Hugo R&D for Innovation Development's AI/ML Initiative

Background

SIL's Innovation Development group (IDX) is beginning a two year initiative to develop a suite of tools to aid in the translation quality checking process for Bible translation. This phase of Bible translation tends to be a significant bottleneck in the overall process, so tools that can help measure specific things about a translation such as readability, similarity to other translations, etc. could be really beneficial to translators.

Part of the AI/ML process is the sharing of experiments, results, and knowledge. To do this the IDX team has requested that the IT Software Engineering (ITSE) team develop and/or host a website for this and assist with any automation around content management and access. This kind of website is often referred to as a "leaderboard". Examples of other AI/ML leaderboards: here, here, and here.

These leaderboard websites are primarily static in nature with occasional updates as new experiments or results are published.

Research & Development

Hugo - https://gohugo.io/

Hugo is a very popular "static site generator". It operates on files with content in formats like YAML/TOML and HTML along with layouts and templates to generate the full HTML, CSS, JS for a site. The process of generating the site is very fast and thus leveraging a static site generator like Hugo is often a better solution than a dynamic content management system like Wordpress. Hosting static websites is extremely cheap and far less prone to hacking since there is no backend code or database to compromise.

We'd like to research if using Hugo along with custom "content types" and/or "archetypes" would be a good solution for a leaderboard site. The idea being that we could easily create new files containing appropriate data about an experiment or test results and rebuild and redeploy the site. Ideally this could happen as part of automation when tests are executed in their automation environment that the results could be sent to a process that creates the content file, rebuilds the site, and deploys it.

Expected Site Layout

- Home
- Tasks (a list of the various tasks available for working on. May also be displayed on homepage)
  - Submissions (for each task a list of submissions)
- Resources
  - Source Control
  - Process Documentation
- About
  - High level information about this initiative, goals, funding, contacts, etc.

Content Types

Task

Property	Type	Notes
Title	Text	
Summary	Text (Markdown?)	
Getting Started Docs	Markdown	
About	Markdown	
Contact Name	Text	
Contact Email	Text	
Icon	Text	Relative path to an image of specific dimensions also hosted as part of site.
Tags	Text (List?)	Ideally support multiple tags to create tag based navigation on site as well.

Submission

Property	Type	Notes
Name	Text	
Date	Date	

Description	Text (Markdown?)	
Submitted By	Text	
Contributors	Text	
Score	Float	Is Score a singular value? Or do we need to support multiple scores for different components of the task? For example
Source Code URL	Text	