

A set of stories

Team 3-CITS3200 Professional Computing-Documentation Identification Software

By Rania Khan, Hazel Wang, James Felstead, Peter Fang, Chunyu Zheng

Overview:

Sprint 2 focuses on delivering a functional prototype capable of:

1. Accepting metadata of census documents from user
2. Performing AI-assisted searches across multiple sources to locate matching documents.
3. Retrieving accessible PDF files.
4. Maintaining a basic search history for reference.

This sprint will provide the core functionality for the document sourcing and retrieval workflow, ensuring core components are in place for subsequent enhancements in Sprint 3.

Stories:

As a researcher	I want to	So that I
Researcher	Input census document metadata into the system	Initiate precise searches with accurate data
Researcher	Receive AI-assisted search results with ranked potential matches	Efficiently identify the most relevant source documents
Researcher	Retrieve PDF files of matched documents when available	Access full content with verification and analysis
Researcher	View a basic history of past searches	Revisit previous results without repeating the search process

Story Breakdown:

Story1: Metadata Input & Validation

Implement the user interface for entering census document metadata and validating inputs before passing them to the search module.

- Metadata form Country, Province, Year, Author, Agency, Title (Subtitle)
- Validation rules (Year between 1860–2025, required fields)
- Data packaging into JSON and passing to search module

Story2: AI-Assisted Search & Matching

Integrate AI API calls and web crawling to search multiple sources, return ranked matches, and remove duplicates.

- Integrate search algorithm with AI-assisted ranking.

- Implement connection to external document repositories.
- Display ranked search results with confidence scores.

Story3: PDF Retrieval, Storage & History

Download matched PDFs, store them in cloud storage with correct naming conventions, and record basic search history. If no PDF is found, provide a list of accessible libraries.


































- Automate retrieval of available PDF files from identified sources.
- Store retrieved files in a secure and accessible location.
- Implement basic search history logging and retrieval.

Rationale

This set of stories was chosen due to it delivering the essential functionality required for a usable prototype by the end of Sprint 2. Breaking down into user-centric stories could make sure that each completed story provides tangible value to end users and moves the project toward its final goal.

Jira Template

In the current development plan, Jira serves as the central platform for coordinating work. During execution, individual sub-tasks may generate additional issues, such as new sub-tasks or bug reports. This flexible workflow allows the team to address challenges, refine deliverables, and keep the project on track.

Type	Key	Summary	Status	Comments	Sprint
	CDI-7	Package data to JSON	TO DO	 Add comment	Sprint 2
	CDI-8	Send JSON to backend	TO DO	 Add comment	Sprint 2
▼ <input checked="" type="checkbox"/>	CDI-2	AI-Assisted Search & Matching	TO DO	 Add comment	Sprint 2
	CDI-9	Integrate AI API	TO DO	 Add comment	Sprint 2
	CDI-10	Implement prompt tuning	TO DO	 Add comment	Sprint 2
	CDI-11	Connect to multiple sources	TO DO	 Add comment	Sprint 2
	CDI-12	Add ranking & deduplication	TO DO	 Add comment	Sprint 2
	CDI-13	Show results with scores	TO DO	 Add comment	Sprint 2
▼ <input checked="" type="checkbox"/>	CDI-4	PDF Retrieval, Storage & History	TO DO	 Add comment	Sprint 2
	CDI-14	Automate PDF download	TO DO	 Add comment	Sprint 2
	CDI-15	Save to cloud storage	TO DO	 Add comment	Sprint 2
	CDI-16	Create search history DB table	TO DO	 Add comment	Sprint 2
	CDI-17	Display history in UI	TO DO	 Add comment	Sprint 2
▼ <input checked="" type="checkbox"/>	CDI-3	Technical Setup & Resources	TO DO	 Add comment	Sprint 2
	CDI-18	Configure AI API keys	TO DO	 Add comment	Sprint 2
	CDI-19	Set up cloud storage	TO DO	 Add comment	Sprint 2
	CDI-20	Install dev frameworks	TO DO	 Add comment	Sprint 2
	CDI-21	Review API docs	TO DO	 Add comment	Sprint 2

Intermediate Acceptance Criteria

Metadata Input & Validation

- Acceptance Criterion 1: All form fields functional and validated.
- Acceptance Criterion 2: Data correctly passed to search module.

AI-Assisted Search & Matching

- Acceptance Criterion 1: Returns results from at least two data sources.
- Acceptance Criterion 2: Removes duplicates and sorts results by relevance.

PDF Retrieval, Storage & Basic History

- Acceptance Criterion 1: Download success rate $\geq 90\%$ for valid sources.
- Acceptance Criterion 2: Files named and stored correctly; basic history visible.