



InterPlanetary Archival Record Object for Decentralized Web Archiving and Replay

Team Members: Benji Bui, Alexander Grigorian, Alexey Kuraev, Patrick Le, Thiyazan Qaissi

Project Background

- Internet Archival:
- Maintain records of web content over time.
 - Websites can vary over time. Can archive versions tied to a date. (Versioning)
 - Stored web content as archive files that maintain the essence of the original content.
 - Internet archival services such as *Internet Archive* are the libraries of the internet.

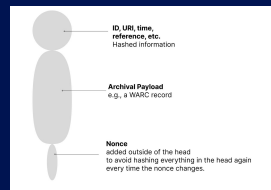
- Internet Archive*:
- Inefficient data storage and retrieval.
 - Prone to cybersecurity attacks.
 - Mass loss of data can be equivalent to the burning of the Library of Alexandria.

- Issues to be tackled:
- Brought to light by Dr. Alam and Dr. Kelly.
 - Lack of better native support for versioning.
 - Centralized databases limit users to institutions and organizations.

Introduction

InterPlanetary Archival Record Object (IPARO) system:

- IPARO can be thought of as nodes.
- Decentralized solution via InterPlanetary File System (IPFS)
- InterPlanetary Name System (IPNS) provides native versioning support.



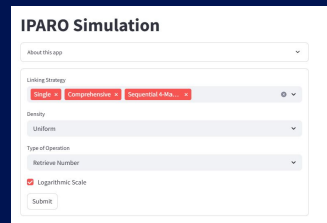
IPARO Node

- Mission:
- Create a website that uses the IPARO system
 - Showcase the ability for entities of any size to adopt the system.
 - Simulate the implementation to better define linking strategies for differing database environments.

Simulation

The IPARO Simulation allows for the curation of environments for IPARO linking strategy testing.

Simulation Input



Environment Fields

- Linking Strategies:
- Sequential
 - Temporal
- Density:
- Volume Distribution
- Type of Operation:
- Counts of specific operations done by IPARO system. (ex. Links stored)

Simulation Output

Results

Linking Strategy	Density	IPARO Archive Count									
		Count	Mean	Std	Min	25%	50%	75%	Max		
Comprehensive	Uniform	100	1.88	0.2056	1	2	2	2	2	3	
Comprehensive	Volume	100	100	2	0	2	2	2	2	2	
Comprehensive	Logarithmic	1000	100	2	0	2	2	2	2	2	
Sequential	Uniform	100	3.42	0.2242	3	3.75	3	4.25	5		
Sequential	Volume	100	3.4	0.1807	2	2	3	4.25	5		
Sequential	Logarithmic	1000	3.46	0.0994	2	2.75	3.5	4	5		
Single	Uniform	100	3.45	0.1094	2	3	3	4.25	5		
Single	Volume	1000	5.04	0.9568	1	3.75	4	13.5	18		

Numeric Results



Graphical Representation

- Operation count per linking strategy and volume density returns:
- Count, mean, and standard deviation.
 - Minimum and maximum possible operations.
 - Average operations in quartiles.

Visual representation of Numeric Results.

Logarithmic scale aids in better visibility of data on the Operation Count axis (Y-Axis)

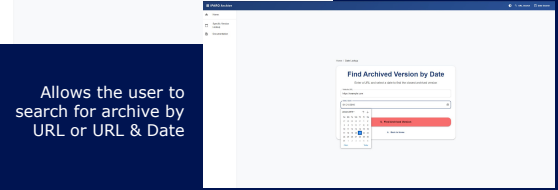
Website Implementation

The IPARO Simulation allows for the curation of environments for IPARO linking strategy testing.

Front Page

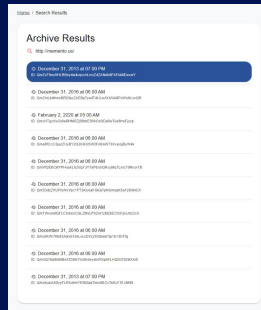


Integrated web page frontend to an IPARO System backend.

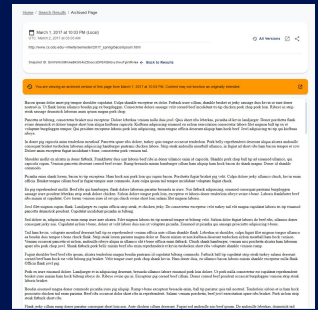


Allows the user to search for archive by URL or URL & Date

Example Search for Archive via URL



Example list of archives versions for a URL



Archive displayed through the Reactive API