ICICLE AI Institute's Training Catalog: Using Open-Source Tools for Documentation

Introduction

ICICLE AI Institute's documentation hub, known as the Training Catalog, hosts detailed guides, tutorials, and references for our AI models, software releases, and datasets. It categorizes content using Tags aligned with our research thrusts, making navigation intuitive. Additionally, the site includes a powerful client-side search, allowing users to quickly find relevant content across the entire catalog.

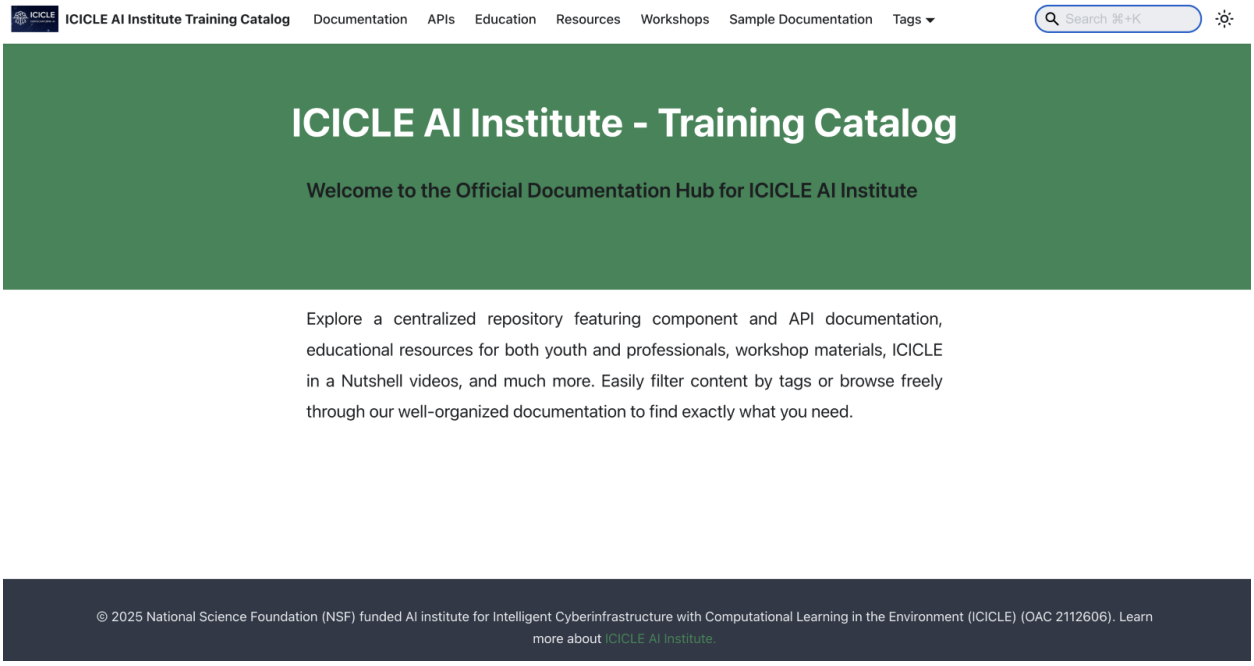

Figure 1. The landing page of the training catalog

The website was built with open-source tools like Docusaurus and Umami, and documentation is structured using the Diátaxis framework.

Building The Training Catalog

The Training Catalog is powered by Docusaurus, a popular open-source tool for creating documentation websites. Chosen for its simplicity and flexibility, Docusaurus allows writing content in Markdown files in a lightweight format, making it easy to maintain and update.

Structured Documentation with Diátaxis

Diátaxis categorizes technical documentation into clear, user-focused segments: Tutorials, How-to guides, Explanations, and References. This structured approach greatly improves clarity and navigation, helping users quickly locate relevant information.

To systematically structure our documentation, we implemented a custom Python parser based on the Diátaxis framework.

Our custom Python parser utilizes README files released by individual projects, organizing them systematically into Markdown (.md) files and folders as required by Docusaurus. This automated process significantly streamlines documentation updates and ensures consistent structuring aligned with the Diátaxis framework.

Education Material

The ICICLE AI Institute is committed to advancing AI and Cyberinfrastructure (CI) literacy through comprehensive education initiatives spanning K–13 and professional development. One of the institute's goal is to train the next generation of researchers, educators, and practitioners by fostering a strong foundation in responsible AI/CI practices. The Education section features resources that promote awareness of human and AI biases, emphasize ethical considerations, and support the development of critical skills in youth and professionals. We also use educational tags like youth, video, etc. for easy filtering.

Search and Thrust-Based Tags

The training catalog organizes software components using thrust-based tags aligned with ICICLE AI's core research areas. Each released component is assigned at least one relevant thrust tag, making it easier for users to filter, and discover components based on specific research focuses.

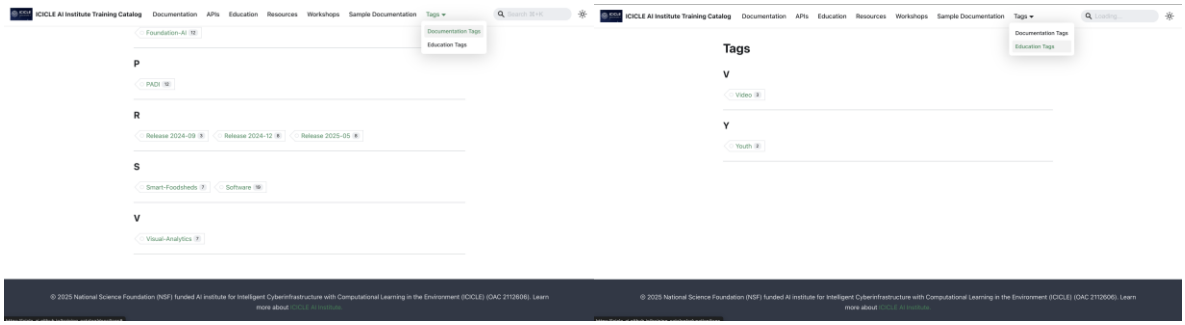

Figure 2. Tags page showing various thrust-based and educational tags.

To enhance discoverability, the site integrates full-text, client-side search allowing users to instantly search across all documentation pages. The search bar supports keyword-based lookups, highlighting matches within titles, and content.

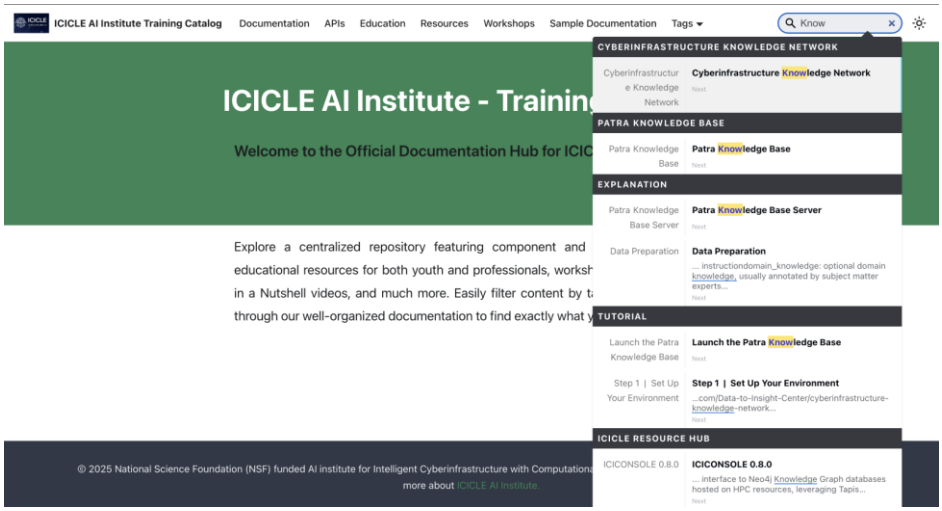

Figure 3. Search functionality

Analytics with Umami

To monitor user engagement and optimize documentation effectiveness, we utilize Umami, an open-source, privacy-focused analytics solution.

Umami delivers insightful analytics on documentation usage without compromising user privacy. Hosted securely on TACC infrastructure using Tapis, Umami provides essential data on page views and user interactions, allowing us to continuously enhance our documentation.

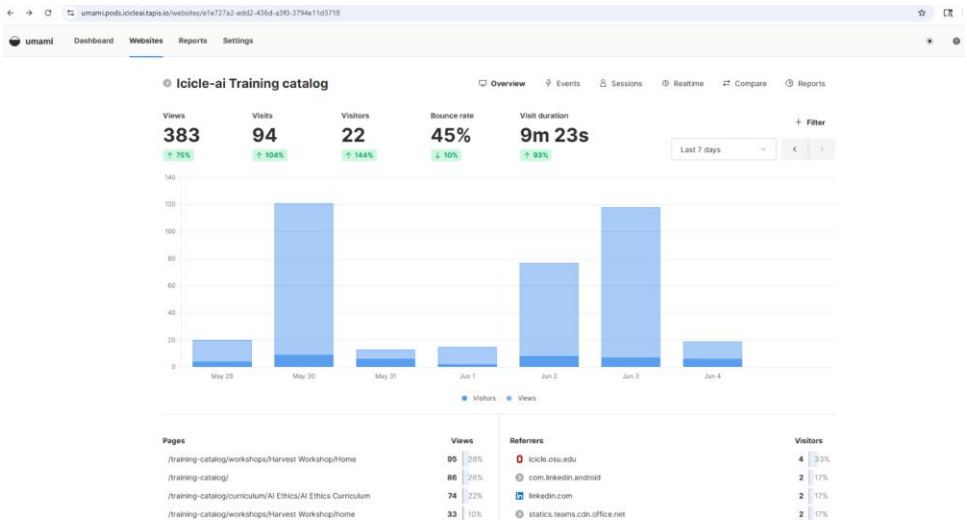

Figure 4. Umami Analytics dashboard

Hosting the Training Catalog

We host the Training Catalog on GitHub, using GitHub Actions to automatically publish updates whenever we make changes. This ensures our documentation remains updated, reliable, and consistently accessible, requiring minimal manual intervention.

Training Catalog Workflow Architecture

The training catalog's workflow architecture is illustrated in Figure 5.

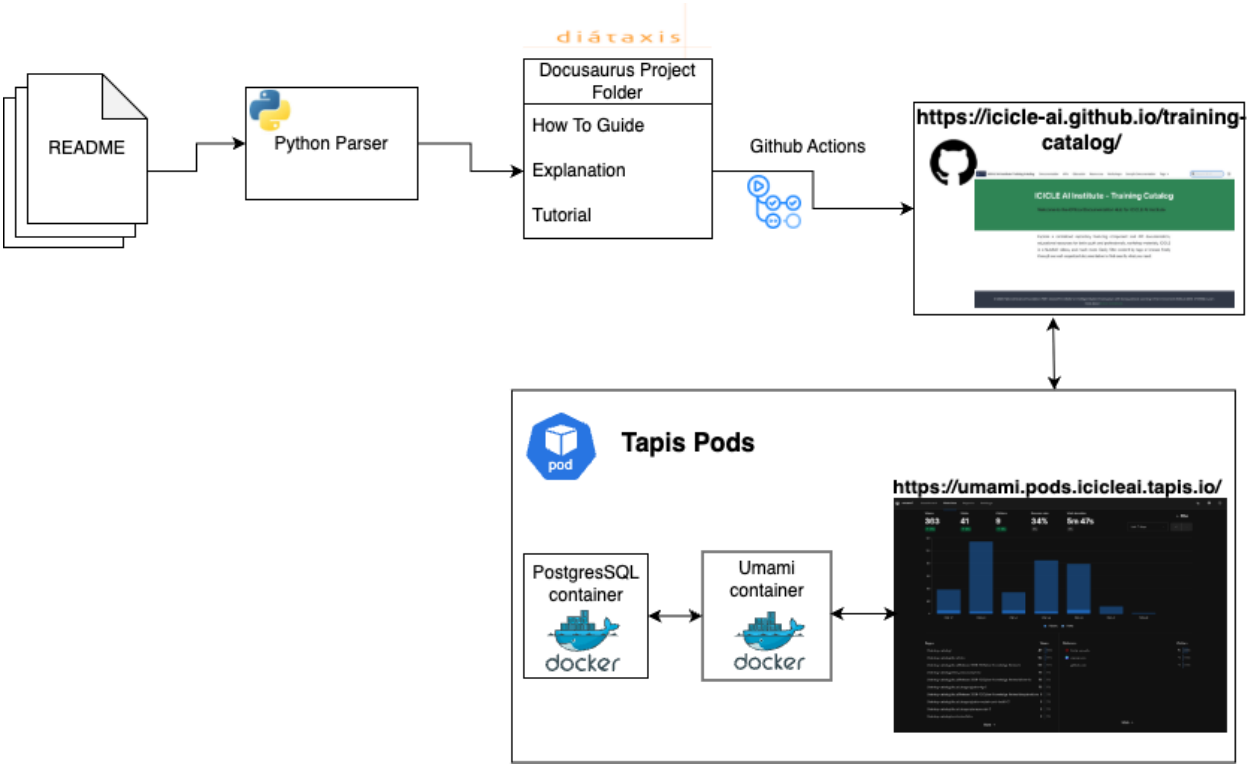

Figure 5. Training Catalog Workflow Architecture Diagram.

The README files from our AI models, Datasets and software components releases, are structured following the Diátaxis documentation standard, and serve as the input to a custom Python parser. This parser automatically processes the content and generates well-structured Markdown (.md) files within appropriate project folders. These files are formatted to take full advantage of Docusaurus's dynamic features, such as documentation cards and auto-generated sidebars.

We also configure and integrate Umami, a privacy-focused analytics tool hosted on TACC infrastructure via Tapis, into our Docusaurus-powered Training Catalog. This integration provides actionable insights into user behavior, such as page views and interaction patterns, allowing us to continuously refine and enhance our documentation based on real usage data.

Conclusion

Our integrated approach—combining Docusaurus, Umami analytics, GitHub Actions automation, and the Diátaxis methodology delivers robust, user-centric documentation. This framework ensures the effective dissemination and practical application of ICICLE AI Institute's innovations.

Visit the Training Catalog to explore more.