User Interface development (Project 1)

In this project we would use Islandora, a collaborative open source framework to manage digitised assets and develop it to showcase digitised content from Noolaham in image, pdf and text formats. The UI will also show metadata and keywords for each document from manual annotations done by annotators. We aim to do keyword labelling via an automated machine learning model for the whole document collection which would be later displayed in the UI.

Detailed Proposal:

Metadata, Keywords and Auto labelling shall be stored in MongoDB in JSON format for each document. Islandora will need to look up this data from MongoDB to display in the UI or Data from MongoDB needs to be migrated to Islandora.

Note: All the existing document details available in the Noolaham needs to be migrated to MongoDB via CSV before starting auto labelling

Islandora Data Structure

Content Type :*Each article/document type shall be called as Content type. Set of metadata fields will be associated with each content type.*

Metadata Fields : *The fields may be associated with content types, and it may have many to many relationships with content types. Fields hold different types like Boolean, text, etc.*

Taxonomy vocabularies *: Taxonomy vocabularies can be created with multiple Taxonomy terms (fields). A Taxonomy vocabulary shall be referenced with one or many content types.*

Add/Edit/Delete options : *Contents, Fields / Taxonomy can be managed by Islandora administration Portal*

MongoDB Structure

MongoDB will be used for storing Auto Labelled Taxonomies

Collections (tables) should be set up in MongoDB to hold the metadata for each category like books , newspapers etc. Each document (row) in MongoDB represents an article/document. The metadata specific to a document shall be kept in Jason format.

Data Access from MongoDB to Islandora

1.Export data from MongoDB to Islandora using Data Migration APIs (needs to be researched)

Data shall be exported in to Islandora Rest API / Data Migration APIs

https://islandora.github.io/documentation/technical-documentation/migration-overview/

https://islandora.github.io/documentation/technical-documentation/migration-migrate-api/

2. Update Islandora manually via Islandora workbench for the data captured in MongoDB

this option needs CSV in the right format with the data captured in MongoDB

https://islandora.github.io/documentation/technical-documentation/migration-overview/

3.Lookup data from MongoDB to Islandora

Currently one of the default content types is called "repository_item" which contains multiple taxonomies. These taxonomy fields may hold authority links. This needs to be researched and confirmed whether Islandora can access MongoDB via authority link is possible or not.

Mentors : Prashanth Sirinivasan
Partners :
Donors :