Webengine

The World Congress on Engineering and Computer Science (WCECS'14)

San Francisco, USA, 22-24 October, 2014

Matthew Harris

Lawrence Livermore National Laboratory

LLNL-CONF-656083

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Software Used

- Hyde
  - Hyde is a static website generator written in python. While Hyde took life as awesome Jekyll's evil twin, it has since been completely consumed by the dark side and has an identity of its own.
  - Hyde desires to fulfill the lofty goal of removing the pain points involved in creating and maintaining static websites.
- Webshooter
  - Webshooter is a command-line assistant to the static website generator Hyde or Jekyll. It produces Hyde or Jekyll ready websites in the user's choice of layout and automates page creation.
- Jekyll
  - Jekyll is a simple, blog-aware, static site generator. It takes a template directory containing raw text files in various formats, runs it through Markdown (or Textile) and Liquid converters, and spits out a complete, ready-to-publish static website suitable for serving with your favorite web server. Jekyll also happens to be the engine behind GitHub Pages, which means you can use Jekyll to host your project's page, blog, or website from GitHub's servers for free.
- Bootstrap
  - Bootstrap is the most popular HTML, CSS, and JS framework for developing responsive, mobile first projects on the web.

Content Management Systems

- Drupal
  - Drupal is an open source content management platform powering millions of websites and applications. It's built, used, and supported by an active and diverse community of people around the world.
- Zope
  - Zope is a free and open-source web application server written in the object-oriented programming language Python.
- Plone
  - Plone is among the top 2% of all open source projects worldwide, with 340 core developers and more than 300 solution providers in 57 countries. The project has been actively developed since 2001, is available in more than 40 languages, and has the best security track record of any major CMS.
  - It is owned by the Plone Foundation, a 501(c)(3) not-for-profit organization, and is available for all major operating systems.
- Moin Moin
  - MoinMoin is an advanced, easy to use and extensible WikiEngine with a large community of users. Said in a few words, it is about collaboration on easily editable web pages.

Generated HTML vs Markdown

- Output from Drupal

```
<div class="link-wrapper">
  <ul class="links inline">
    <li class="node-readmore first">
      <a href="/?q=node/11" rel="tag" title="ArticleTwo"> Read more
        <span class="element-invisible"> about ArticleTwo </span>
      </a>
    </li>
    <li class="comment_forbidden last">
      <span>
        <a href="/?q=user/login&amp;destination=node/11%23comment-form">Log in</a> to post comments
      </span>
    </li>
  </ul>
</div>
```

- Output from Webengine (markdown)

```
* [Read more](/?articleid=11) about ArticleTwo
* [Log in](/?login=user) to post comments
```

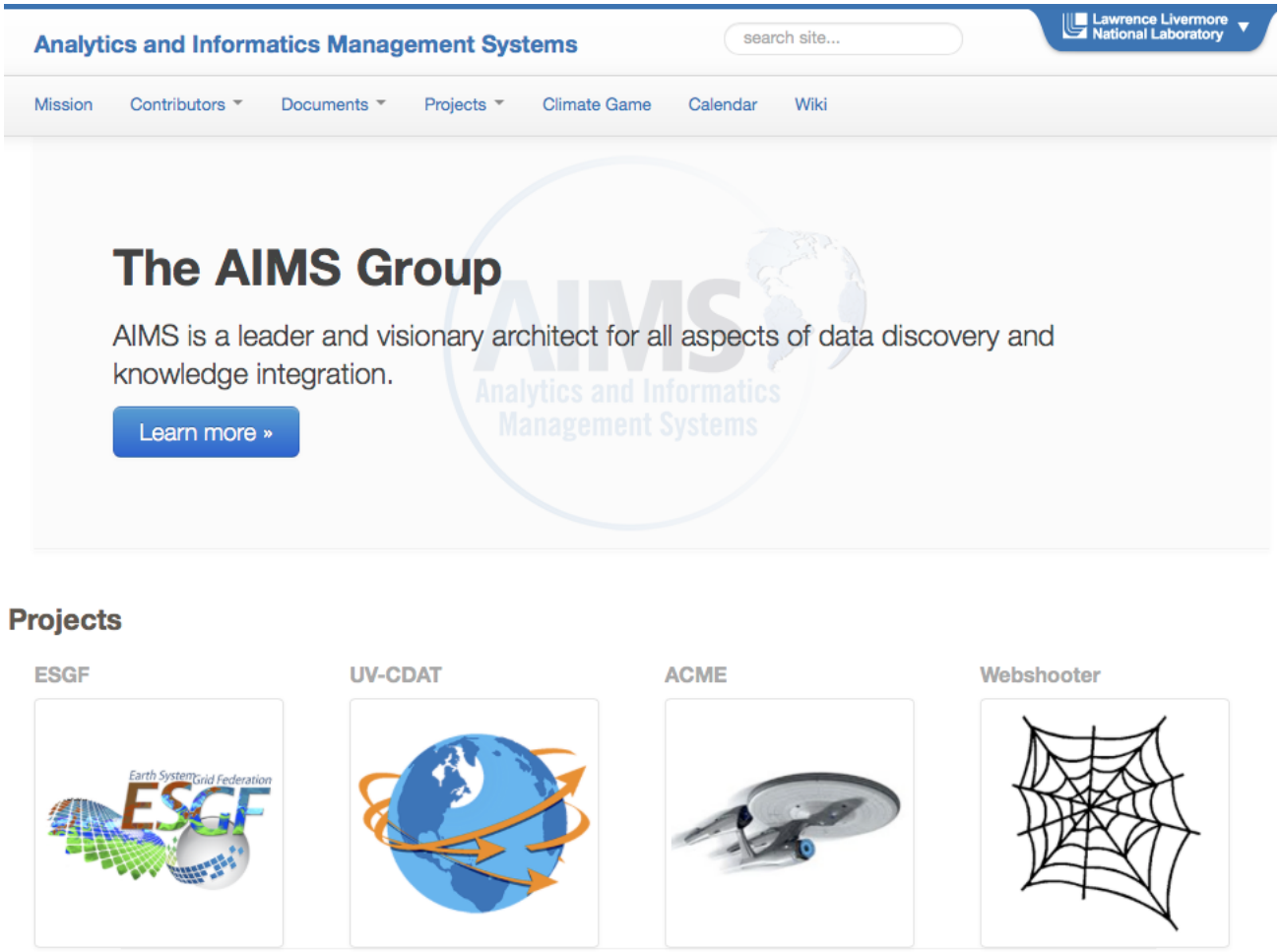
webengine.py

- *url_gatherer*: Takes a site-index.html page and generates a list of URLs
- *file_gatherer*: Fetches (WGET) all the pages in the URL list
- *site_gatherer*: Takes a site URL and fetches (WGET) the whole site
- *file_extractor*: Recursively reads all site files, and saves a little of the content as possible
- *file_corrector*: Reads all new files changing href and src to user input
- *html_table_2_markdown*: Parses all files changing html tables to GitHub flavored markdown
- *file_converter*: Converts the remaining html to markdown (plain text – Arron Swartz html2text.py)
- *image_gatherer*: Creates a unique list of all site images and downloads them
- *bold_cleanup*: Removes all ** and __ in all files
- *head_adder*: Adds a Hyde or Jekyll header and footer to each page

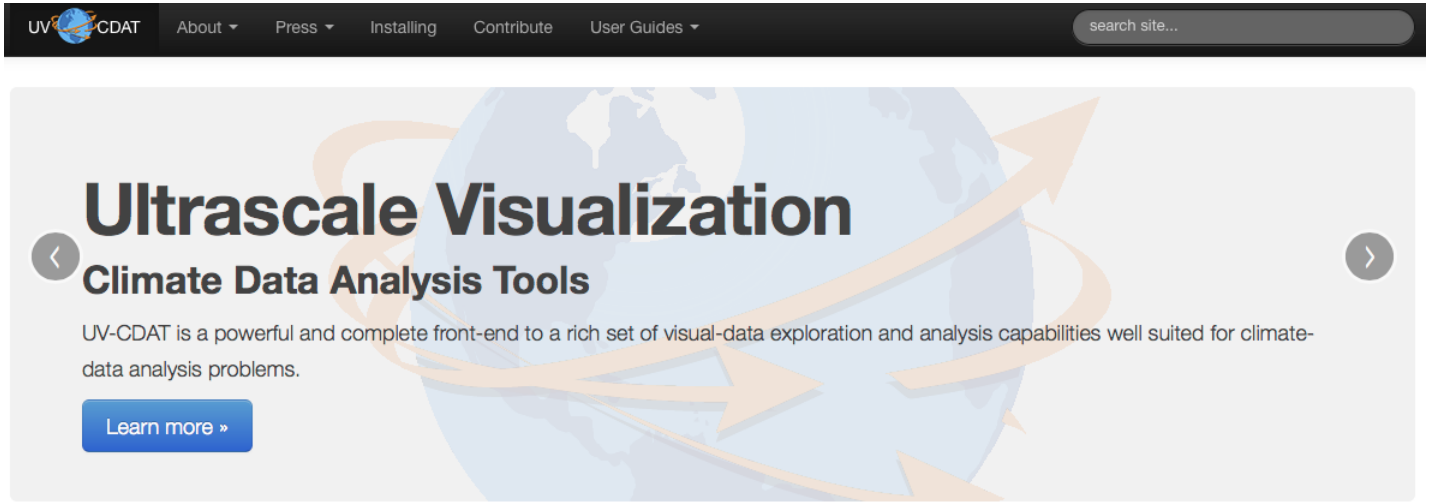
webshooter.py

- Now with Jekyll Support!
- Creates a Hyde or Jekyll static generating website with one of three site styles.
- Customizable, easy to add new styles.

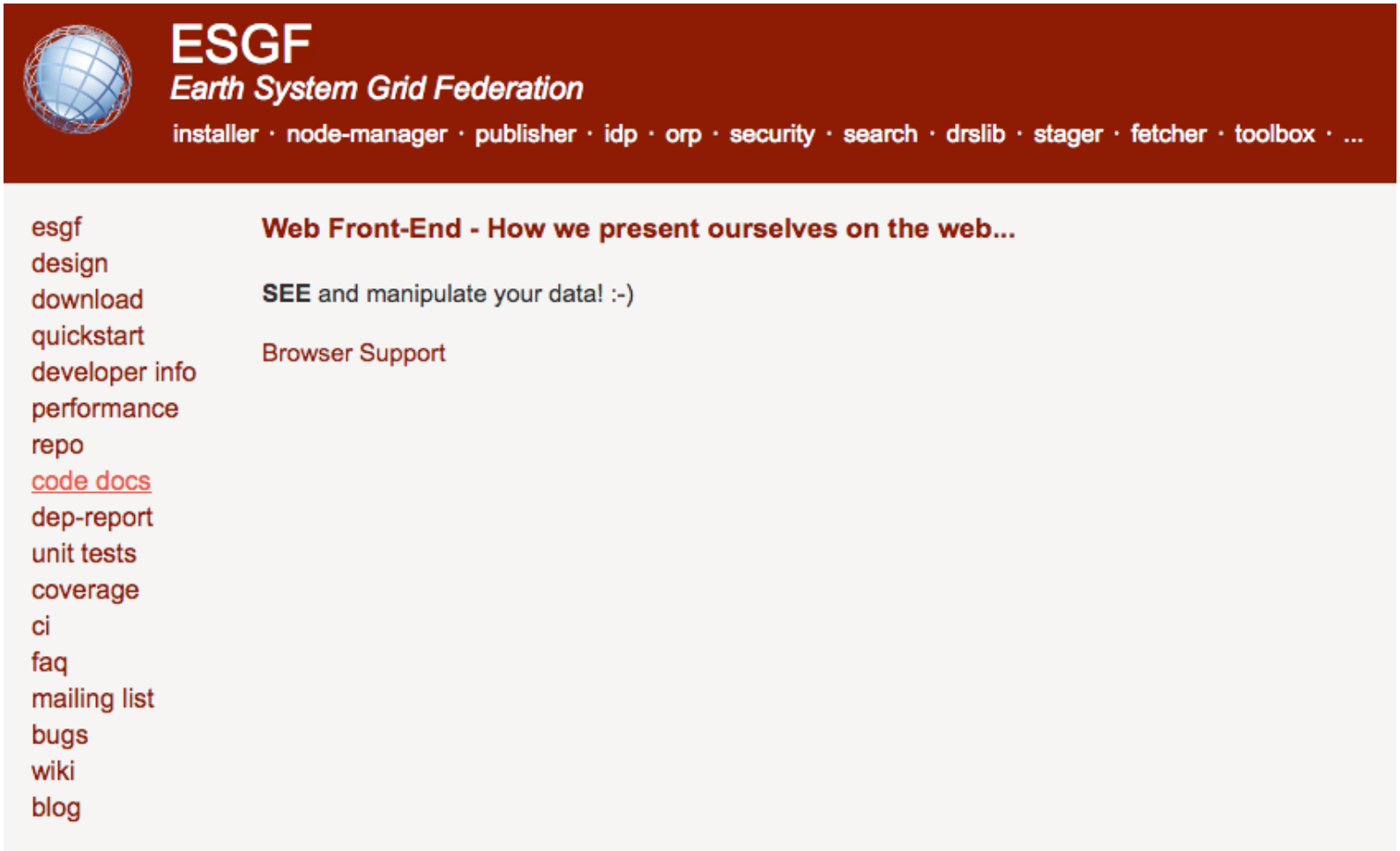
LLNL One Lab Bootstrap



Basic Bootstrap



LinkedIn tshirt



Experience

- Time savings
- Creating updated websites
- Dynamically created static webpages
- Hyde vs Jekyll
- Power of Markdown

Future Work

- Automated Testing
- Corner Cases / One off error handling
- Interactive webengine and command line options
- webshooter command line options
- Automation of style updating and changing

Conclusion

- The Webshooter repository started with one file in June 2013.
- I hope others will find it useful and even contribute to it.

Acknowledgments

I would like to thank Sam Fries for his time debugging, testing, and contributing to the code base. I would also like to recognize Ben Carlsson, Dean Williams for their support and contributions. Special thank you to Katie Walter, and Brian Harris for their editorial skills.

Questions?

http://github.com/webshootertk
http://github.com/mattben
harris112@llnl.gov