACM) # UK81403

Institute of Electrical and Electronics Engineers (IEEE) # 97636430

International Information System Security Certification Consortium (ISC)² # 307483

Awards

2018	Old Dominion University <i>Dominion Scholar</i>
2007	Naval Facilities Engineering Systems Command (NAVFAC) <i>Certificate of Appreciation for Navy C&A Process Training</i>
2006, 2003, 2002, 2001	Space and Naval Warfare Systems Command (SPAWAR) <i>Lightning Bolt Team Excellence Award</i>