Potato: the POrtable Text Annotation TOol

TECHNOLOGY NUMBER: 2023-442

Easy to deploy

- 10+ templates and schemas
- Easy multilingual/multi-task annotation
- Fit for both internal annotation and crowdsourcing
- Editable UI

Improving productivity

- Active learning
- Conditional highlighting
- Keyboard shortcut input
- Label tooltips

Better quality control

- Built-in attention test
- Qualification test
- Pre- and post-screening questions
- Built-in time tracker

High accessibility

- Free for everyone
- Fully open-sourcing
- Easy sharing annotation logs and configurations

Technology ID
2023-442

Category
MOSS - Michigan Open Source Software

Inventor
David Jurgens
Jackson Sargent
Naitian Zhou
Xingyao Wang
Aparna Ananthasubramaniam
Jiaxin Pei

Learn more

OVERVIEW

User-friendly web-based tool for streamlined text annotation tasks

- Simplifies annotation setup with minimal configuration, no coding needed
- Used for data labeling, linguistic research, machine learning, team-based text annotation

BACKGROUND

Data annotation is crucial for machine learning model training, with a history rooted in manual labeling and traditional software. Previously, data labeling was often arduous and time-consuming, relying heavily on rudimentary tools or bespoke scripts. As machine learning exploded, the need for rapid, scalable annotation solutions became more pressing. Previous methodologies were limited by inflexible user interfaces and data privacy concerns, necessitating data sharing with third-party platforms. This posed challenges in terms of data security, especially when dealing with sensitive or proprietary datasets. Additionally, the rigidity of existing tools imposed restrictions, obstructing the tailoring needed for distinct labeling tasks. Thus, the demand surged for a more adaptable annotation tool that could efficiently harness crowd-based resources while retaining data control and offering customizable user experiences.

INNOVATION

Potato is an intuitive web-based annotation tool that addresses the challenges of traditional text annotation by requiring minimal configuration and no coding expertise. It operates as a locally hosted web server, with a front-end accessible via a web browser, allowing quick setup and easy access for annotators. Driven by a single configuration file, Potato enables users to specify tasks and data types without delving into complex coding or web design. While it provides a default interface to get users started immediately, Potato also offers customization options for tailoring the user experience. This innovation significantly reduces the time and technical knowledge required to initiate annotation tasks, making it suitable for both individual annotators and small teams. Real-world applications include data labeling for machine learning models, linguistic research annotation, and efficient collaborative text annotation projects. Potato democratizes the annotation process, promoting broader participation and expedited project workflows.

ADDITIONAL INFORMATION

PROJECT LINKS:

- [Potato Code Repository](#)
- [Potato Documentation](#)

DEPARTMENT/LAB:

- [David Jurgens, School of Information/Electrical Engineering and Computer Science (EECS)](#)

LICENSE:

- Polyform-Shield