Data Engineering · Project Proposal

Advice and Paper Recommendations from Experience Paths

Team 13

San ghun Kang

Se hyeok An

Jin Oh Kim

Seung beom Nam

2026.10.20

Contents

01	**Background and Problem Definition**	Background · Example · Problem · Goal
02	**User Scenario**	Overview · Question · Suggestion · Record
03	**Data Pipeline**	Collection, Recording, and Delivery
04	**Database Design**	Experience Path DB · Paper DB · DB Example

Appendix · Related Work · References

Background

Papers and research experience accumulate, but their application context is hard to connect

Published Research

Papers and Technical Documents

- Proposed Methods
- Experimental Results
- Limitations and Conditions

Practical Experience

Applying Methods in Practice

- Reasons for Choosing a Method
- Attempts, Changes, and Decisions
- **Problems Encountered Next**

Key **The challenge is connecting paper knowledge, experience, and context**

Example

What if we could learn from someone who made the same choice before?

Researcher A · Real Case (Team Member)

High Noise
A · Initial Problem

→

Filter Applied
B · Method Tried (Noise Reduced)

→

About 10 ms of Delay
C · Next Problem (1–2 Months to Find Cause)

Noise fell, but later experiments kept giving poor results. It took 1–2 months to identify the filter delay as the cause.

Researcher B · Current Situation

Sensor Noise Problem → **Preparing an Experiment After Filtering** → **“Should I check the filter delay before the experiment?”**

Goal **Use past experience to suggest checks now, without assuming a future failure**

Problem Definition

Researchers struggle to reuse the experience behind others' decisions

01

Fragmented Experience

Attempts and decisions beyond papers are scattered in personal notes and chats

Collection and Sharing

02

Disconnected Steps

The order of attempts and reasons for moving to the next step are unclear

Data Representation

03

Limited Reuse

Past follow-up problems and conditions are hard to connect to a current choice

Retrieval and Use

Core Task **A data structure that preserves the order and context of experience for use in similar research situations**

Goal

Use other researchers' experience to inform the next decision

Two Types of Data

Paper DB

Paper metadata: titles, keywords, and public links

Linked Papers

Experience Path DB

Who tried what, and what happened afterward

Using Practical Experience

Current Research Situation

A → B

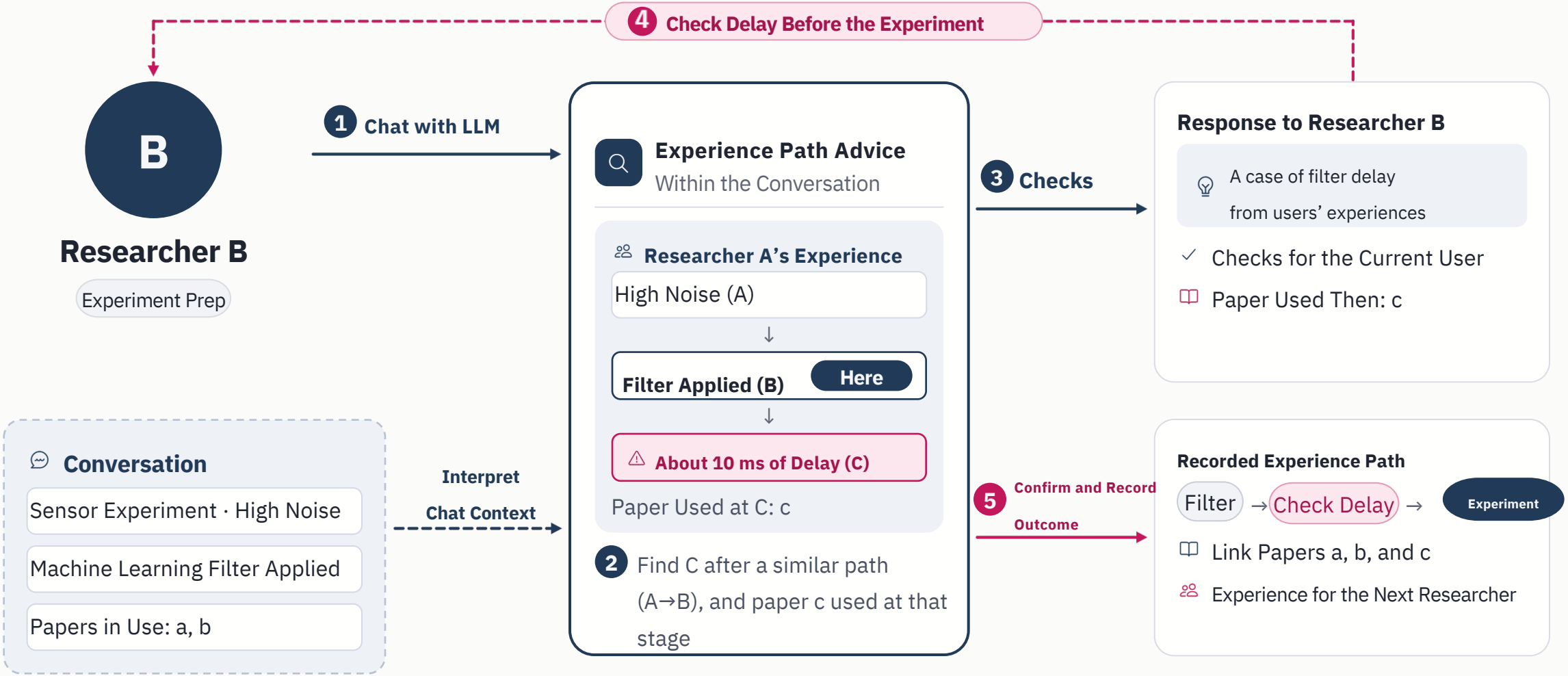
Past Research Experience

A → B → C

Checks for C and Related References

Scenario Overview

From one question to an experience that helps the next researcher



‹ Researcher B's scenario: preparing the next experiment after applying a filter ›

02 User Scenario

Scenario 1 · Question

Researcher B simply asks a question as usual

Thinking

“There is a lot of noise. A filter should fix it, right?”

Researcher B

Sensor Experiment

Preparing the Next Experiment After Filtering

Research Assistant · Usual LLM Chat

Conversation, No Search Box

I am collecting experimental data with a sensor, but there is too much noise.

High noise can make it hard to distinguish real changes in the measurements.

Researcher B · Today

I read papers a and b and applied a machine learning filter. The noise clearly decreased. I want to use this filter in my next experiment. What should I check first?

Context Identified from the Conversation

Sensor Experiment | High Noise (A) | Filter Applied (B) | Papers Used: a, b

Scenario 2 · Suggestion

Find checks in the experience of someone who made the same choice before

Research Assistant · Response

Question: Pre-Experiment Checks

Experience of a Researcher Who Made the Same Choice

Researcher A reduced sensor noise with a machine learning filter, but only discovered its roughly 10 ms delay one or two months later.

What to Check Now

1. How much delay can your experiment tolerate?
2. Have you measured the signal delay before and after filtering?
3. Have you compared filter settings with less delay?

\+ ... Other experiences may suggest additional checks

Paper Researcher A Used at C

Paper c

How to measure and reduce signal delay introduced by filtering

Scenario 3 · Record

Report the outcome, and your experience can help the next researcher

After Use · Follow-Up

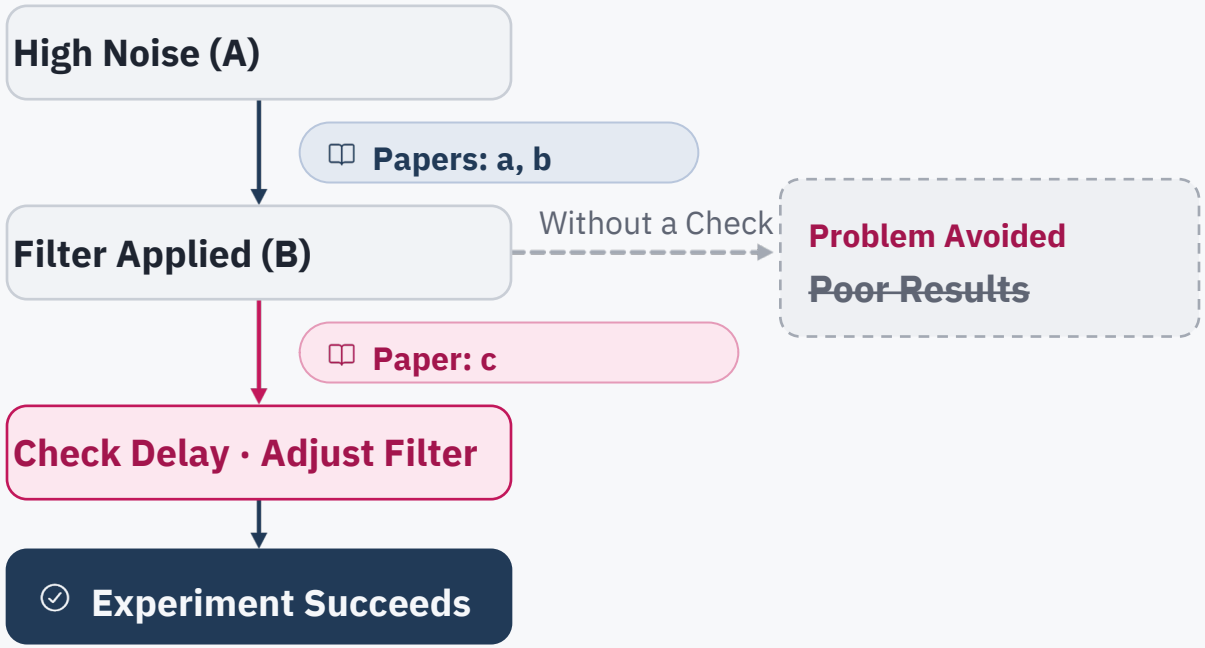
Researcher B

I checked the delay before the experiment. I adjusted the filter settings to meet its tolerance, and the results were stable.

Please confirm the summary. I will then save it under your sharing settings.

Confirmed Experience Path

Experience for the Next Researcher

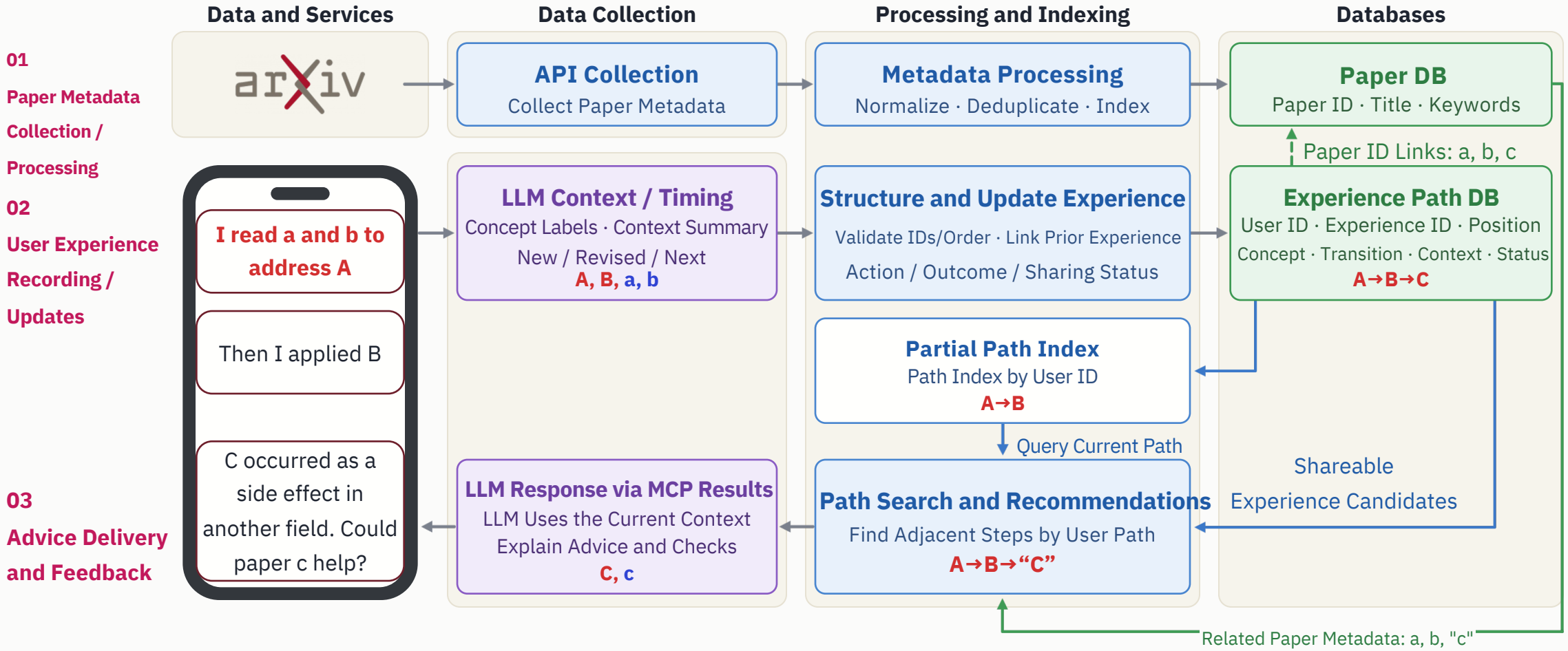


This experience and paper c can help the next researcher following the same path (A→B).

03 Data Pipeline

Collection, Recording, and Delivery

Three Flows: Paper Collection · Experience Recording and Updates · Advice and Feedback



Notation: Uppercase = concepts, lowercase = papers · A high noise → B filter applied → C delay (~10 ms) · c = paper used at C
Rows: 1 paper collection · 2 user context recording · 3 advice and feedback · Follow-up messages return to row 2, distinguished by timing

Experience Path DB

Experience Info

Column	Description
Experience ID	Identifies the Experience
User ID	Author
Title	Experience Title
Field	Research Field
Sharing Status	Sharing Scope

Experience Steps

Column	Description
Step ID	Identifies the Step
Experience ID	Parent Experience
Position	Step Order
Concept ID	Shared Concept Label
Context	Situation / Attempt
Role	Step Role
Action Status	Whether Attempted
Outcome	Attempt Outcome
Conditions	Conditions at That Stage

Paper Links

Column	Description
Link ID	Identifies the Link
Start Step ID	Reference Start
End Step ID	Reference End
Paper ID	Referenced Paper
Purpose	Reference Purpose

Storage Candidates: PostgreSQL (Service) · SQLite (Initial Testing)

Collection and

User Conversation
Questions · Attempts · Outcomes

→

Structure Experience
Organize Steps by Experience

→

Store Experience
PostgreSQL / SQLite

→

Retrieve Next Steps
Next Steps · Referenced Papers

Query Example: In E01, retrieve C (delay) and its paper after A (high noise) → B (filter applied)

04 Database Design

Paper DB

Paper Info

Column	Description
Paper ID	Identifies the Paper
Source ID	Source Paper ID (arXiv ID)
Version	Paper Version
Title	Paper Title
Abstract	Paper Summary
Authors	Author List
Category	Source Category

Column	Description
Year	Publication Year
Public Link	Paper / Full-Text Link
Keywords	Search Terms from Category / Abstract
Verification Level	Metadata / Full-Text Checked
Source	Paper Source
Collection Time	Time Collected

Proposed setup: separate experience and paper data in one store, linked by Paper ID

Collection and

Storage: PostgreSQL (Service) · SQLite (Initial Testing)



Link Example: An experience references Paper ID P01 to retrieve its title and public link

DB Example

Experience	Field / Conditions	A · Excess Information (e.g., Noise)	B · Filtering (e.g., Applying a Filter)	Next Recorded Step (Own Concept Label)	Condition Match
Now · Researcher B	Sensor Experiment Real-Time Response	High Noise	Filter Applied	? Searching	—
E01 · Researcher A	Sensor Experiment Real-Time Response	High Noise	Filter Applied	Response Delayed ~10 ms Paper c	Match → Prioritize
E03	Demand Forecasting Fast Response Needed	Fluctuating Order Volume	Smoothing by Averaging	Demand Changes Reflected Late Paper c'	Match → Prioritize
E05	Image Analysis	Many Image Specks	Blurring	Fine Detail Lost	Different → Lower

1 **Find the Same Concept Path**
A Excess Information → B Filtering

→

2 **Compare Conditions**
Real-Time Response Important?

→

3 **Next Step of That Experience and Its Referenced Paper**

Returned to Researcher B

“After filtering, how much response delay can your experiment tolerate?”

Paper c (Sensors) · Paper c' (Demand Forecasting)

Thank You

Q&A

Appendix · Related Work · References

Appendix · Related Work

Prior Work / Technology	What It Does	Relevance to Our Project
Process-Oriented Case-Based Reasoning POCBR / ProCAKE	Stores past work and problem-solving processes as directed graphs with semantic labels, then finds similar processes or substructures for reuse and adaptation	**Closest to our core idea: storing experience as paths and comparing them with the current context** ProCAKE, 2019
Derivational Analogy Derivational Analogy / PRODIGY	Records the decisions and reasons leading to a solution, as well as the solution itself, then reexamines and reuses that process for a new problem	Relevant to recording why someone moved from A to B and which decisions failed Veloso · Carbonell, 1993
Relational Analogy Structure-Mapping Engine (SME)	Compares relational structures across two cases rather than surface names, producing mappings and candidate inferences	Links experiences with similar relations across fields, such as sensor experiments and demand forecasting Northwestern Research Team Overview
Meta-Path Meta-Path-Based Retrieval	Defines an ordered sequence of node types and relations as a meaningful composite relation for retrieval and similarity calculation	Provides a formal basis for giving meaning to A→B→C paths. Additional information is needed to represent decision reasons and side effects. PathSim, 2011

Our Project's Position Drawing on process storage, decision reasons, relational structure, and path meaning, we propose a data structure and pipeline linking researchers' trial-and-error paths to the papers they used

Appendix · References

Appropriate Uses For SQLite

https://www.sqlite.org/whentouse.html

Storage choice: SQLite for initial testing. Consider a client/server engine for a multi-user service.

PostgreSQL Documentation — Schemas

https://www.postgresql.org/docs/current/ddl-schemas.html

Basis for separating experience and paper data into different schemas within one database

PostgreSQL Documentation — JSON Types

https://www.postgresql.org/docs/current/datatype-json.html

Basis for storing conditions as JSONB