Project Acceptance Tests.

Team 3-CITS3200 Professional Computing-Documentation Identification Software
By Rania Khan, Hazel Wang, James Felstead, Peter Fang,Chunyu Zheng

Objectives

The purpose of this document is to outline the general approach Group 3 will take towards the testing of our document identification software. The testing will require the system to receive various identifiable document information and metadata relating to a migrant census of a specific province over the last 150 years and then using that input search the internet for a matching source.

These tests will be conducted using sample census documents with already known/located sources to verify and investigate whether the software matches expected functionality. These tests will also require the software to function under reasonable time constraints and appropriately identify if an input has no matching source.

Test Summary

Our system should function as an efficient source search tool with inputs relating to attributes of a census for a user to enter such information into fields such as "Country", "Province", "Year" and "Author". By entering valid information as inputs the system using AI software should function as a deep search across the internet matching all fields specified by the user to a document source.

The output of our tests should download the corresponding PDF if one is found into a structured directory. If no such PDF exists then the most accessible library containing the document should instead be located. The most accessible library should take into account aspects such as if a library accepts remote data requests and its geographic location. If neither an online source or library match is found then the system must clearly notify the user that no census document matches could be identified.

Testing Strategy

The testing of the implemented document identification application will be carried out in three main phases. These three phases are unit testing, integration testing and system testing. Unit testing will be applied to each individual module of the software to verify all subsystems functionality. Integration testing will be applied to verify each subsystem of the application properly incorporates and interacts with other components in order to produce an expected output. Finally system tests will be used to replicate a user's complete interaction with the software and require the entire system to be implemented. These will be completed on a test version of the software that is separate from the live version and produce information such as runtime for each process during the testing.

Subsystems:

1. User Interface including text fields - Interacted with by the user. Receives identifiable attributes of a census document and passes the information to the Automated search subsystem.

2. Automated internet source search - Powered by AI deep search software. Utilised census metadata and other provided information to search through web resources for matches to the specific document.

3. Source handling module - Retrieves a matched PDF if found. Otherwise uses scraped metrics to create a ranking of accessible libraries for the document source.

4. File management system - Automates the organisation of downloaded PDF files into a structured directory located in cloud storage.

Test A - Basic User Interface and Input Handling

Test A focuses on verifying the functionality of the user interface and that it correctly receives and validates document attributes entered by the user such as Country, Province and Year. It ensures that the inputs are valid and passed onto the next subsystem correctly.

Test Specification

Requirements tested:
UI input fields are functional and responsive for users to enter document information..
Inputs are correctly validated. (Year is within a reasonable range.)
Data is correctly passed downstream in an expected and practical structure.

Methods tested:
Validate inputs in every appropriate field.
Data transfer to AI internet search subsystem.

Test Description

- Means of Control: Manual inputs via user interface form fields
- Data
    - Input Data: Set of census meta data
    - Input Commands: User types input and clicks the "Search button"
    - Output Data: Form validation message to user, data packet forwarded to AI system.
    - System Messages: Invalid input message for incorrect inputs ("Year must be between 1860 and 2020"). Confirmation message or indicator that the AI software is processing the request.
- Procedures: 1. Open application, 2. Enter valid data into all fields and submit, 3. Verify the data is passed onto the AI subsystem correctly. 4. Enter invalid data and submit. 5. Verify appropriate error messages appear and submission is not continued.

Test Analysis Report

- Function: User inputs are correctly received, validated, and passed to the next subsystem without alteration.
- Performance: Input validation occurs instantaneously or within acceptable UI response times (<1 second).
- Data measures: All fields accept expected data types and ranges. Errors correctly flagged for invalid inputs.

Any validation failures,or data transfer issues will impact assessment (If Year validation fails downstream searches will yield inaccurate results).

Test B - Source Retrieval and File Management

Test B verifies that if the AI search subsystem identifies and locates an accessible matching PDF the system correctly retrieves the PDF, downloading it and organising it into a directory located in cloud storage. This test ensures the ability and functionality of the application to retrieve specific PDFs from the internet and moves them into the right location with a standardised naming convention.

Test Specification

Requirements tested:
Ability to retrieve a located PDF and download the file.
File management system's ability to label, save and organise the file into the cloud storage directory.

Methods tested:
URL retrieval method.

PDF download process.
File processing and automatic directory structuring.

Test Description

- Means of Control: Automated test driver simulating many successful matches and PDF URL retrieval.
- Data
    - Input Data: Location of a known accessible PDF source.
    - Input Commands: System triggering download process after receiving PDF.
    - Output Data: PDF downloaded and stored within cloud stored in structured directory system.
    - System Messages: Confirmation of the file successfully being downloaded and its placement and name.
- Procedures: 1. Simulate the system acquiring the information and location of a known and accessible PDF source 2. Trigger automated PDF retrieval 3. Confirm the PDF saved has been stored in the correct directory within cloud storage 4. Verify the file is uncorrupted and retrievable from cloud storage by users and check system logs for successful operation.

Test Analysis Report

- Function: Correct PDF retrieved and downloaded without corruption. Files organised into an efficient user specified directory system for census source search applications.
- Performance: Download from web location to cloud storage within acceptable time based on total file size and internet connection speeds. (lower bound around 3min/GB for user application)
- Data measures: File integrity check, Naming convention and location match expected output.

If the download completes but the file is corrupted, the system should retry or flag the file for manual review or remove the file. Otherwise the entire software functionality requiring a complete database of census sources is undermined.

Test C - Library Location Fallback

Test C requires the system to handle a nonoptimal case where no online PDF sources exist for the census specified. In this test the system must recognise this and identify libraries that hold the document and then preference a library that allows for requested remote access of the document.

Test Specification

Requirments tested:
The system correctly identifies when a specified document has no existing PDF match.
The search identifies libraries that have been determined to contain the document match.
Using ranking logic the system identifies libraries that allow for remote access as the best accessibility match.

Methods tested:
PDF availability check without matching false positives.
Deep search for all libraries containing specified census document.
Web scraping of Library accessibility attributes.
Ranking algorithm of most accessible source location.

Test Description

- Means of Control: Manual user entering of an already located document known to have no existing pdf matches and multiple library sources..
- Data
    - Input Data: Meta data of census available exclusively in libraries.
    - Input Commands: System triggering search of library matches after no PDF existing is determined.
    - Output Data: Display most accessible Libraries in a visual list at the user's specifications.

    - System Messages: Clear output on an extensive document search finding no matches.
- Procedures: 1. Input metadata and other text fields relating to a PDF known to have no online matches. 2. Have an automated extensive search be run that concludes that no match could be found. 3. Libraries are automatically detected to contain the document. 4. Retrieve and rank locations of libraries based on their accessibility attributes. 5. output a clear ranking to the end user of such locations.

Test Analysis Report

- Function: PDF search correctly identifies no false positives before library search. Library search is accurate and collects accessibility attributes.
- Performance: PDF search is extensive while occurring in a reasonable runtime as this case is relatively common during usage for obscure census data (less than 2min). Collecting and ranking of libraries should be much faster.
- Data measures: Determination of no false positives. Accuracy of libraries identified and accessibility attributes of those libraries. Correct ranking of output.

Common determinations of a false positive PDF match will undermine the entire purpose of the required function of the document identification software. A search that incorrectly determines all libraries that contain the census document could cost unnecessary resources if retrieval is attempted. Unclear or incorrect ranking of accessibility of such libraries could also cost unnecessary resources if retrieval is attempted.

Test D - Non Existing Census Document

Test D checks all system fall backs in the case of a document being unreachable on the internet or any libraries. It should run through all fallback methods with extensive searches before clearly notifying the user that no such document exists. This allows for the user to be aware of which fraction of census documents are unreachable by automatic means.

Test Specification

Requirements tested:
The system correctly identifies without false positives that no matching PDF exists.
The system correctly identifies without false positives that no libraries exist that contain the specified document.
The user receives clear notification that the document is unreachable.

Methods tested:
Extensive AI powered PDF search.
Extensive AI powered document library search detection.
User front end notification system of non-existent documents.

Test Description

- Means of Control: Automatic testing simulating searches of nonexistent census documents. Using permutations of countries, provinces and years known to have no census data.
- Data
    - Input Data: Artificial text fields as inputs for a fictitious census document.
    - Input Commands: Simulated user input and search function.
    - Output Data: Notification that such a document has no match.
    - System Messages: Log message that extensive searches occurred with no results.
- Procedures: 1. Enter artificial metadata into the search system. 2. Run a census search on the document using the AI system. 3.Verify no matches for PDF document. 4. Verify no matches for library containing the document. 5. Confirm that negative notification is displayed to user and outputted correctly.

Test Analysis Report

- Function: Failure detection without false positives works extremely consistently at both PDF and library searches.
- Performance: Total search time within a reasonable timeframe.
- Data measures: Percentage of fictitious documents that have falsely found a match at either stage close to zero.

If such documents are not notified to the user then the functionality of determining the sources and availability of sources of census documents are incomplete. Any false positives indicate poorly implemented match searching that poisons the reliability of the systems output.