

LADA Open Discovery Toolkit

LDD Genie

Yang Li (yli546@usc.edu)

Yizhou Sheng (yizhoush@usc.edu)

Yulun Zhang (yulunzha@usc.edu)

Jincheng Zhou (jinchenz@usc.edu)

Changyu Zhu (changyuz@usc.edu)





Logical Document Determination (LDD)

LDD is the process of establishing the boundary of a document by identifying logical breaks.

WHAT'S THE PROBLEM?



PROBLEM

It requires a lot of manual labor to identify and assign logical boundaries, as well as fill in semantic labels.

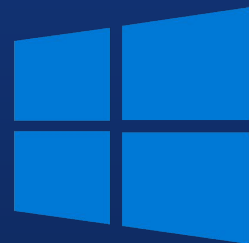


SOLUTION

LDD Genie: The next generation document management software, designed for LDD.



LDD Genie



A standalone Windows® application specialized in document boundary & semantic label management

*Windows® and the Windows logo are registered trademarks of Microsoft Corporation

WHY LDD GENIE?



COMPATIBILITY

Compatible with the concordance archives used by Nuix Workstation



EASY & FAST

Create boundary type & semantic label presets and use hotkeys to assign pages at lightning speed



ARTIFICIAL INTELLIGENCE

Automatically assigns boundaries. Learn from your corrections and evolve over time



Configure Concordance Header

Make sure these fields match your load file configuration.

DOCID

DOCID

PARENT_DOCID

PARENT_DOCID

BEGINBATES

BEGINBATES

ENDBATES

ENDBATES

BEGINGROUP

BEGINGROUP

CANCEL UPLOAD

Getting Started

Create new project

+ Import pdf file

+ Import concordance archive

START WITH PLAIN PDF OR CONCORDANCE ARCHIVE

A PDF is all we need to get started

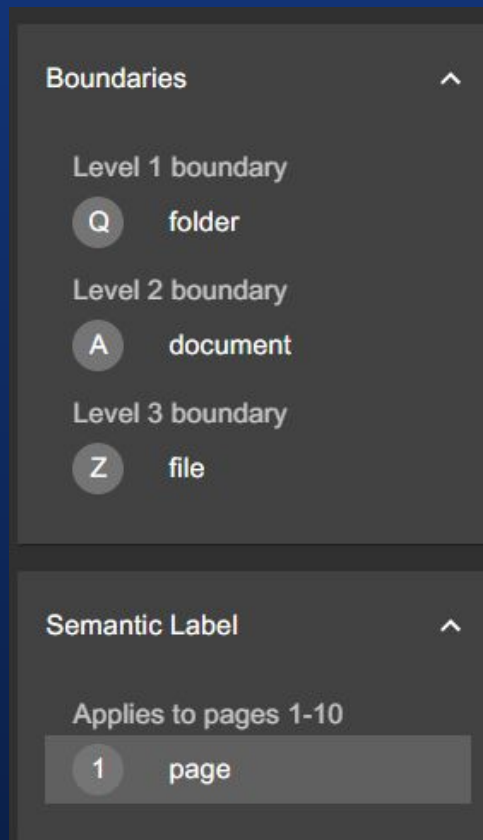
Or import existing archives easily with flexible configuration



HOTKEYS

No more clicking around

Finish the work with the power of keyboard





Configuration

VISUAL	JSON
Semantic labels	+
page x	
Level 1 boundary	+
folder x	
Level 2 boundary	+
document x	

CANCEL SAVE

Create presets of boundary names and semantic labels to boost your productivity.

Quickly import a template with JSON.

CUSTOM PROJECT CONFIGURATION





GALLERY VIEW

View all pages in a glance

Identify different document bundles by background colors

The screenshot displays a gallery view of a document. At the top, it shows 'PAGE VIEW < 1 / 1 >' and 'Items per page 10'. The main area contains a grid of document pages, numbered 1 through 6. Page 2 is highlighted with a yellow border. A tooltip over page 2 shows 'Start of document', 'file', 'Semantic Label: page', and 'Single page 2'. The right sidebar includes 'Project #1', 'Project name Sample Project', and a list of actions: 'Config', 'Export', 'Save', 'Predict Boundaries', and 'Back'. Below these are 'Boundaries' with 'Level 1 boundary folder', 'Level 2 boundary document', and 'Level 3 boundary file'.





MULTI-SELECT

Assign labels
and boundaries
in batch

PAGE VIEW < 1 / 1 > Items per page 10

Project #1

Boundaries

- Level 1 boundary folder
- Level 2 boundary document
- Level 3 boundary file

Semantic Label

Applies to pages 2, 3-4, 6-10

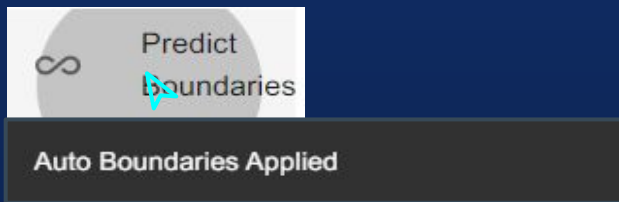
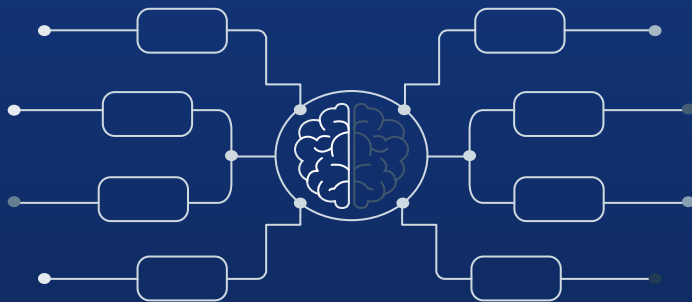
- 1 page
- 2 invoice





Meet your new assistant...

ARTIFICIAL INTELLIGENCE



Detect boundaries with one click

```
\ml_models>python train_model.py|
```

Evolve over time at your command





MATERIAL DESIGN

Material is an adaptable system of guidelines, components, and tools that support the best practices of user interface design

The screenshot displays a document viewer interface. At the top, it shows 'PAGE VIEW < 1 / 1 >' and 'Items per page 10'. The main area contains a grid of document pages, with page 2 highlighted. A sidebar on the right is titled 'Project #1' and includes 'Project name Sample Project' and a list of actions: 'Config', 'Export', 'Save', 'Predict Boundaries', and 'Back'. A metadata overlay on page 2 shows: 'Start of document file', 'Semantic Label: page', and 'Single page 2'.



DARK / LIGHT THEME

Configuration

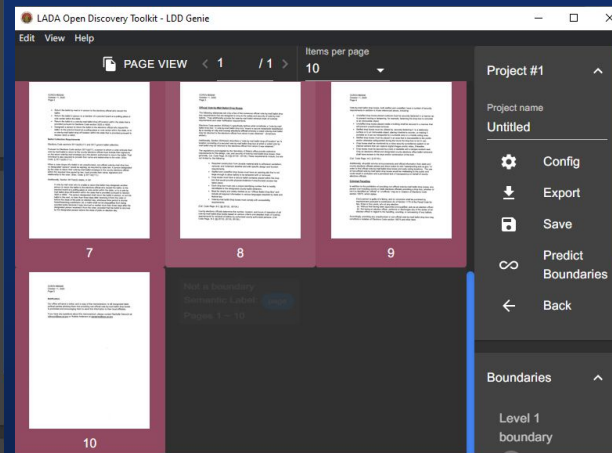
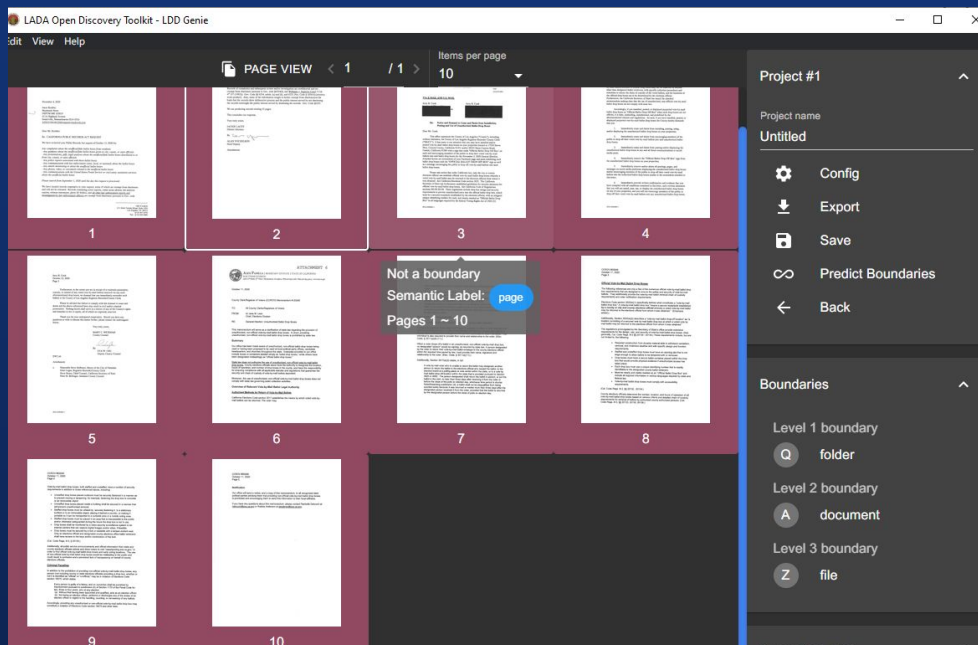


The following are user-specific settings that only affect your workspace.
Changes take effect immediately.

Theme



RESPONSIVE DESIGN



Automatically adapts to the window size

Zoom in / Zoom out support





TECHNOLOGY AND FRAMEWORKS





INDUSTRY LEADING TECHNOLOGY

NestJS: Reliable server



React: Component-based stateful views

Tensorflow: Comprehensive, flexible platform for machine learning



Electron: Build cross-platform desktop apps with web technology



PORTABLE SOFTWARE



NO INSTALLATION

We won't modify anything outside the working directory, nor change your PC settings/registry.

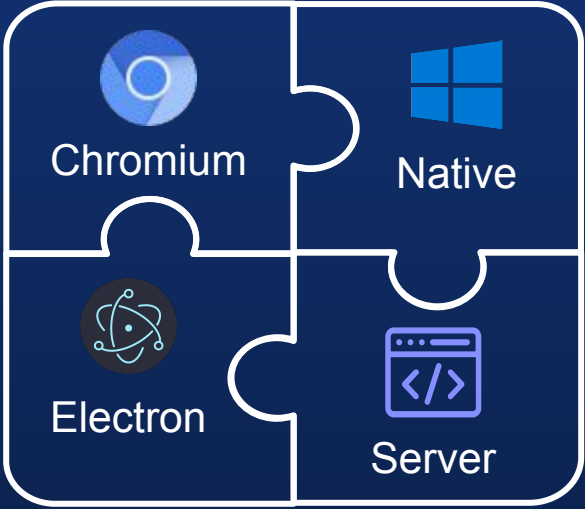


UNZIP AND USE

All dependencies are bundled.
Uninstallation is as simple as deleting the folder.



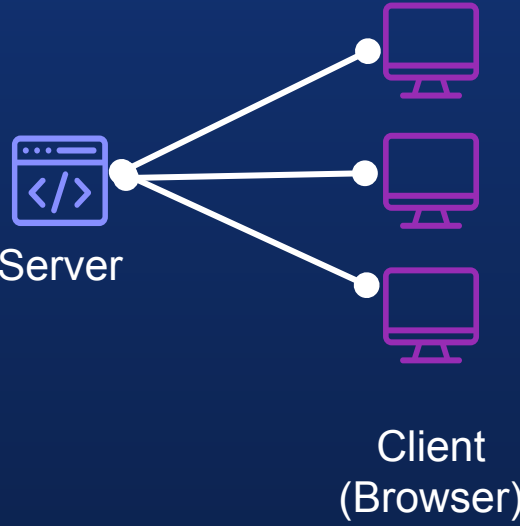
WEB APP AS STANDALONE APP



Native Windows App



This software is designed to be smoothly migrated to a web service with minimal code changes



Dedicated Server For Web App





DEMO





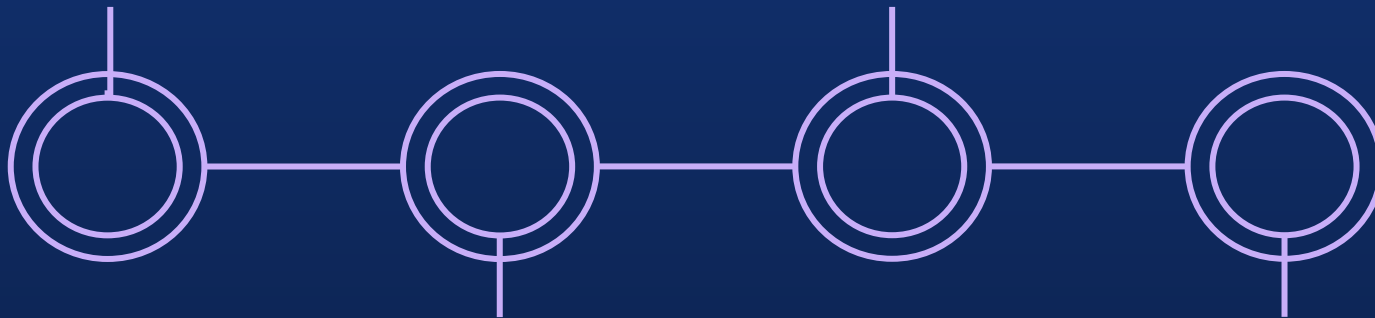
PLAN MOVING
FORWARD





Current release


Deploy as web
app




User access
control system

Online ML model
training with new
data





Q&A





THANK YOU

