jamk.fi

Film Archive Management System

Sami Antila

Bachelor's thesis
December 2016
Technology, communication and transport
Degree Programme in Software Engineering

Jyväskylän ammattikorkeakoulu
JAMK University of Applied Sciences

jamk.fi

Description

Author(s)	Type of publication	Date
Antila, Sami	Bachelor's Thesis	December 2016
		Language of publication: English
	Number of pages: 70	Permission for web publication: x

Title of publication
Film Archive Management System

Degree programme
Software Engineering

Supervisor(s)
Salmikangas, Esa

Assigned by
King Mongkut's University of Thonburi, CPE, ESIC-Laboratory; Dr. Priyakorn, Pusawiro

Abstract

The aim of this thesis was to research and design a web-based film archive management system for National Film Archive (Public Organization), The Thai Film Archive (FAPOT). The purpose of the system was to be used to manage and view metadata, documentation and objects of archived films.

The research goal was to find what was available, what could be bought and what was needed to be designed and developed within the project. To accomplish this, archiving standards used by other national film archives were studied. Websites with similar functionality were studied and researched using reverse-engineering to find out useful features for the project. The research included both private and public national film databases. The differences between private and public film databases were discussed. Cur-rent technologies were researched, compared and the system architecture was designed. The best web techniques and practices were also studied.

The design concentrated on the critical features for the system from user standpoint, and a viable front-end system based on research was created. The main focus in design was on usability, which answered how users can find the requested information from the database. The basic functionality for back-end system was also designed. Different standards were researched, and it was discussed if a standard should be used, should a new standard be created or should the system be created without any standard. The design concentrated on European EN 15744 and EN 15907 standards.

The results of the thesis show what is important in the usability of the film archive, what kind of standards and technologies the film archive should use and what features should be implemented.

Keywords/tags (subjects)
Film, Archive, Web, Standard

Miscellaneous

jamk.fi

Kuvailulehti

Tekijä(t)	Julkaisun laji	Päivämäärä
Antila, Sami	Opinnäytetyö, AMK	Joulukuu 2016
	Sivumäärä 70	Julkaisun kieli Englanti
		Verkkojulkaisulupa myönnetty: x

Työn nimi
Filmiarkiston hallintajärjestelmä

Tutkinto-ohjelma
Ohjelmistotekniikan koulutusohjelma

Työn ohjaaja(t)
Esa Salmikangas

Toimeksiantaja(t)
King Mongkut's University of Thonburi, CPE, ESIC-Laboratory; Dr. Priyakorn, Pusawiro

Tiivistelmä

Opinnäytetyön tavoitteena oli tutkia ja suunnitella verkkopohjaisen elokuva-arkiston hallinta järjestelmää Thaimaan kansalliselle elokuva-arkistolle (National Film Archive (Public Organization), The Film Archive, FAPOT). Järjestelmän tarkoituksena oli hallita ja tarkastella filmeihin liittyviä dokumentteja, esineitä ja metadataa.

Tutkimuksen tavoitteena oli selvittää, mitä oli saatavilla, mitä on mahdollista ostaa ja mitä tarvitsee suunnitella ja kehittää projektin sisällä. Tämän saavuttamiseksi tutkittiin arkistointistandardeja, joita muut kansalliset elokuva-arkistot ovat käyttäneet. Sivustot, joiden toiminnallisuudessa oli yhtäläisyyksiä projektin kanssa, tutkittiin ja mallinnettiin takaisinmallinnusta hyväksi käyttäen. Tutkimus sisälsi sekä yksityisiä että kansallisia elokuva-arkistoja ja tietokantoja. Pohdittiin, miten yksityiset ja kansalliset elokuva-arkistot eroavat. Työssä tutkittiin verkkosovellusten kehittämiseen liittyviä teknologioita käytäntöjä, sekä suunniteltiin järjestelmäarkkitehtuuria.

Suunnittelu keskittyi kriittisten ominaisuuksiin järjestelmän käyttäjän näkökulmasta ja käyttökelpoisen käyttöliittymän luontiin. Suunnitteluun sisältyivät käyttöliittymähahmotelmat, joissa pääpaino oli käytettävyydessä.

Eri standardeja tutkittiin ja käsiteltiin pitäisikö standardia käyttää tai onko uuden standardin kehittäminen tarkoituksenmukaista. Suunnittelu keskittyi CEN:n (European Committee for Standardization) kehittämiin EN 15744 ja EN 15907 standardeihin.

Työn tuloksena saatiin selville, mikä filmiarkiston käytettävyydessä on tärkeintä, millaisia standardeja filmiarkisto voisi käyttää, millä teknologioilla filmiarkiston hallintajärjestelmä voidaan toteuttaa ja millaisia toimintoja sen tulee sisältää.

Avainsanat (asiasanat)
Filmi, Arkisto, Web, Standardi

Muut tiedot

Contents

Terminology 5

1 Introduction 6

1.1 Background 6

1.2 Problem Statement 6

1.3 Scope 7

2 Research 8

2.1 Film Archive (Public Organization) The Thai Film Archive 8

2.2 Soffico (Thailand) limited 8

2.3 Internet Movie Database 8

2.4 Rotten Tomatoes 10

2.5 British Film Institute 10

2.6 EN 15744 & EN 15907 Standards 10

2.7 Collections Management System 11

3 Technical Research 12

3.1 Collective access 12

3.2 Internet Movie Database 30

3.3 British Film Institute Collection Search 36

4 Results 48

4.1 Outcome of Research 48

4.2 Front-end Features 49

4.3 Administration Features 50

5 Design 52

5.1 System overview 52

5.2 Front-end 54

5.3 Administration 60

6 Conclusion 61

References 64

Appendices 66

Appendix A 66

Appendix B 67

Appendix C 68

Appendix D 69

Appendix E 70

Figures

Figure 1. Sitemap of Collective Access demonstration website 13
Figure 2. Collective Access login page screenshot 14
Figure 3. Collective Access login page flow chart 15
Figure 4. Collective Access login page sequence diagram 16
Figure 5. Collective Access browse page screenshot 17
Figure 6. Collective Access screenshot of browse/object page 18
Figure 7. Collective Access screenshot of object edit page 18
Figure 8. Collective Access browse and edit object flow chart 19
Figure 9. Collective Access browse and edit item sequence diagram 20
Figure 10. Collective Access screenshot from creating collection 21
Figure 11. Collective Access add collection flowchart 22
Figure 12. Collective Access add collection sequence diagram 23
Figure 13. Collective Access screenshot from new object page 24
Figure 14. Collective Access add new object flow chart 25
Figure 15. Collective Access add new object sequence diagram 26
Figure 16. Collective Access screenshots from deleting an object 27
Figure 17. Collective Access deleting an object flow chart 28
Figure 18. Collective Access deleting object sequence diagram 29
Figure 19. IMDB search results sort by methods 30
Figure 20. IMDB advanced search, results, movie page map 31
Figure 21. IMDB advanced search flowchart 32
Figure 22. IMDB advanced search sequence diagram 33
Figure 23. IMDB search results flowchart 35
Figure 24. IMDB search results and movie page sequence diagram 35
Figure 25. BFI Collection Search sitemap 37
Figure 26. BFI screenshot of search page 38
Figure 27. BFI screenshot of search results 38
Figure 28. BFI screenshot of item page 39
Figure 29. BFI search, search results and details page flowchart 40
Figure 30. BFI search, search results and details page sequence diagram 41

Figure 31. BFI search results and selection screenshots and notes....42
Figure 32. BFI Selection flowchart....43
Figure 33. BFI Selection sequence diagram....44
Figure 34. BFI Search history screenshots....45
Figure 35. BFI search history flowchart....46
Figure 36. BFI Search history sequence diagram....47
Figure 38. System diagram with planned technologies.....52
Figure 39. FAPOT Film Collection front-end sitemap....54
Figure 40. FAPOT Film Collection user journey flowchart....55
Figure 41. FAPOT Film Collection introduction page mockup....56
Figure 42. FAPOT Film Collection search page mockup....57
Figure 43. FAPOT Film Collection search results page mockup....57
Figure 44. IMDB Compact and Detailed search results....58
Figure 45. FAPOT Film Collection search history page mockup....58
Figure 46. FAPOT Film Collection browse page mockup....59
Figure 47. FAPOT Collection item page mockup....60

Terminology

API

API is an acronym for Application Program Interface, which is “a set of routines, protocols, and tools for building software applications” (Beal, V.).

Archive

An archive is a place where often earlier material is stored. Usually anykind of media with a historical or informational value (Archive).

JSON

JSON is an acronym for JavaScript Object Notation, which is a language independent data interchange format (The JSON Data Interchange Format).

MySQL

MySQL is an open-source database management system. MySQL uses relational database to store data in separate tables. (What is MySQL?)

PHP

PHP is a recursive acronym of PHP: Hypertext Preprocessor. It is a general purpose scripting language, used commonly for web-development (What is PHP?).

Standard

“A standard is a document that sets out requirements for a specific item, material, component, system or service, or describes in detail a particular method or procedure.” (What is a European Standard (EN)?)

Relational Database

“A relational database is a collection of data items organized as a set of formally-described tables from which data can be accessed or reassembled in many different ways without having to reorganize the database tables.” (Relational database).

XML

XML is an acronym for eXtensible Markup Language. “XML is a software- and hardware-independent tool for storing and transporting data.” (What is XML?)

1 Introduction

The introduction chapter gives overview of the project, explain background information and defines the scope of the project.

1.1 Background

National Film Archive (Public Organization) The Thai Film Archive has collected and received in form of donations a collection of films and items related to films. They have documented every donation, loan of a film or item and restoration of film.

Everything has been done on paper or in an analog way. Film is an analog medium: useful for the purpose it is made for, however, not long lasting. By digitalizing their collection, documenting all items and moving their paper files to digital form. Improvement will be made for preservation of national Thai history. Information and films will be better available for viewing and research purposes. Thai Film Archives will also have a more efficient system for their log keeping and documentation. Thai Film Archives has five teams: registration, film, digital, IT and museum teams.

1.2 Problem Statement

FAPOT has films, materials, documentations and logs on an analog medium located in FAPOT's headquarters. FAPOT's goal was to digitalize and make material available in digital form. To help achieve this, the goal was to find useful standards, technologies and features that suit the project. Study if It was required to create, modify or use existing film archiving, cataloging and collection standards and frameworks.

The project approached this problem by first researching for film archiving, cataloging and collection standards around the world. Researching what was available for free, what can be bought and what is needed to design and develop within the project.

Client stakeholders of the project were National Film Archive (Public Organization) The Thai Film Archive and KMUTT Document and Media Archive. The project was based on ESIC Laboratory located in King Mongkut's University of Technology Thonburi's Computer Engineering department.

1.3 Scope

The primary objective was to research standards, features and technologies to help FAPOT project. The purpose was to research features of the web-based system to help FAPOT project with digitalizing, preserving, cataloging, archiving and collecting all material and information in their facility with ease-of-use for normal end-user knowledge level with low difficulty to learn basic tasks. The system enables better preservation, ease-of-sharing, more efficient and easier organizing of all material. Research what is used in similar project and consider commercial solutions. The goals are listed below:

1. Research the current software focused on Film Archive.
2. Explore database schematic and XML metadata format or standard from various archive system, based on IMDB.com.
3. Study search algorithm and functions from feasible software then compare the best solution for media archives.
4. Study “National Archives” websites from various countries then analyze selected website structures and each web function.
5. Design possible database schematics and relation database management system (RDBMS) for media archives.
6. Design possible functions for users to query or search for archives efficiently.
7. Document all processes.

2 Research

The research chapter discusses the completed research and all topics researched during the thesis work. The chapter begins with information on FAPOT and then continues with information about associate company, standards and other film archives.

2.1 Film Archive (Public Organization) The Thai Film Archive

The Thai Film Archives has collected and received a collection of films and all kind of items related to films in form of donations. They have documents about every donation, loan of a film or item and restoration of film.

Thai Film Archives has five teams: registration, film, digital, IT and museum. Registration team takes donations and documents donations. Each of the incoming items are documented and sorted to each team accordingly. Film team restores and documents films. Digital team handles all digital material, e.g. DVDs and movie posters are forwarded to the digital team as well. They restore and maintain them in good condition. Restored items are forwarded to museum team. Museum team manages everything else except films: including documents, books, tools, magazines, clothes and props. Museum team also keeps record about loaned items.

2.2 Soffico (Thailand) limited

Soffico is a software consulting company located in Bangkok that helped to consult in the FAPOT project. On 22nd of May 2016 the project had a meeting at Soffico's office in President building, Bangkok. In the meeting, the main topic was to go through the inception report and to get a professional opinion as well as advice on how it could be improved. At the meeting, the document was also translated for the non-Thai speaking team member and the project was explained more widely.

2.3 Internet Movie Database

Internet movie database or more commonly known by its acronym: IMDB. IMDB is a web-based database of films, video games and television programs. Internet movie

database has most of the database freely available on Alternate Interfaces page (Alternative Interfaces). Alternative Interfaces page offers files of the database. The dumb file contains different files in XML and JSON formats. They are easy to access, however, going through the data does not give away much of the structure of the database. Only data is available, only some connections are visible and can be logically guessed. Dumb files do not contain all entries in the database, only most popular and/or new items. (*Alternative Interfaces*).

One logical step could be to use the metadata of films in the project. More information, terms and conditions were available on Internet Movie Database's website (Can I use IMDb data in my software?).

With further research, an ORM-UML diagram of IMDB's database structure from 2007 was found (Appendix 1). The main table in this database diagram is the title table followed by the tables film, series and episode, which are made for determining the type of title by looking at title identification. From series and film tables the next are year, film item and type. The database structure is functional; however, it seems over complicated. If title is being looked from the film table first and found, it can could save processing time. In the worst case, it is not found in the film table and is the last compared item in the series table. This seems a reasonable solution. If ID is a number and is the last compared item in the series table. This seems reasonable solution. If ID is a number, table has a lot items and is sorted in numerical order.

When the user loads the movie page it searches and fetches information from the database. If the user wants to see additional information, it will be fetched from the additional database. A diagram of the additional database can be found from Appendix 2.

It was later discussed that FAPOT is not interested in using metadata from IMDB. However, IMDB might have interest to use metadata from FAPOT's upcoming database, which should be discussed with IMDB and thus is not within the scope of this project.

2.4 Rotten Tomatoes

Rotten Tomatoes is another popular internet movie database. The major difference to Internet Movie Database is that they have a working interactive API available and it is well documented. It can be accessed by creating an account and by getting a personal API key. The process was straightforward and the data set is more complete than the one offered by Internet Movie Database.

Working with API does not allow an inspection of database architecture; however, from the JSON string can be seen that the result of the query was similar to that of Internet Movie Database. Rotten Tomatoes seems to be more protective about their database structure. However, Rotten Tomatoes looks more promising on automatically fetching metadata for films. During the research, it appeared that Rotten Tomatoes does not limit access to the database as much as imdb.com. Having an API might be therefore being beneficial for FAPOT project for research purposes.

2.5 British Film Institute

The British Film Institute, acronym BFI, is a public organization that preserves and promotes filmmaking and television produced in the United Kingdom. They decided to modernize their infrastructure in 2008, and their infrastructure was similar to the one that this project is trying to achieve. The British Film Institute chose Adlib as their system supplier. During their modernization process a new European standard for the description of moving image works was finalized and ratified. The British Film Institute decided to be-come the first to adopt EN 15907 standard. By using these four systems/standards, most of what the Thai Film Institute is looking to achieve can be achieved.

2.6 EN 15744 & EN 15907 Standards

Most of the film databases store metadata of the films in two datasets. One for basic information and another for more comprehensive description of the film. EN 15744 standard is used for basic identification of the film and EN 15907 standard is used for more comprehensive information.

The metadata representation has been constructed from the EN 15907 preliminary version of an XML schema (EN 15907 preliminary version of an XML schema). It gives a good idea where the standard is heading. All data is not visible in this representation, only the most critical information. A view like this would be useful, for example, when a normal user would be viewing single item in the catalogue. Metadata representation and XML schema also give a good overview to the standard compared to the documentation available in filmstandards.org website. This gives a much clearer view of the standard on a single page. The metadata representation can be found from Appendix 3.

2.7 Collections Management System

There are many collections management systems commercially available. Two collections management systems looked promising: Open-source Collective Access (Collective Access) and Collection Space (Collection Space) the development of which is led by Museum of the Moving Image of New York. There is a live demo available for Collective Access collections management system (Collective Access Live Demo).

Using a readymade solution to manage collections can in the best case save plenty of design and development time; nevertheless, in this case, a specified type of information that goes into system might take time for modify. In some solutions, the software might not be modifiable. Currently there is no Thai language support available, a translation would need to be made. Open-source solutions are looking most promising as licenses allow to modify the code. The next chapter describes the function of a similar website more in detail.

3 Technical Research

The technical research chapter delves into the topic more in-detail with a technical standpoint and presents reverse-engineering of systems. In this chapter, every system is first presented as a map and then different functions are studied using flow chart and sequence diagram.

3.1 Collective access

Introduction

Collective access is a collection management system similar to what FAPOT project could use in the administrative side of the web service. Collective Access is an open-source content management system which can be modified for FAPOT's requirements.

Sitemap

Collective access has a simple login screen. For example, the forgot password option was missing. After login the user is redirected to the main page where user can add new collections. The user is also able to edit and search for existing documents. Figure 1 is a presentation of the sitemap of Collective Access demonstration website (*Collective Access Live Demo Login Page*).

The Collective Access demonstration website sitemap is presented in Figure 1 below. On the top left corner there is a login page where the user arrives. After valid credentials, the user will have entered and logged in. The user is allowed to access the main page with the red background. From the main page the user can move to administrate objects and collections

Figure 1. Sitemap of Collective Access demonstration website

Login

Figure 2 illustrates a screenshot from Collective Access demonstration website login screen. On the demonstration website, the user and password are told above the text fields which is not part of the actual product and is there only for demonstration purposes. The login screen contains logo, title, username and password text fields and login button.

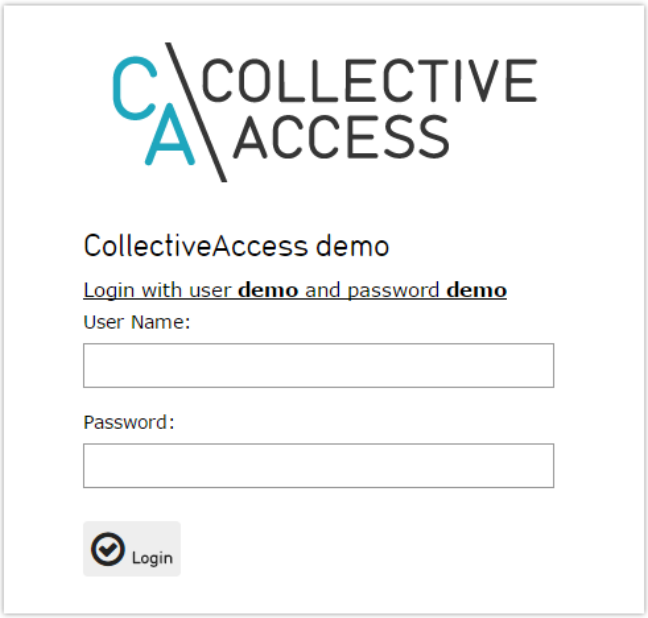


Figure 2. Collective Access login page screenshot

In Figure 3, the login page functionality is presented as an UML flow chart. The users input their credentials and presses the login button. The system checks for credentials and if the credentials are invalid it shows a message and empties the text fields. If the credentials are valid access is granted and the user is redirected to the main page.

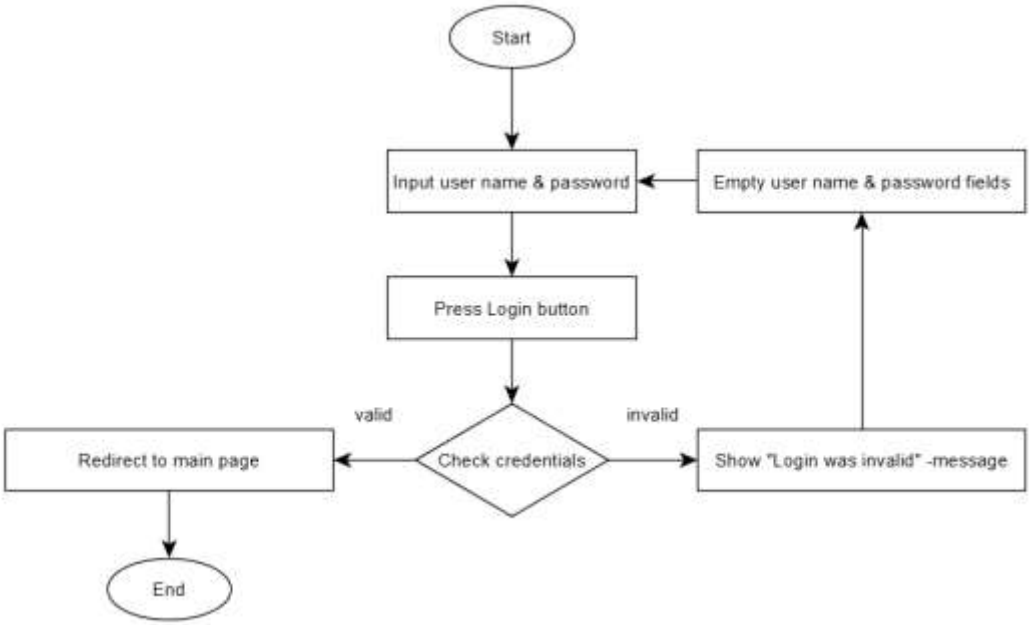


Figure 3. Collective Access login page flow chart

Figure 4 presents the login screen functionality in a UML sequence diagram. In addition to Figure 3 flow chart, the sequence diagram goes more in detail when and how the database is used and in which order the user is redirected to different pages.

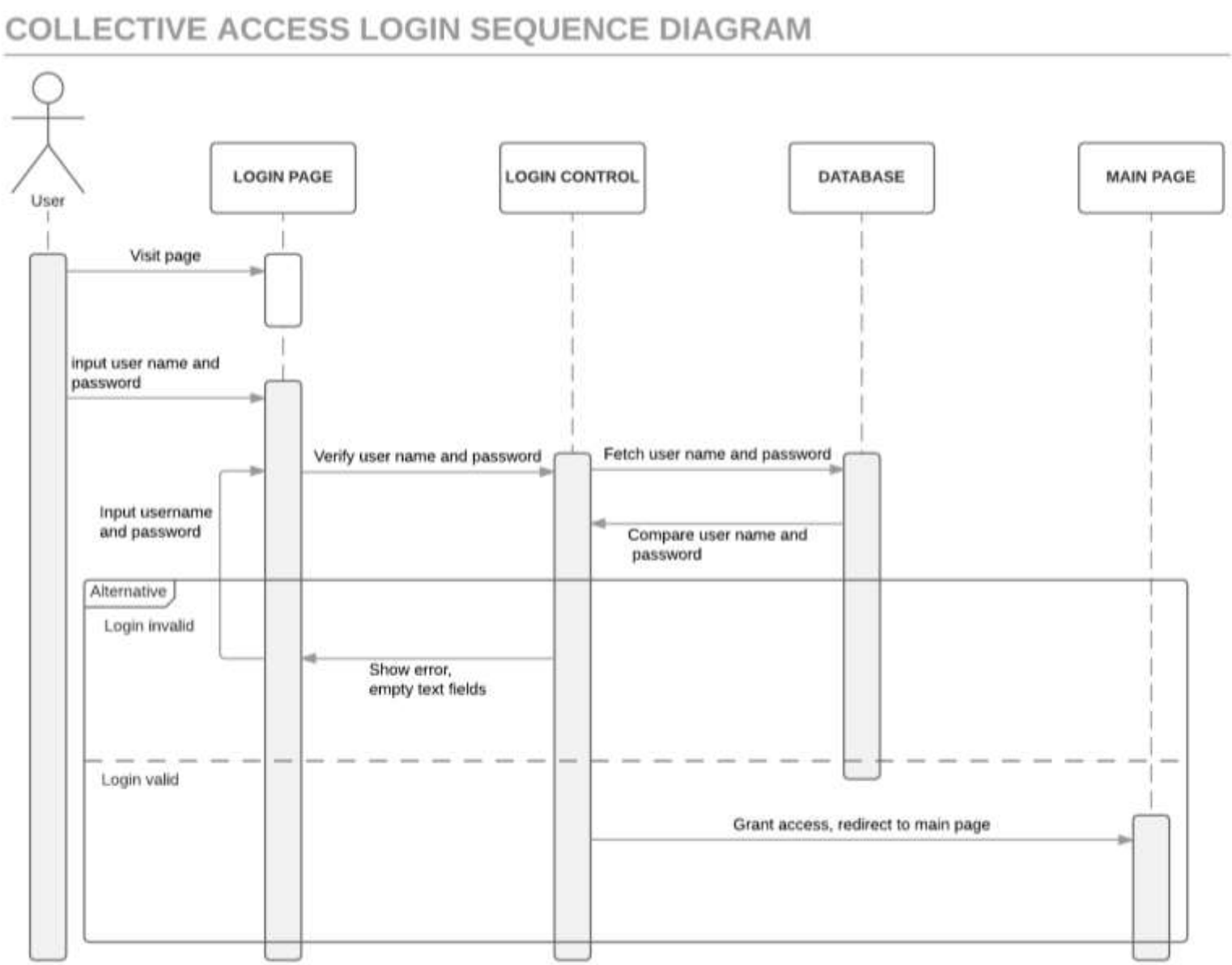


Figure 4. Collective Access login page sequence diagram

Browse and Edit Object

Figure 5 is a screenshot of the browse page on the Collective Access demonstration website where the user can browse objects in alphabetical order.

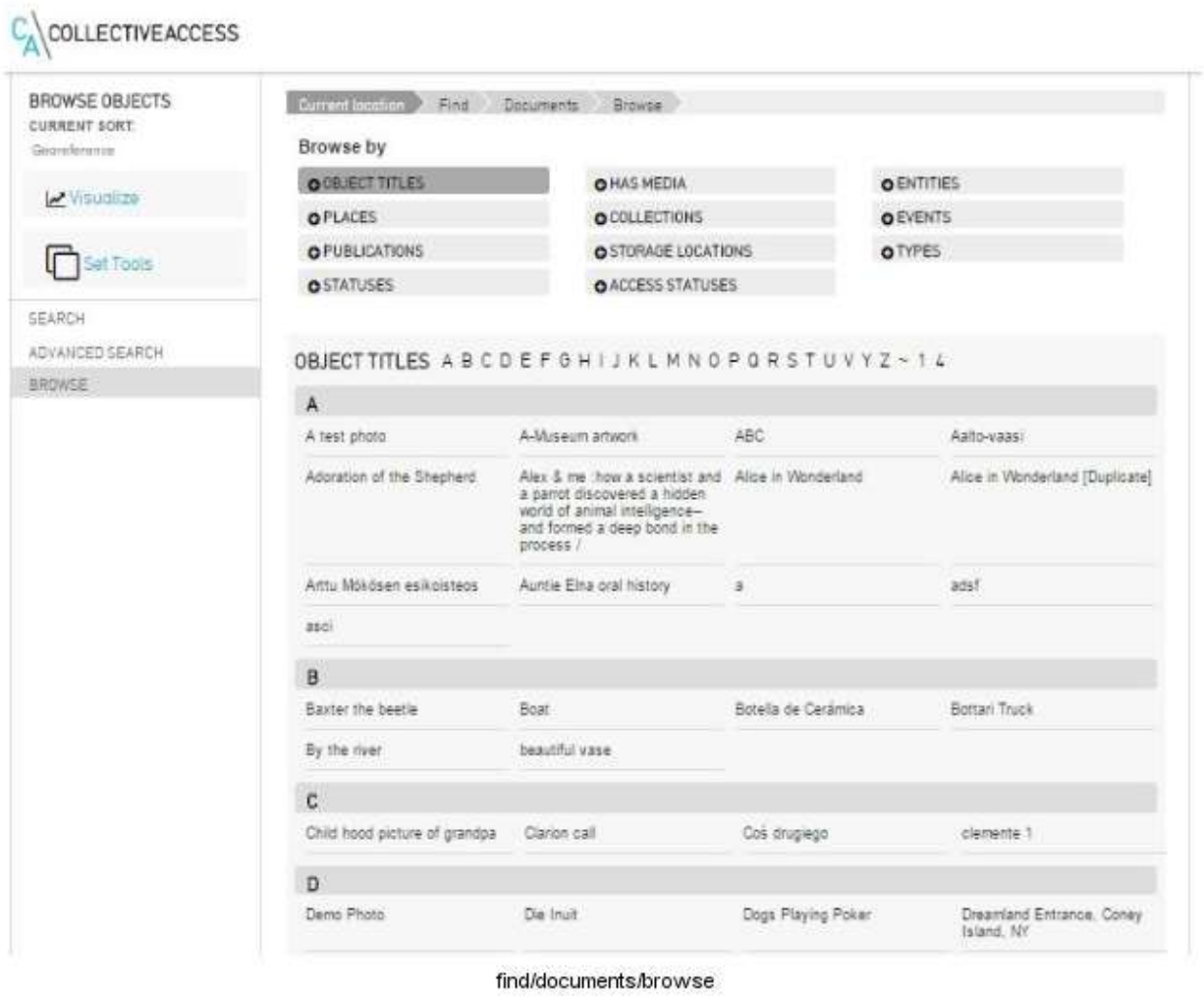


Figure 5. Collective Access browse page screenshot

Figure 6 is a screenshot of the browse page on Collective Access demonstration webpage after an object has been selected.

Figure 6. Collective Access screenshot of browse/object page

Figure 7 is a screenshot of the object edit page of Collective Access demonstration webpage where the user can edit and update an object's information.

Figure 7. Collective Access screenshot of object edit page

Figure 8 presents browsing, selecting and editing of object(s) on Collection Access website as a UML flow chart. It presents the user's journey from browsing to selecting and to editing an object.

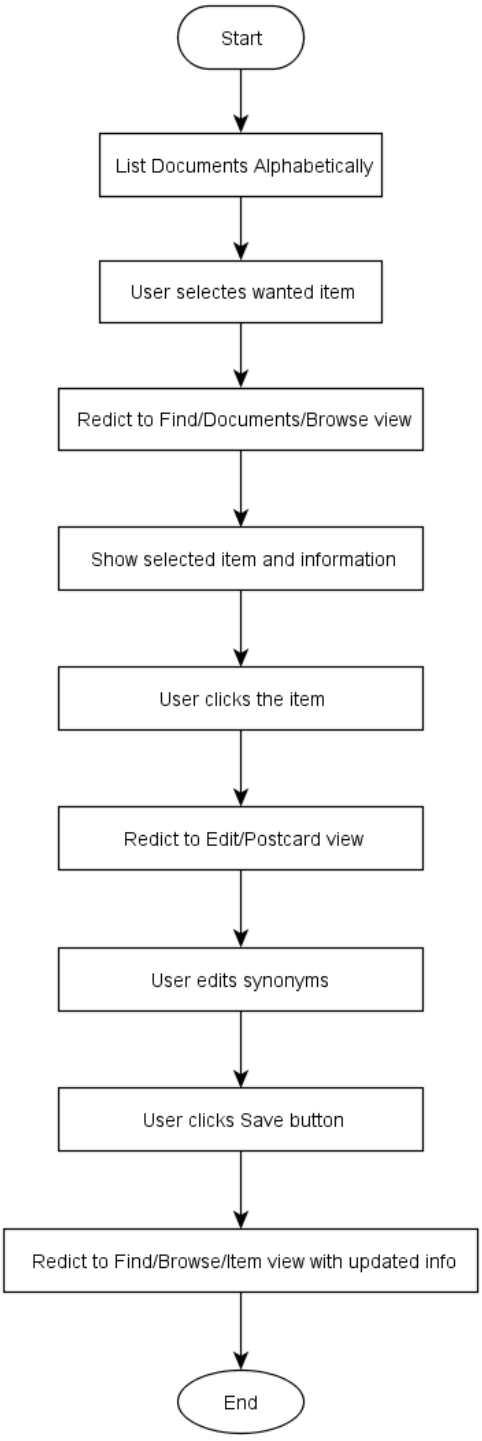


Figure 8. Collective Access browse and edit object flow chart

Figure 9 presents browsing, selecting and editing of object(s) on Collection Access website as a UML sequence diagram. It presents the user's interaction from browsing to selecting and to editing an object. The chart also explains the logic between different pages.

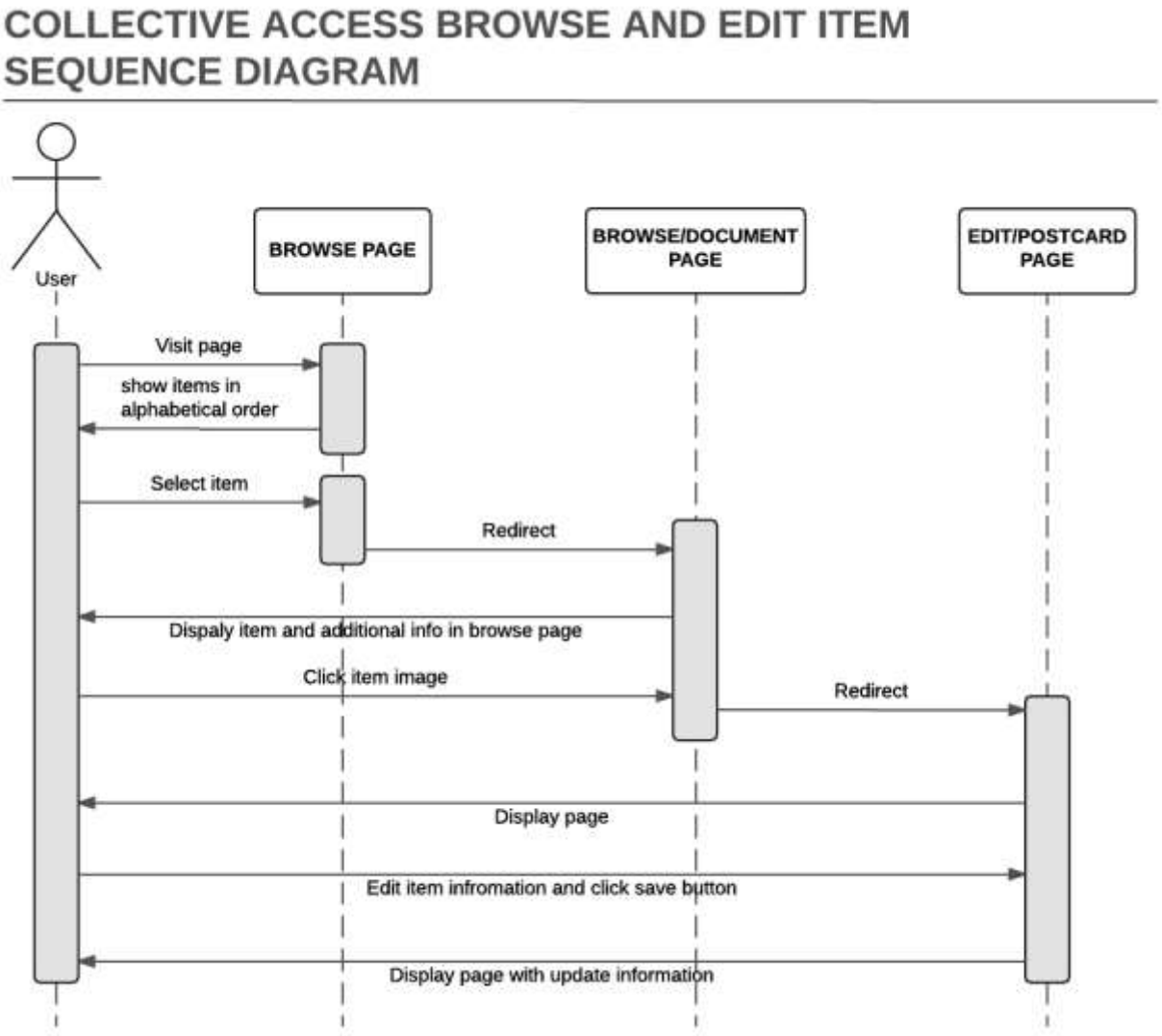


Figure 9. Collective Access browse and edit item sequence diagram

Creating Collection

In Figure 10 presents a screenshot from the new collection page of Collective Access website where the user can create a new collection by filling title text field, optional metadata and by pressing save.

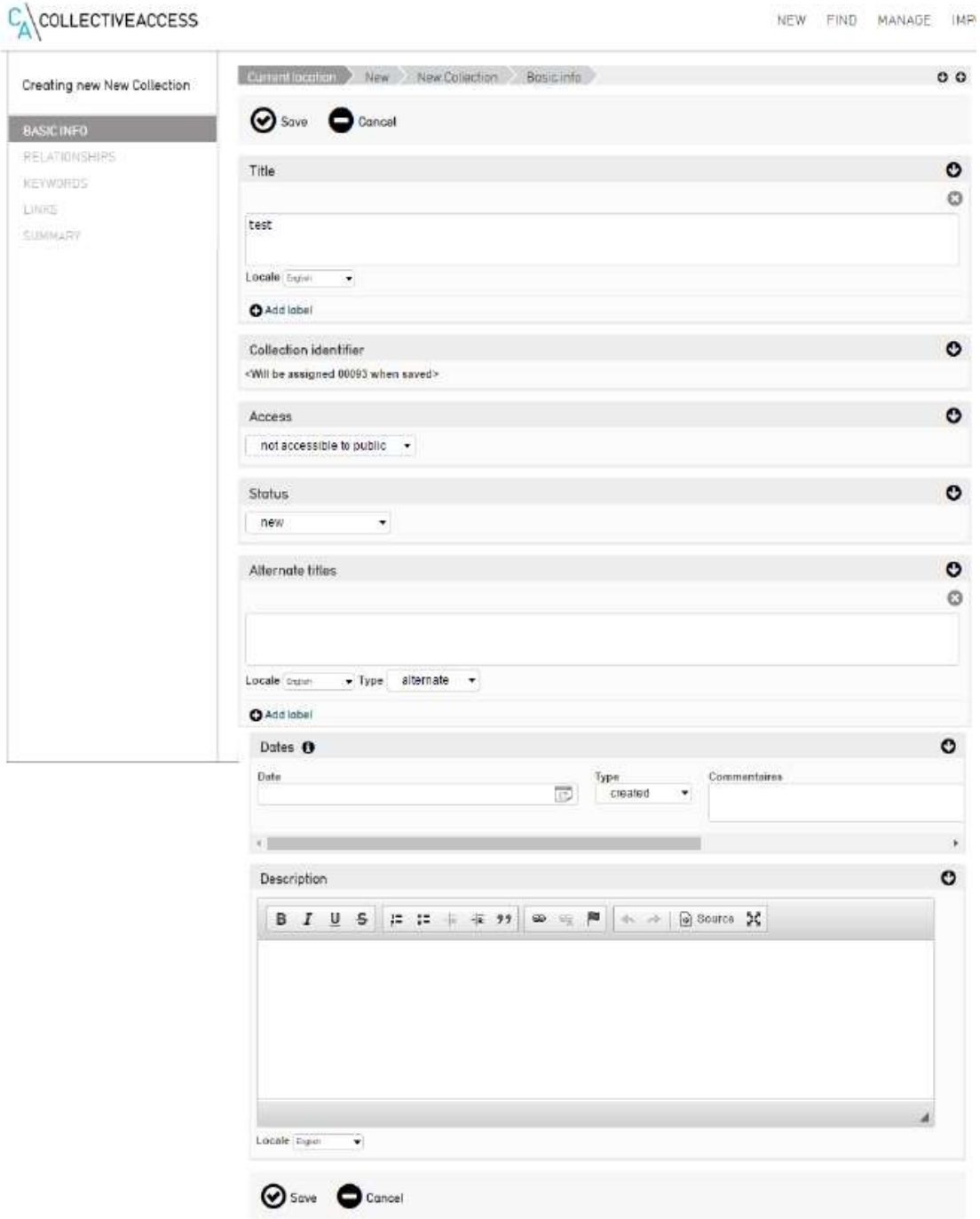


Figure 10. Collective Access screenshot from creating collection

The function to create a new collection is presented in Figure 11 by using UML flow chart. The user fills in the title text field and optional metadata. In the end the user presses save button. Collective Access checks if the title already exists and if it already exists, the user is presented an error message. Otherwise, a new collection is created.

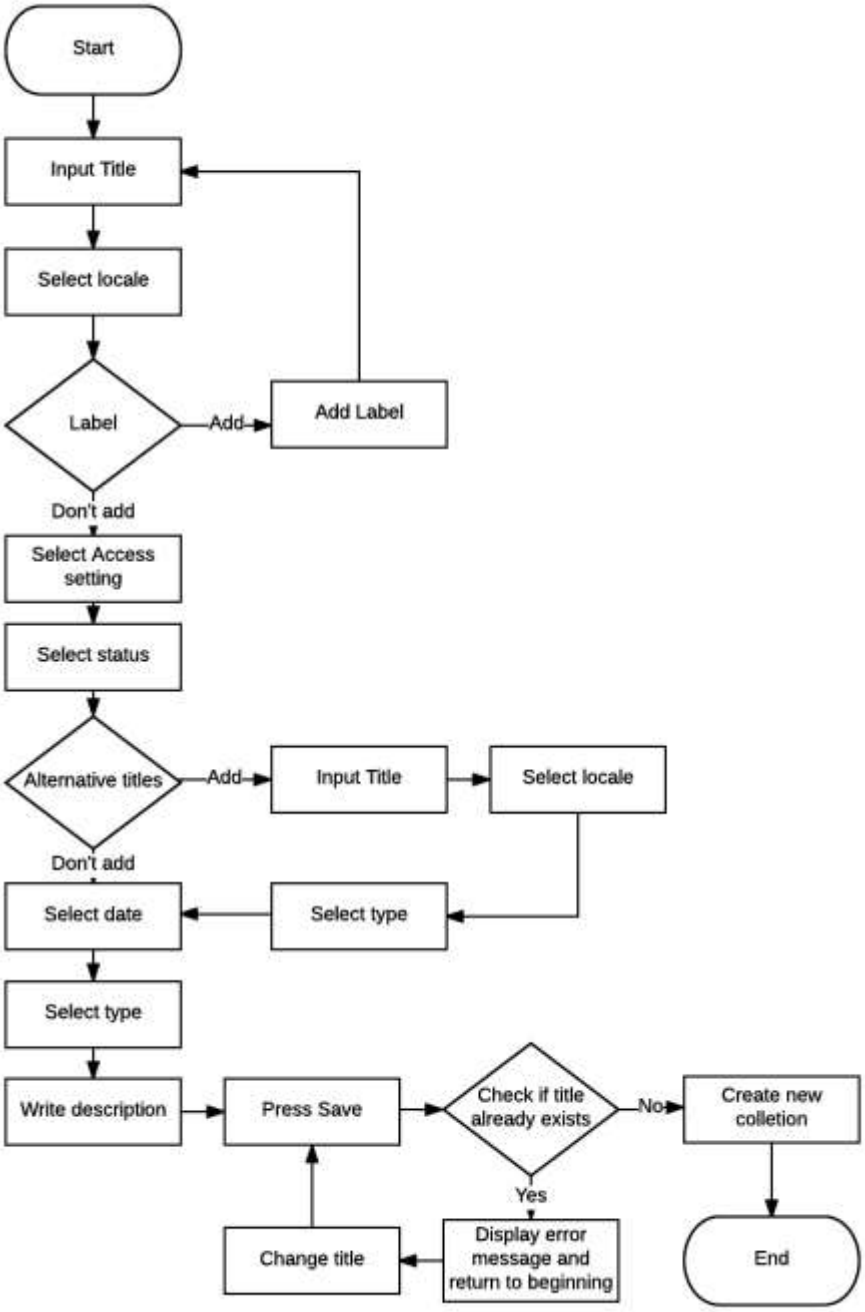


Figure 11. Collective Access add collection flowchart

UML sequence diagram in Figure 12 extends the presentation from Figure 11 by going more into detail of creation of a new collection. The diagram displays more clearly the difference between page and controller.

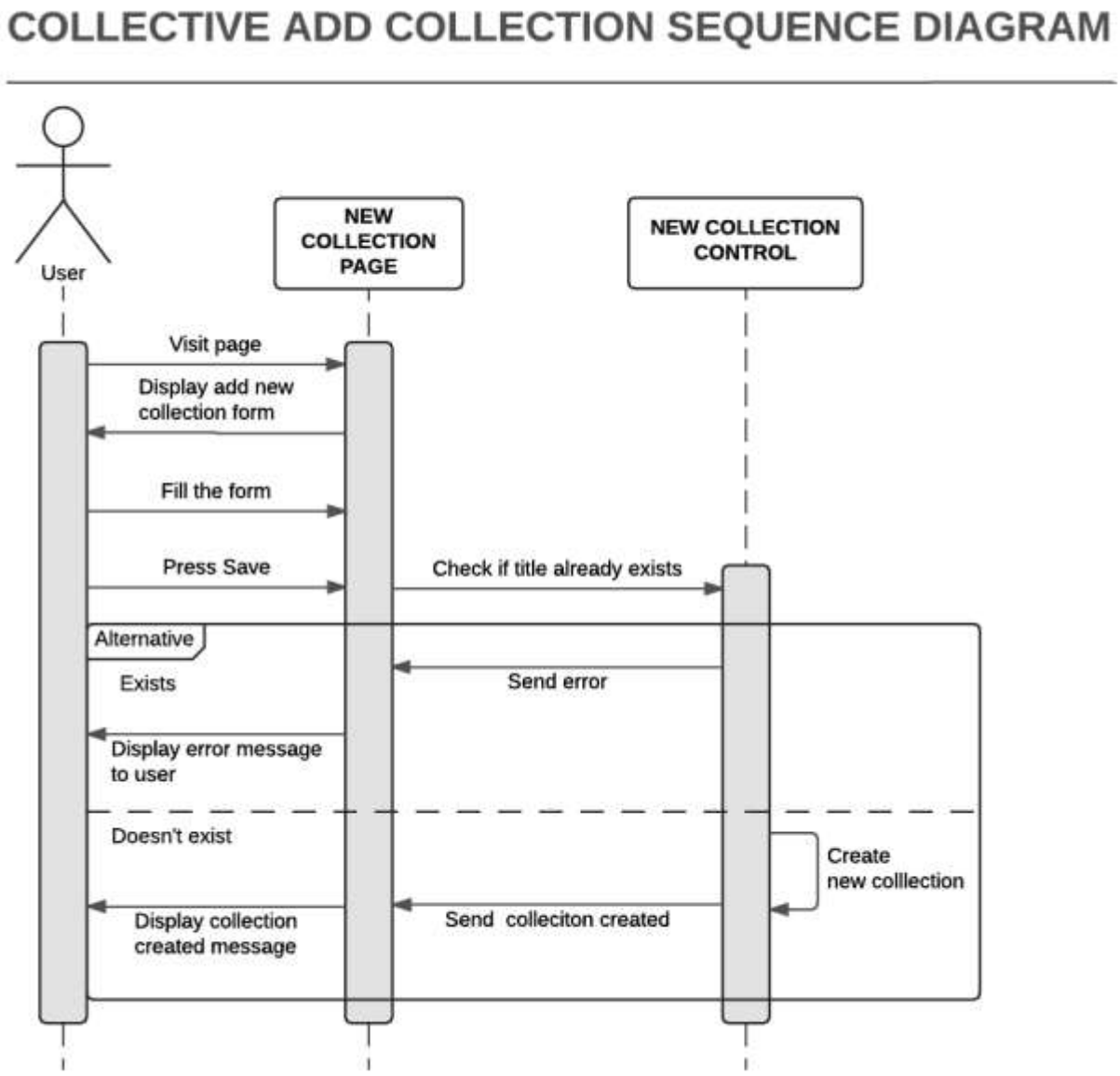


Figure 12. Collective Access add collection sequence diagram

Creating Object

Figure 13 is a screenshot of a new object page. The user can use it to create new objects. The functionality of adding a new object is studied more in the flow chart of Figure 14 and the sequence diagram in Figure 15.

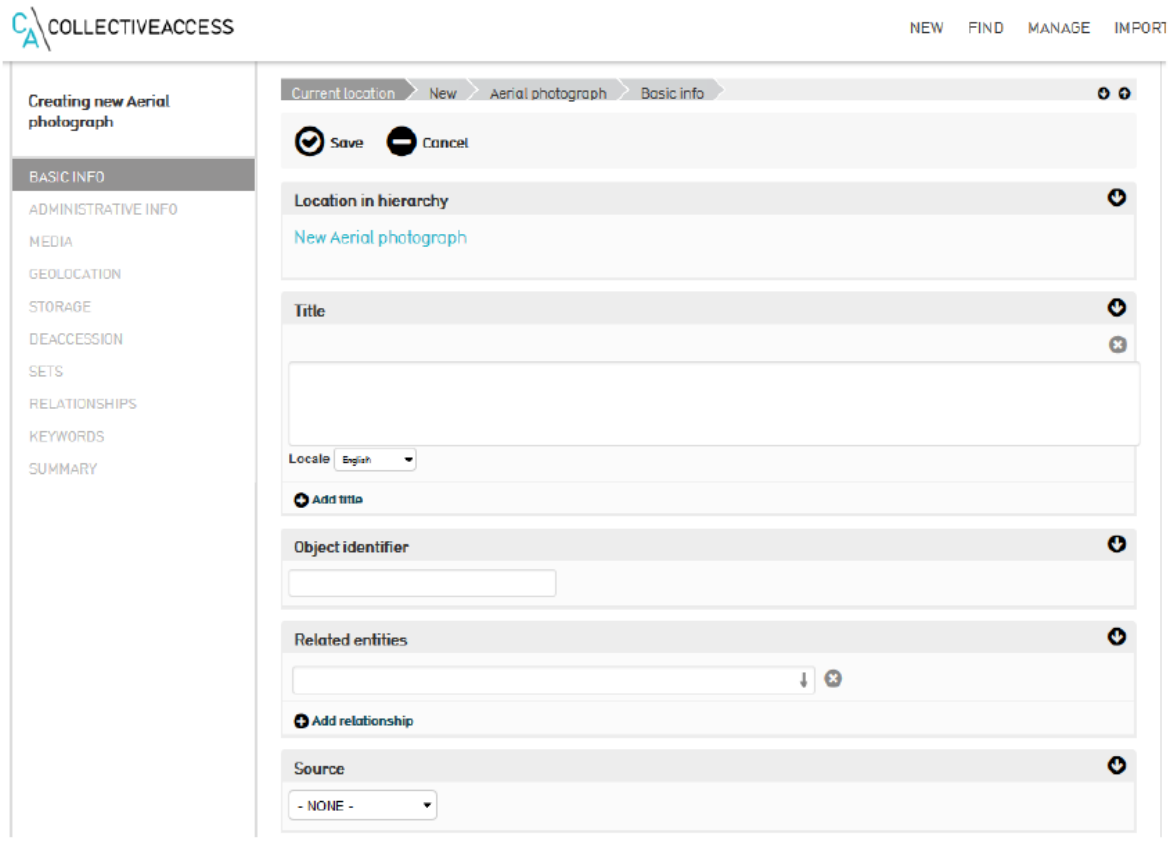


Figure 13. Collective Access screenshot from new object page

The function to add a new object is presented in Figure 14 as a UML flow chart. The user arrives at a new object page, fills in the title text field with optional metadata and clicks save button. Collective Access checks if the title already exists. If a matching title is not found, a new object will be added.

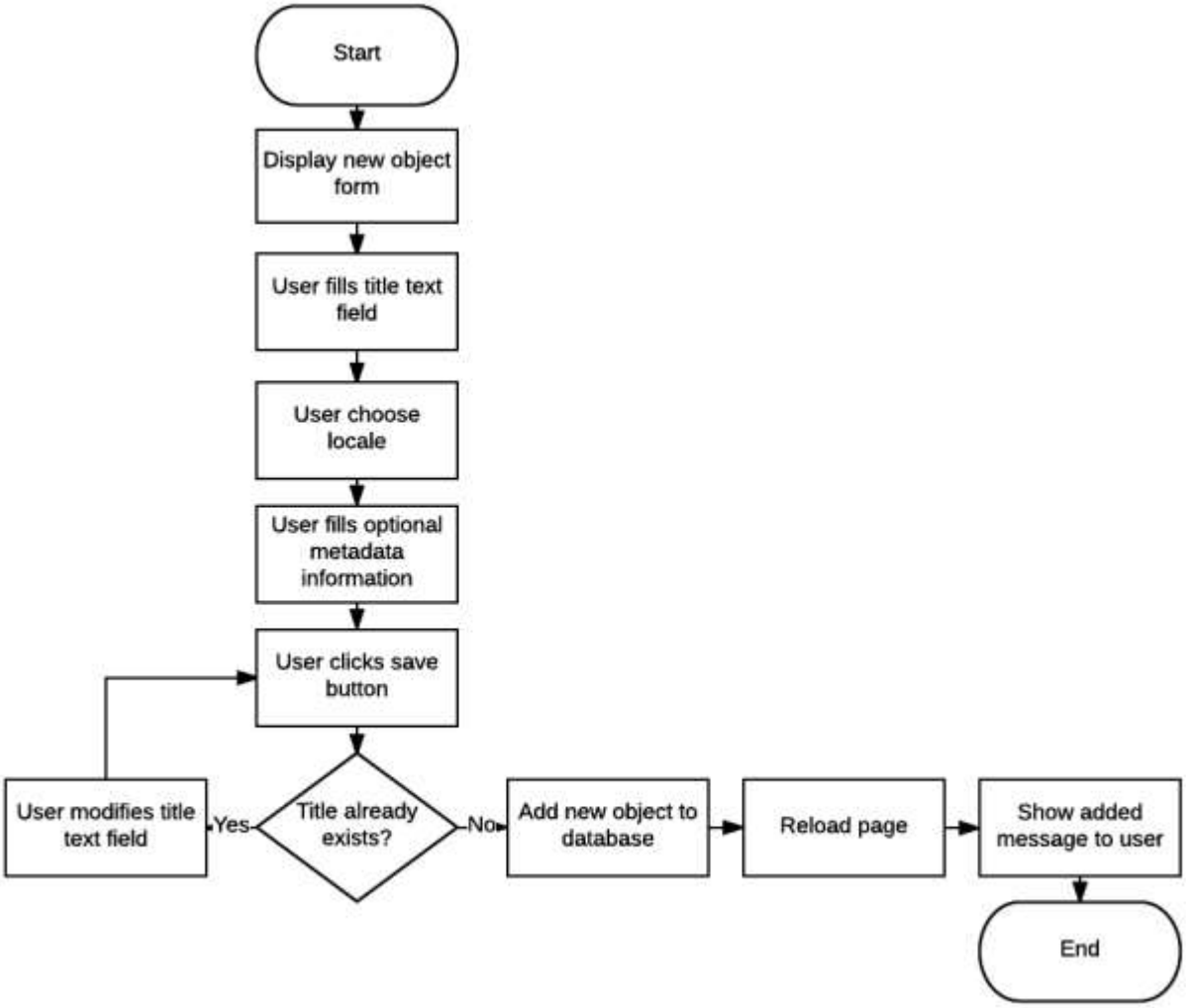


Figure 14. Collective Access add new object flow chart

The Figure 15 below shows a presentation of how a new object is created in Collective Access. The sequence diagram explains better how a new object page interacts with database.

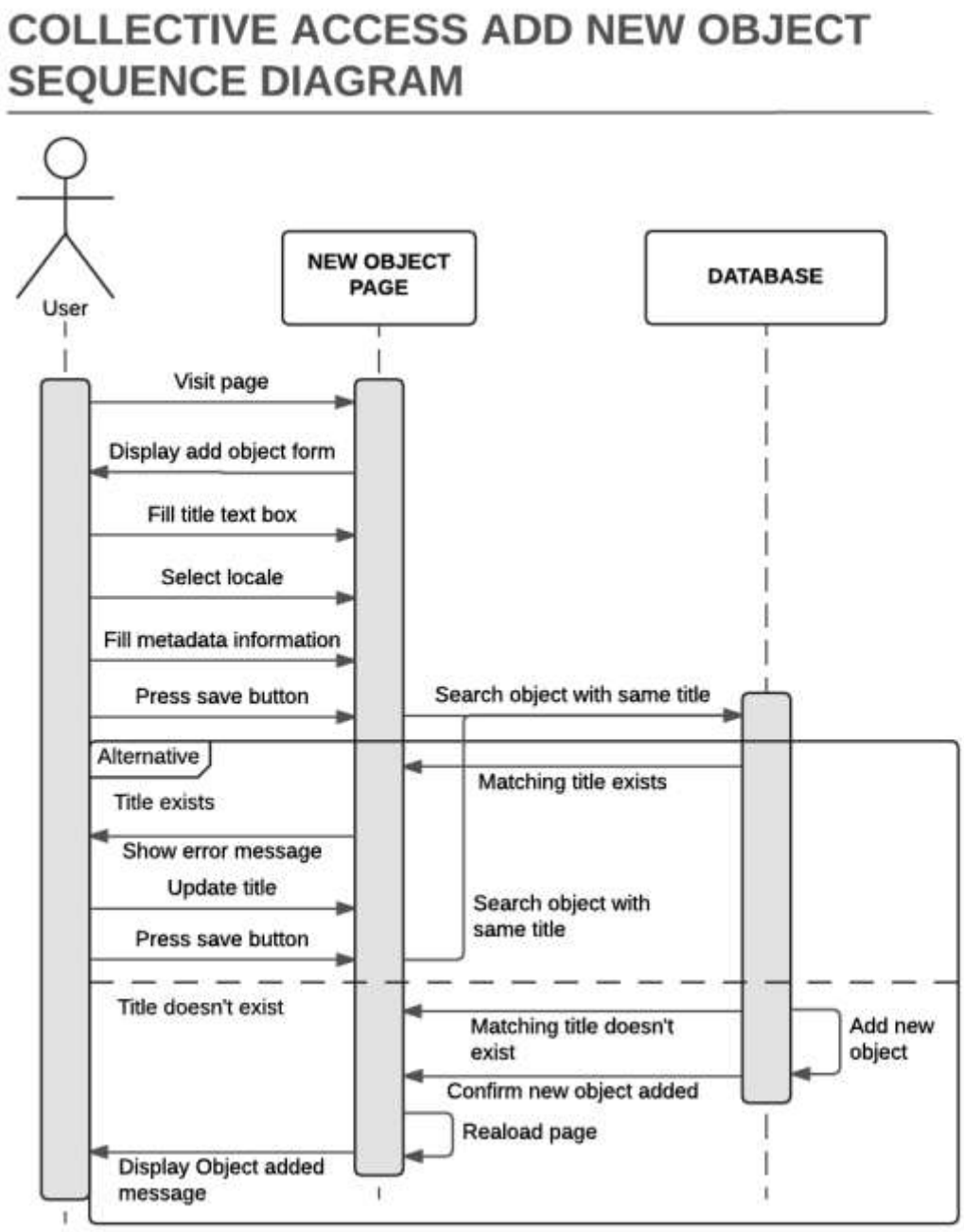


Figure 15. Collective Access add new object sequence diagram

Deleting Object

Figure 16 is a collection of screen captures from the process of deleting an object. The first part is a screen capture from basic information page, where the user is presented a delete button. After the user clicks the delete button he/she is asked for confirmation. After the confirmation the user is directed to delete page, where user is given confirmation messaged from the deletion.

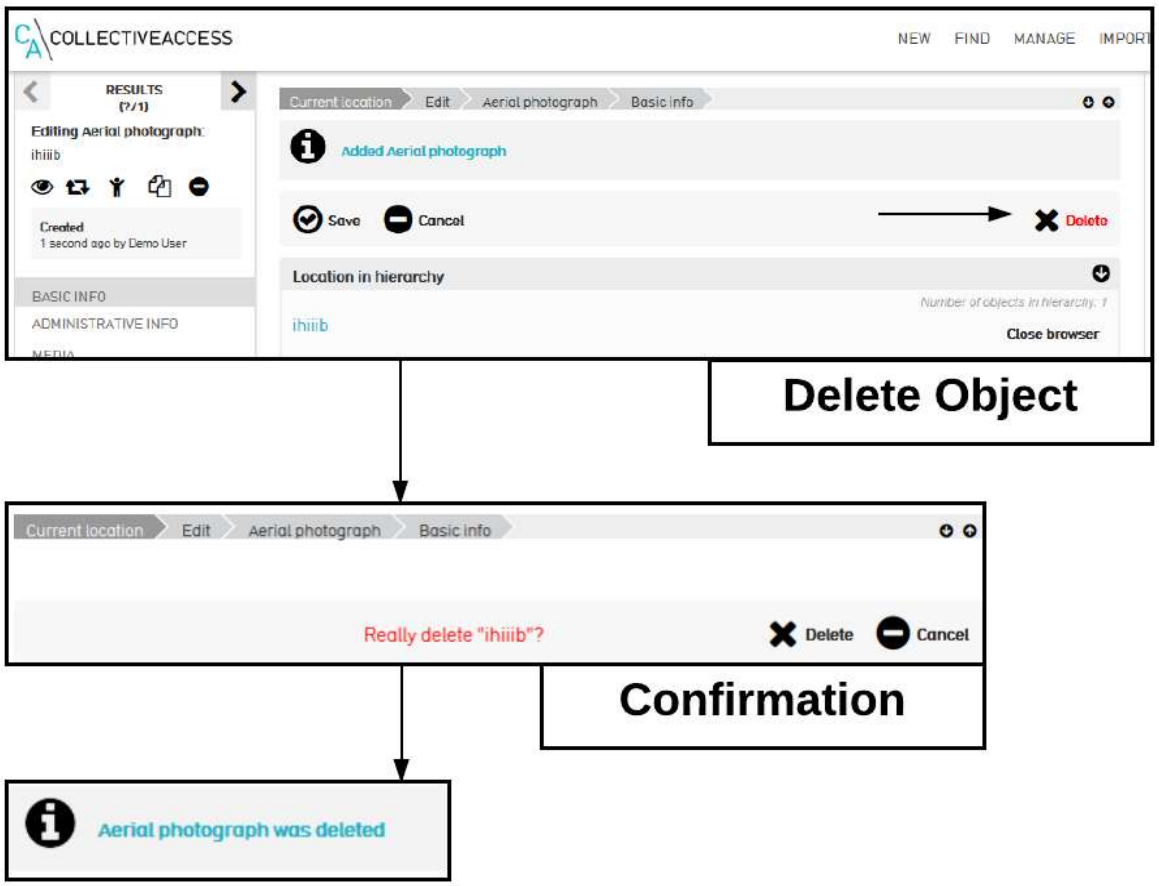


Figure 16. Collective Access screenshots from deleting an object

The Figure 17 below explains how an object is deleted. The user can find the deletion button on the object's basic information page. After the user presses the deletion button he/she is redirected to confirmation page where the user can select to cancel, or delete. If the user selects cancel, he/she will be returned to the basic information page. If the user chooses delete, he/she will be taken to the delete page which Collective Access uses to display confirmation message.

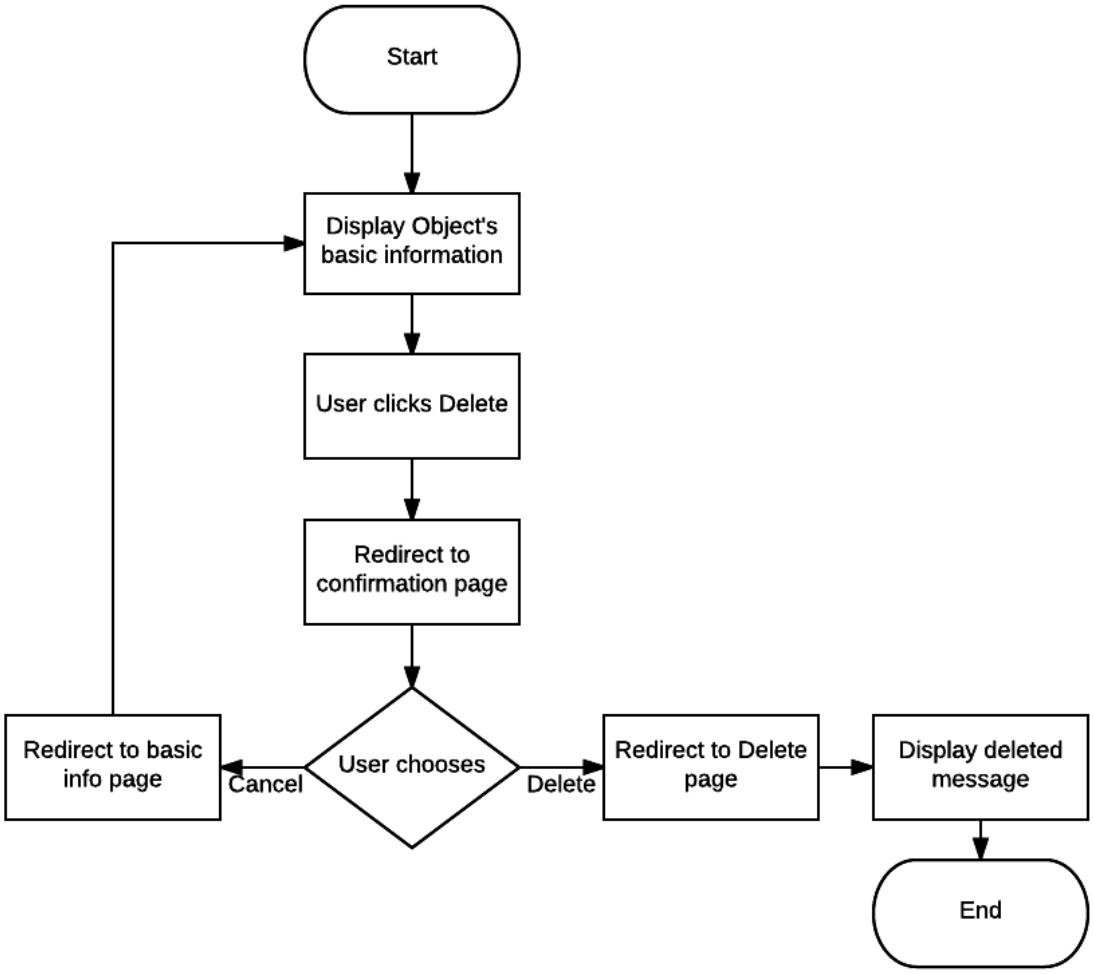


Figure 17. Collective Access deleting an object flow chart

Deleting an object is studied in the Figure 18 below. In the UML sequence diagram the process of deleting of an object is studied in more detail and it extends the flow chart in the Figure 17. It describes better how the database is used and in what order the actions are initialized.

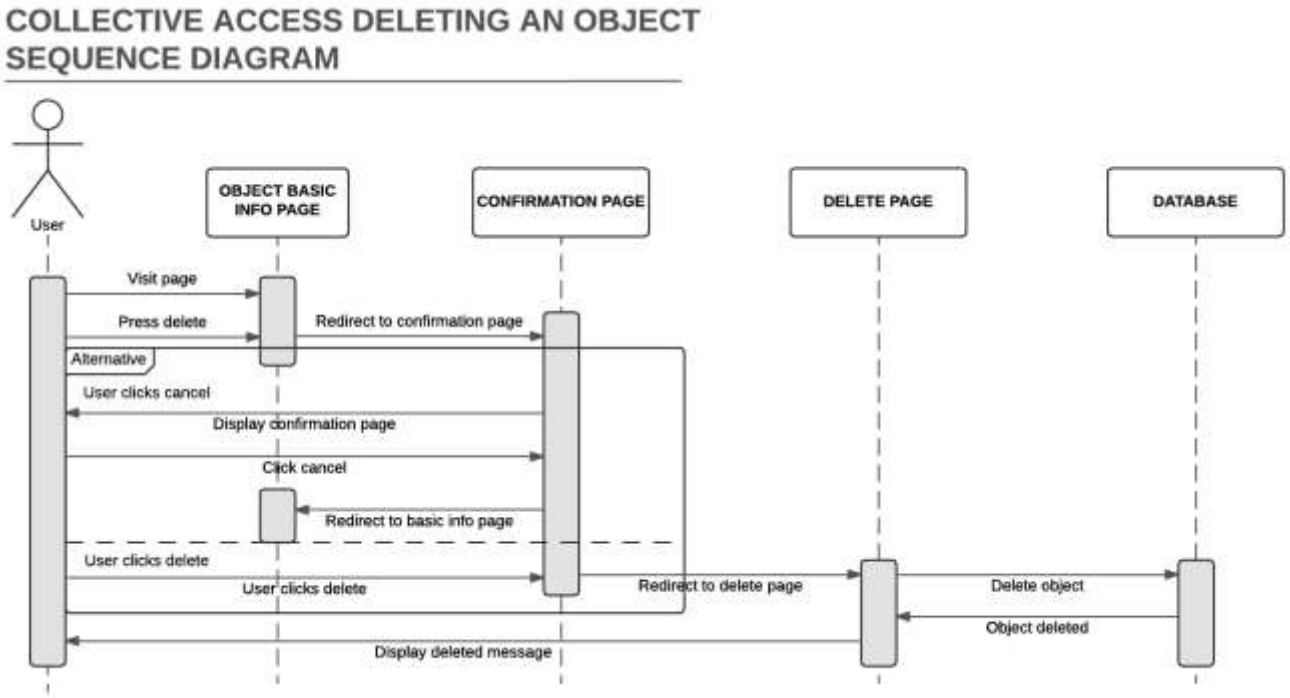


Figure 18. Collective Access deleting object sequence diagram

Summary

Collective Access is more suitable for museum collections of items and can be used in FAPOT project. For film collection itself, please take closer look at British Film Institution Collection Search part of this chapter.

Collective Access is an open-source project and can be modified to meet FAPOT's needs. However, the complete structure of the site was not clear from the demostration version. For example, the login screen is lacking "forgot password" or "create new user options", which indicates that there must to be an administration page. Adding these options would greatly reduce the work for the administrator.

3.2 Internet Movie Database

Introduction

On the IMDB advanced search, the search results and viewing film information were studied. The advanced search checks if the user has filled any part of the form before redirecting them to the search results page. Surprisingly, there were no alert or error messages. The page was only reloaded and the user was let in at the beginning of the page.

Overview

On search results IMDB shows basic information of the film and loads the poster image for the item. If the image is not available, IMDB uses a placeholder image. The search results are put in popularity order from most to least popular results by default. The user can change order by the method as shown in Figure 19. The user can reverse the results by clicking the selected sort by method again.

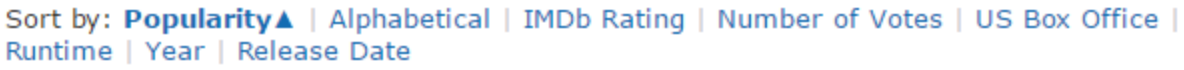


Figure 19. IMDB search results sort by methods

On top of the movie page IMDB shows the basic information of the movie. Under the basic information IMDB expands the page with additional information of the movie. This way the data can be divided to two parts. Basic information can be used twice and it can be expanded with additional information.

Figure 20 contains a site map of Internet Movie Database structure from advanced search through search results to the movie page, which is also the route, how user can find their desired movie.

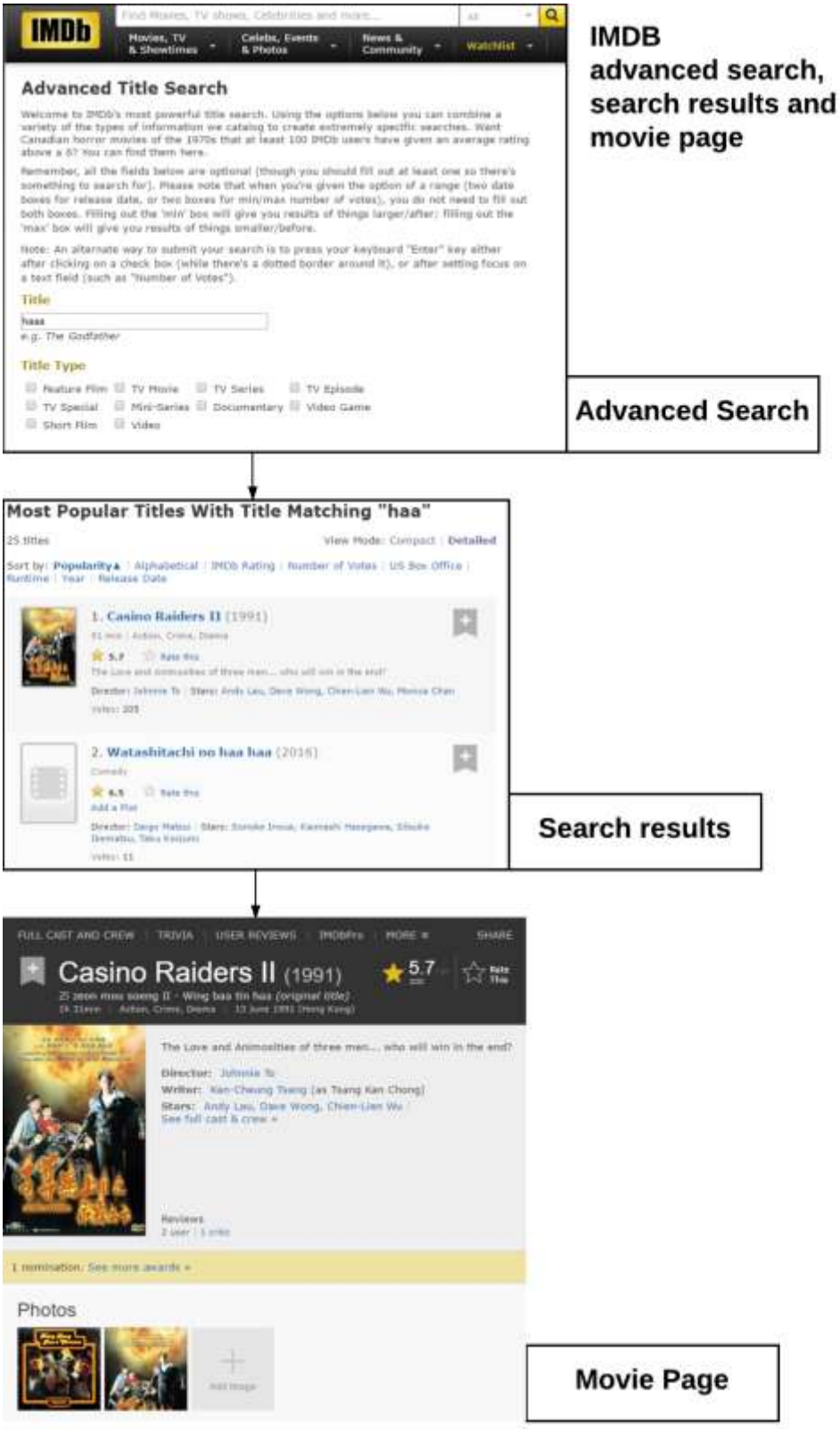


Figure 20. IMDB advanced search, results, movie page map

Advanced Search

The advanced search function in Internet Movie Database is described in Figure 21. The user can use the advanced search for a better definition of what kind of results are desired and minimizing the results to find the desired movie.

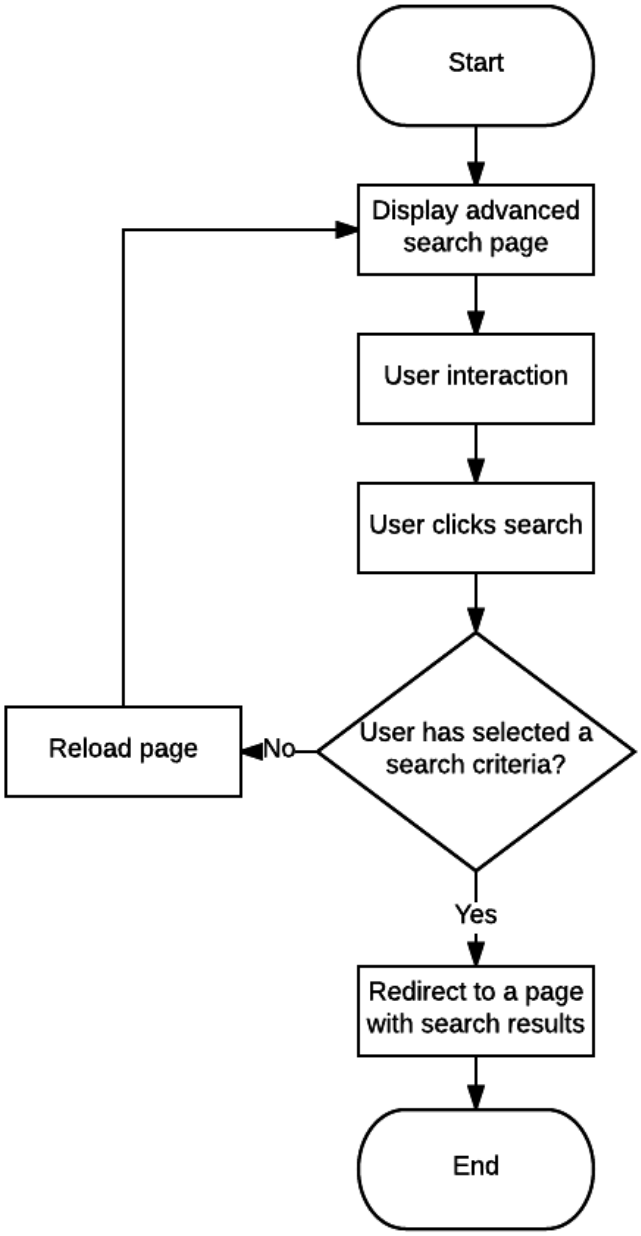


Figure 21. IMDB advanced search flowchart

In Figure 22, a sequence diagram examines further what the flowchart in Figure 21 examined. The sequence diagram describes the function of search controller in more detail as well as how it interacts with the advanced search page.

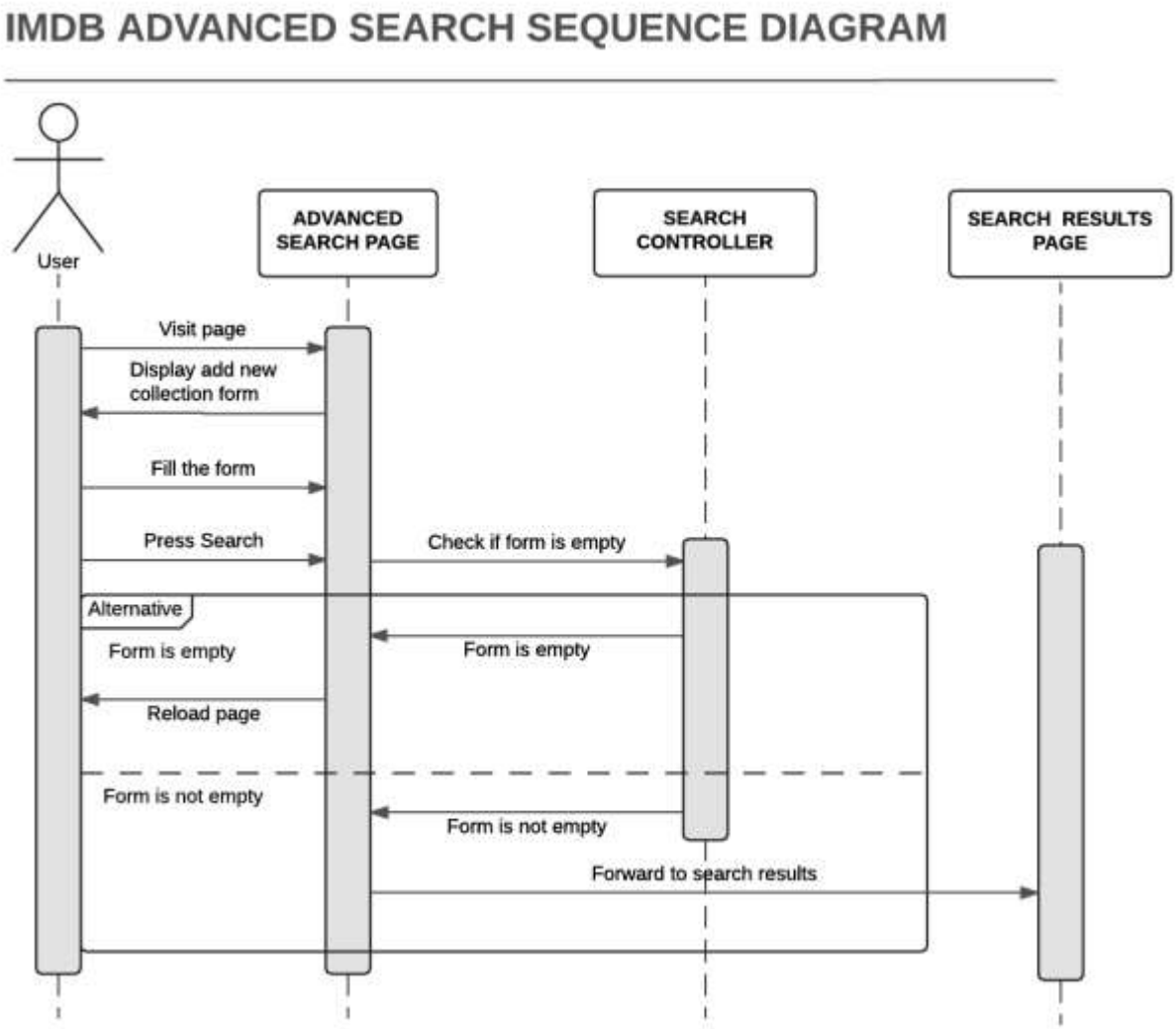


Figure 22. IMDB advanced search sequence diagram

Search Results

Internet Movie Database search takes place on the search results page where database is queried with the given search criteria, and the results are presented as a list. The search results page is examined as a flowchart in Figure 23.

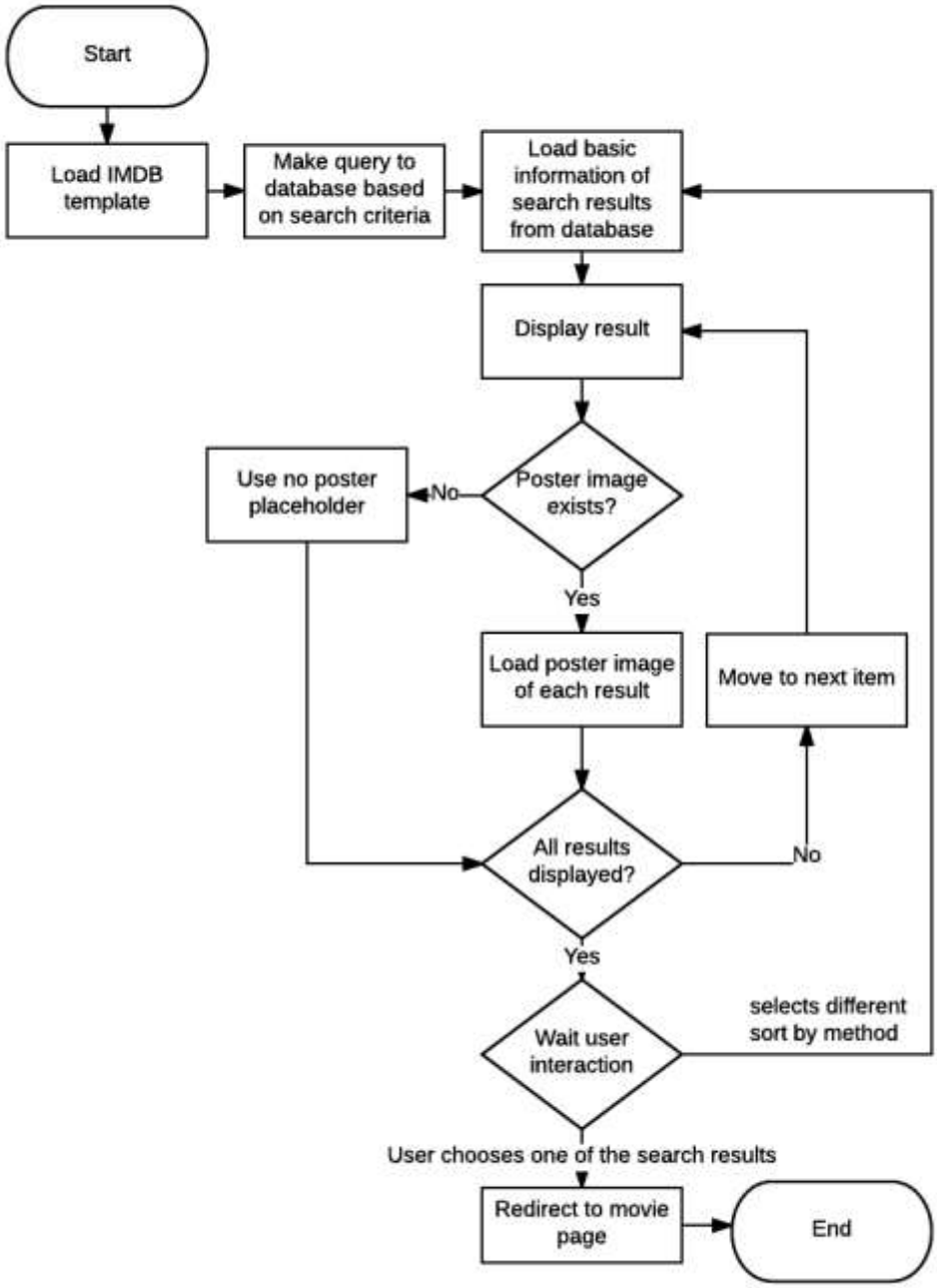


Figure 23. IMDB search results flowchart

Internet Movie Database search results page is further examined in Figure 24 below. The sequence diagram goes more into detail on how the search results controller and database interact.

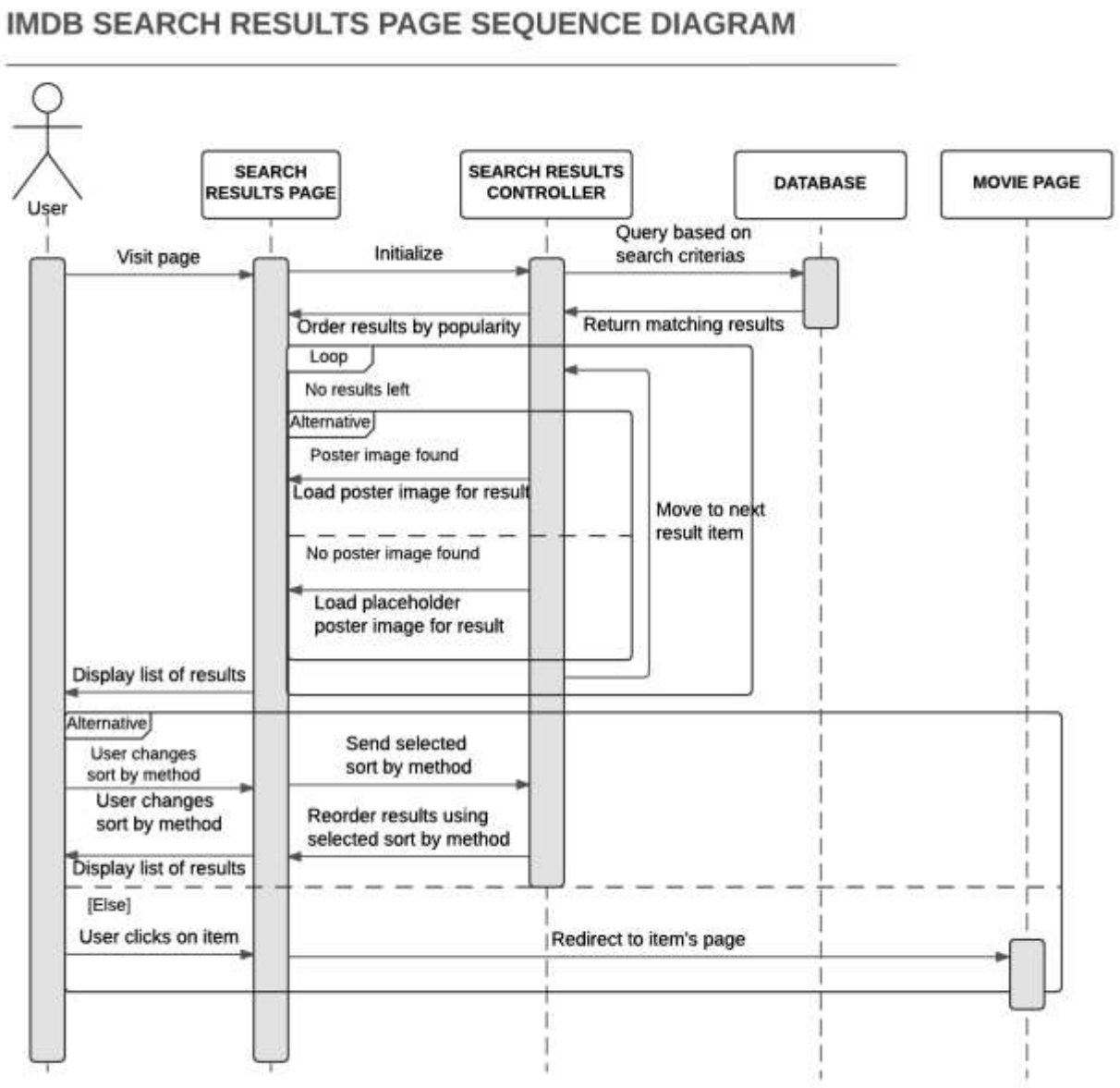


Figure 24. IMDB search results and movie page sequence diagram

Summary

This concludes the study of Internet Movie database. The research concentrated on how users interact with the website and find the movie they want. Internet Movie Database is heavy on social media integration and user interaction such as voting, commenting and movie reviews. These kinds of features are not typical for national film archives, which is discussed in the next section with an examination of British Film Institute Collection Search website.

3.3 British Film Institute Collection Search

Introduction

On the British Film Institute (BFI) Collection Search website, introduction, search, search history, search results, selection and details pages were studied. The introduction page gives an overview of the site and helps the user to understand the icons used in the collection search website, the search page is used to search the collection. The results page displays the results of the search. The details page shows details of a single item chosen from the results. The selection page shows items selected using the results pages during a user's session. The search history page displays completed searches in a user's session. If there is no search history, the page will be disabled and the link will be gray and unresponsive.

Overview

Below in the Figure 25 shows the British Film Institute Collection Search sitemap to be examined in this section. The user lands on an introduction page which explains how the website works. The user can use search, history or browse pages to find the desired objects.

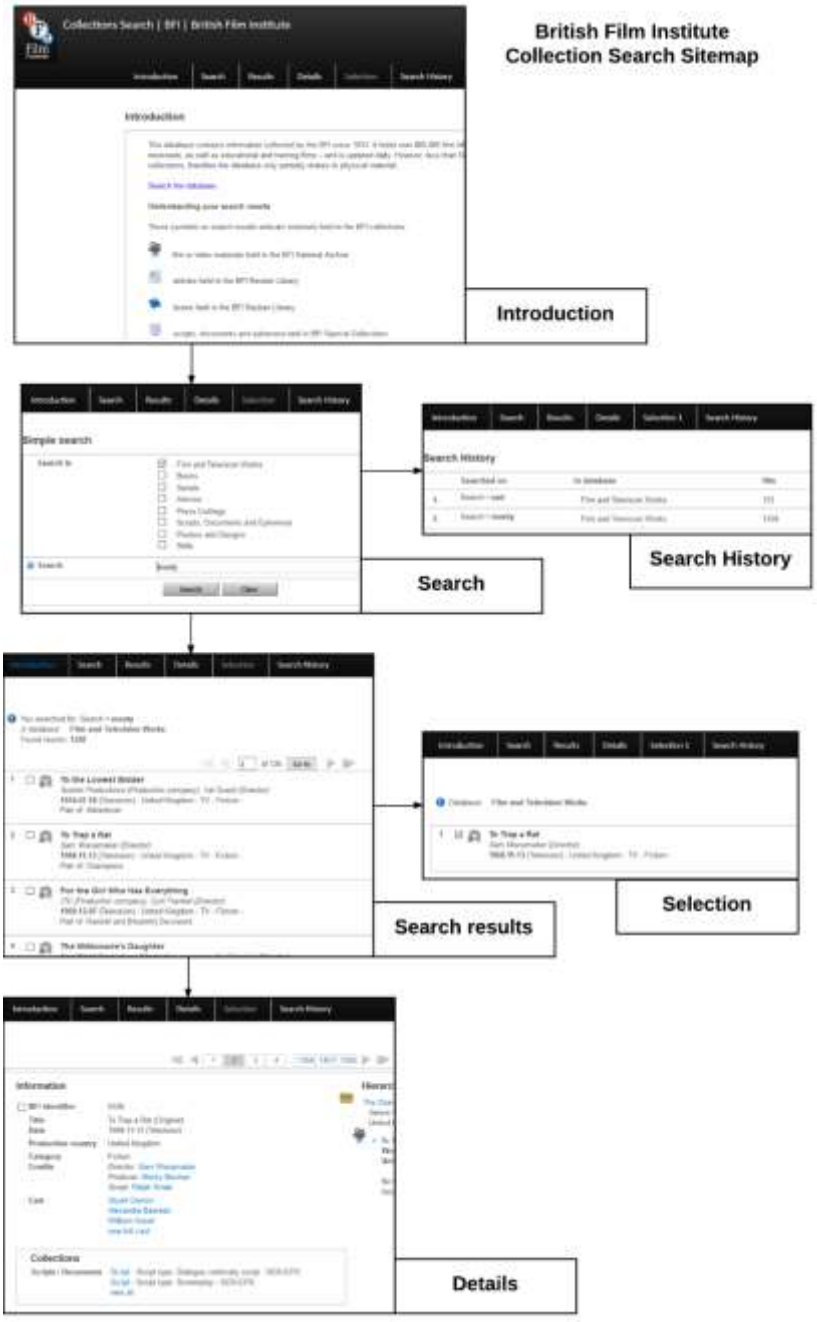


Figure 25. BFI Collection Search sitemap

The search page is used by the user to search the titles and the user is given checkboxes to filter the search to the desired categories. The screen capture of the search page is presented in Figure 26.

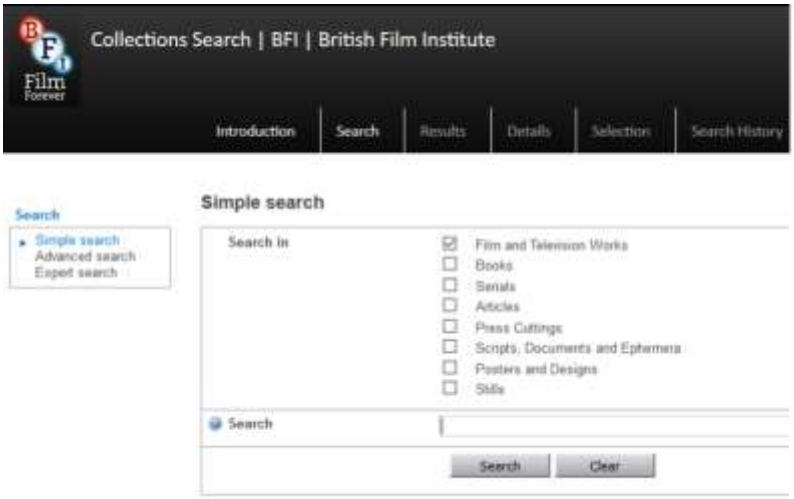


Figure 26. BFI screenshot of search page

After a successful search, the results will be presented on the search results page. The results page includes a listing of results in a paginated list. Each result also has a checkbox that user can use to select the wanted result and save it on the selection page for later viewing. A screenshot of the search results page is presented in Figure 27.

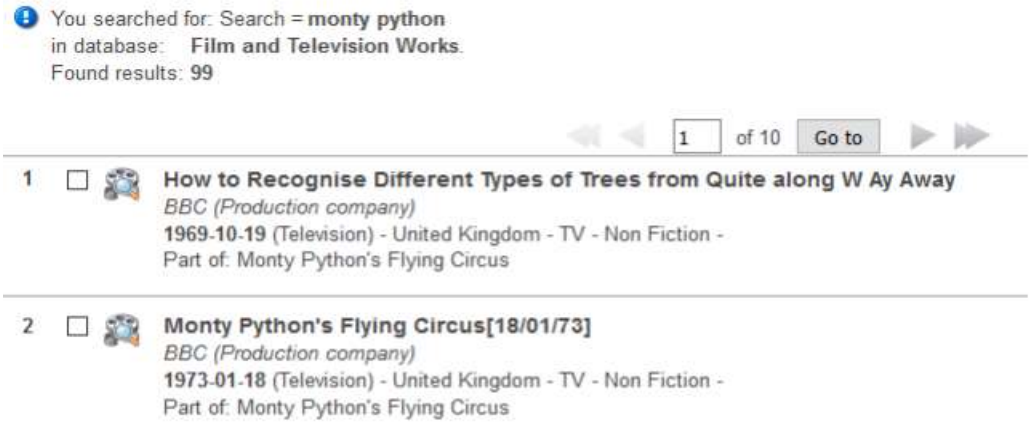


Figure 27. BFI screenshot of search results

From the results page the user can select an item by clicking the title, which redirects the user to the item page where users can view more detailed information of the item, A screenshot of Item page is presented in Figure 28.

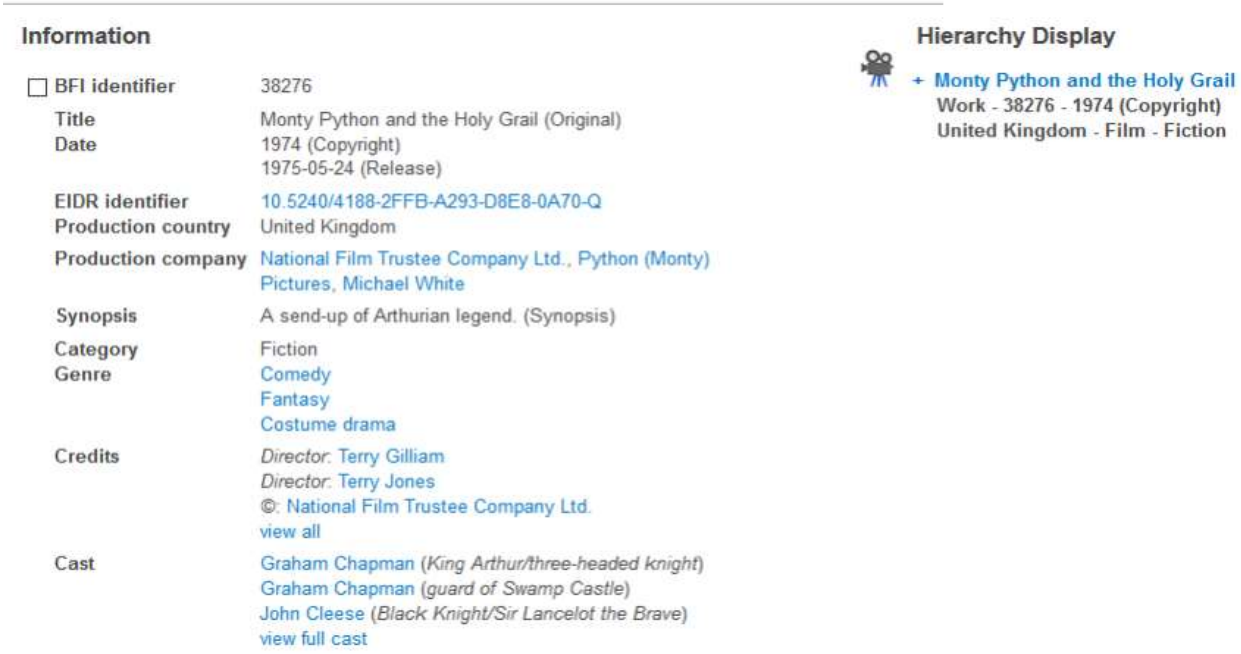


Figure 28. BFI screenshot of item page

When the user arrives on the search page, user can interact with the search form. By clicking clear, the form will be returned to empty state. By clicking search, the form is sent if the search text field is not empty. Otherwise an error message is displayed. After a successful search the user is redirected to the results page, where the user can select an item. After selecting an item the user is redirected to the item page. The flowchart of this process can be viewed in Figure 29 below.

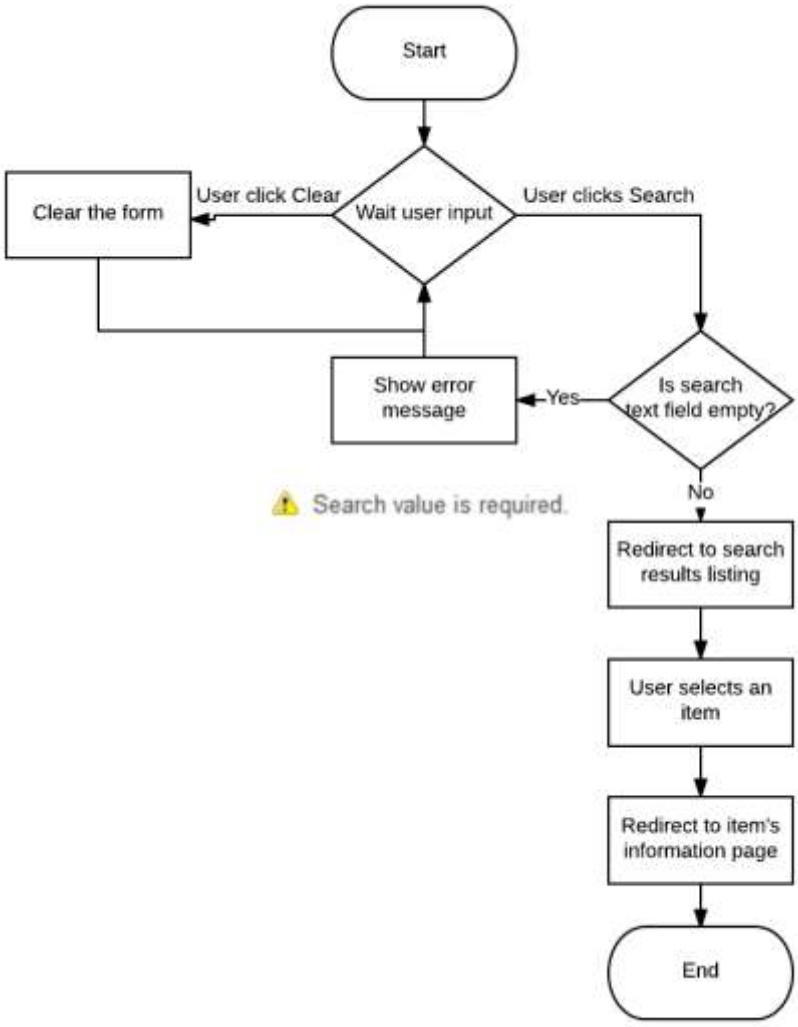


Figure 29. BFI search, search results and details page flowchart

Figure 30 , the sequence diagram below further expands the functions of search, results and item page from flow chart in Figure 33.

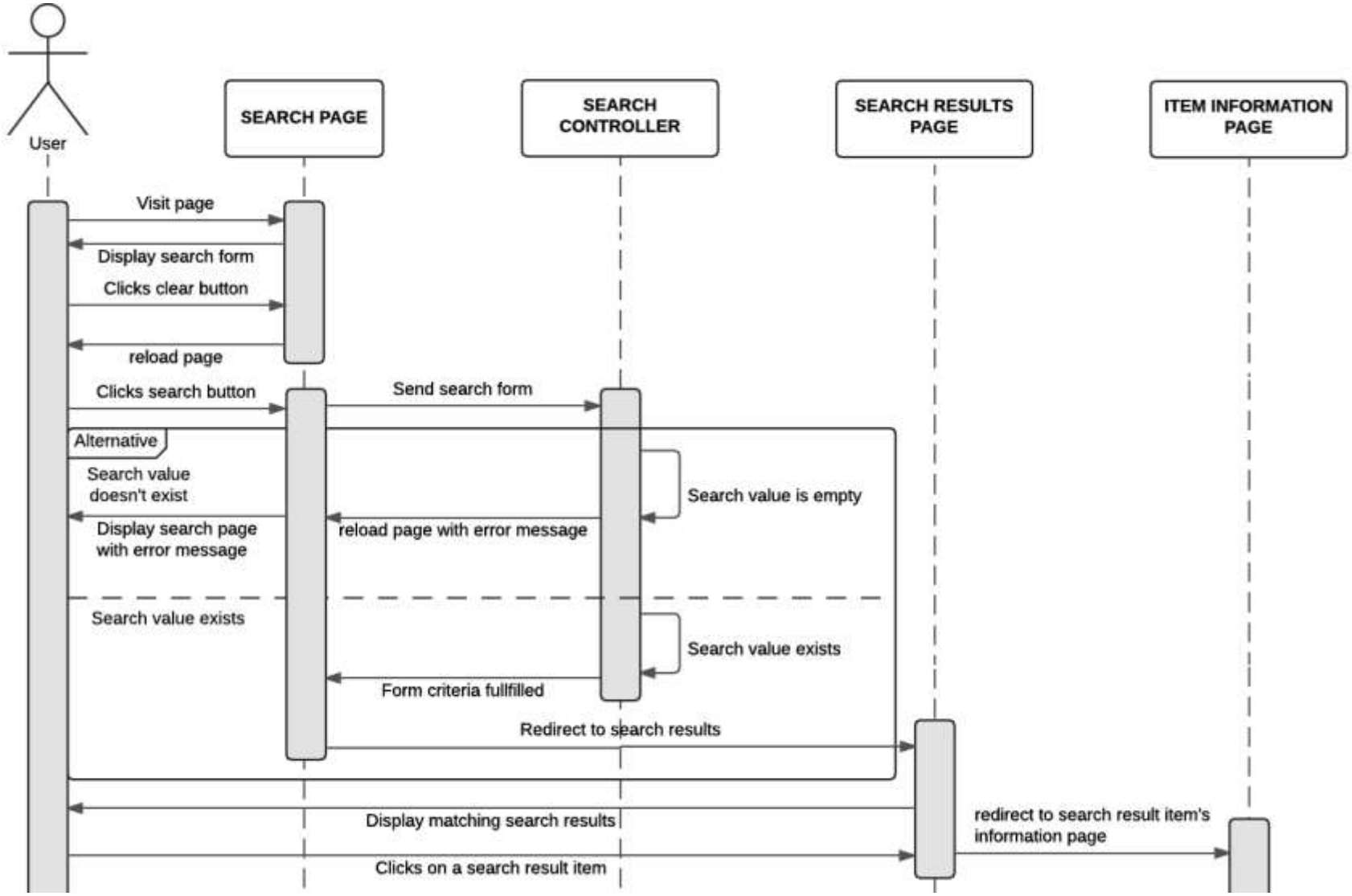


Figure 30. BFI search, search results and details page sequence diagram

Selection

In the search results page the user can check a checkbox of the item, which will add the item to the selection page and increment counter in the selection menu bar item. In the selection page the user can view all selected items. In Figure 31 below screen captures from both pages and added notes on items that are critical for this function can be viewed.

Figure 31. BFI search results and selection screenshots and notes

The process of selection in the search results and selection page are further explained in Figure 32 below. The flowchart explains in what order each action is made and gives an overview of the user experience.

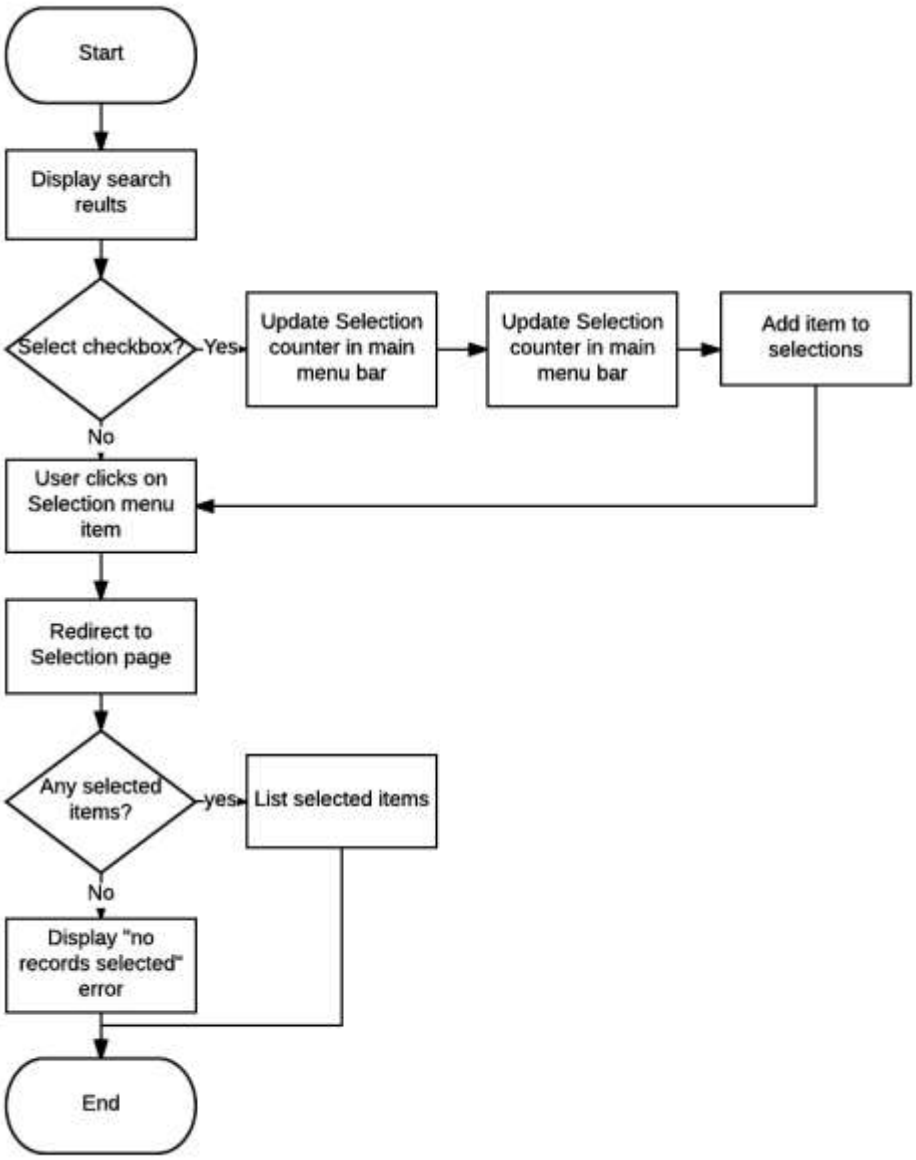


Figure 32. BFI Selection flowchart

The Sequence diagram in Figure 33 below explains further the concept of selection described in Figure 31 and Figure 32.

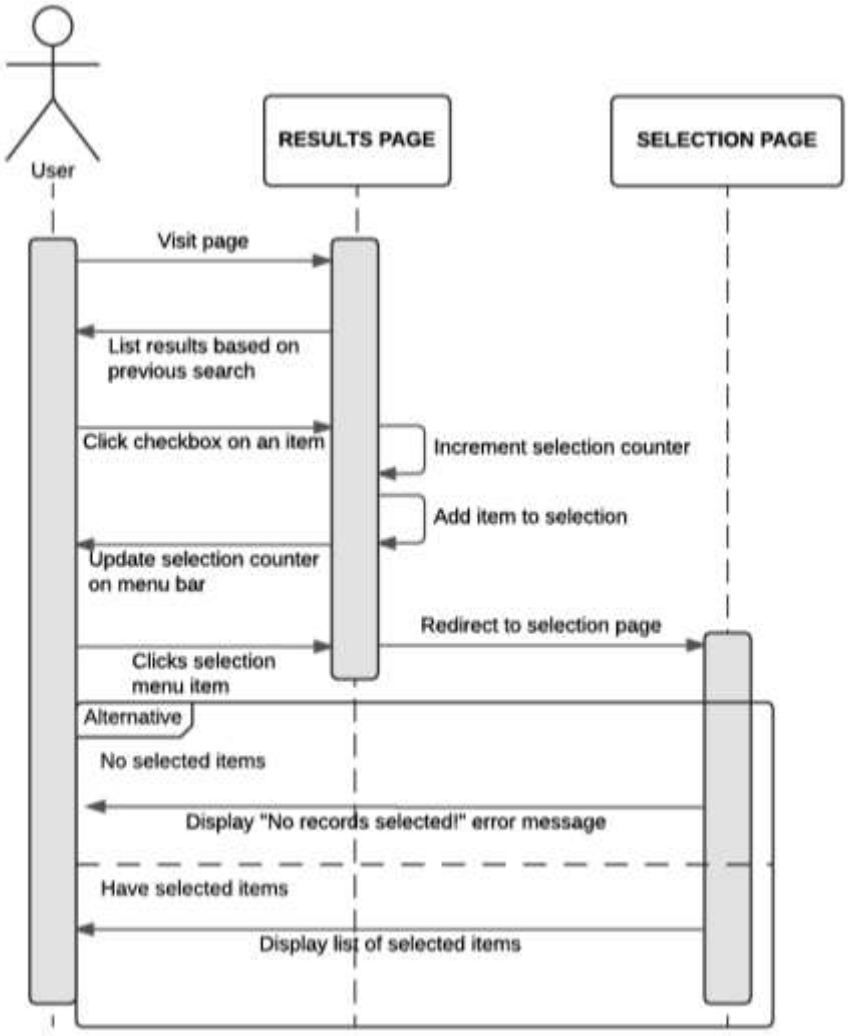


Figure 33. BFI Selection sequence diagram

History

The search history page records all searches a user has made. Every time a user makes a successful search the search item is added to the search history. Screen-shots of the process can be seen in Figure 34.

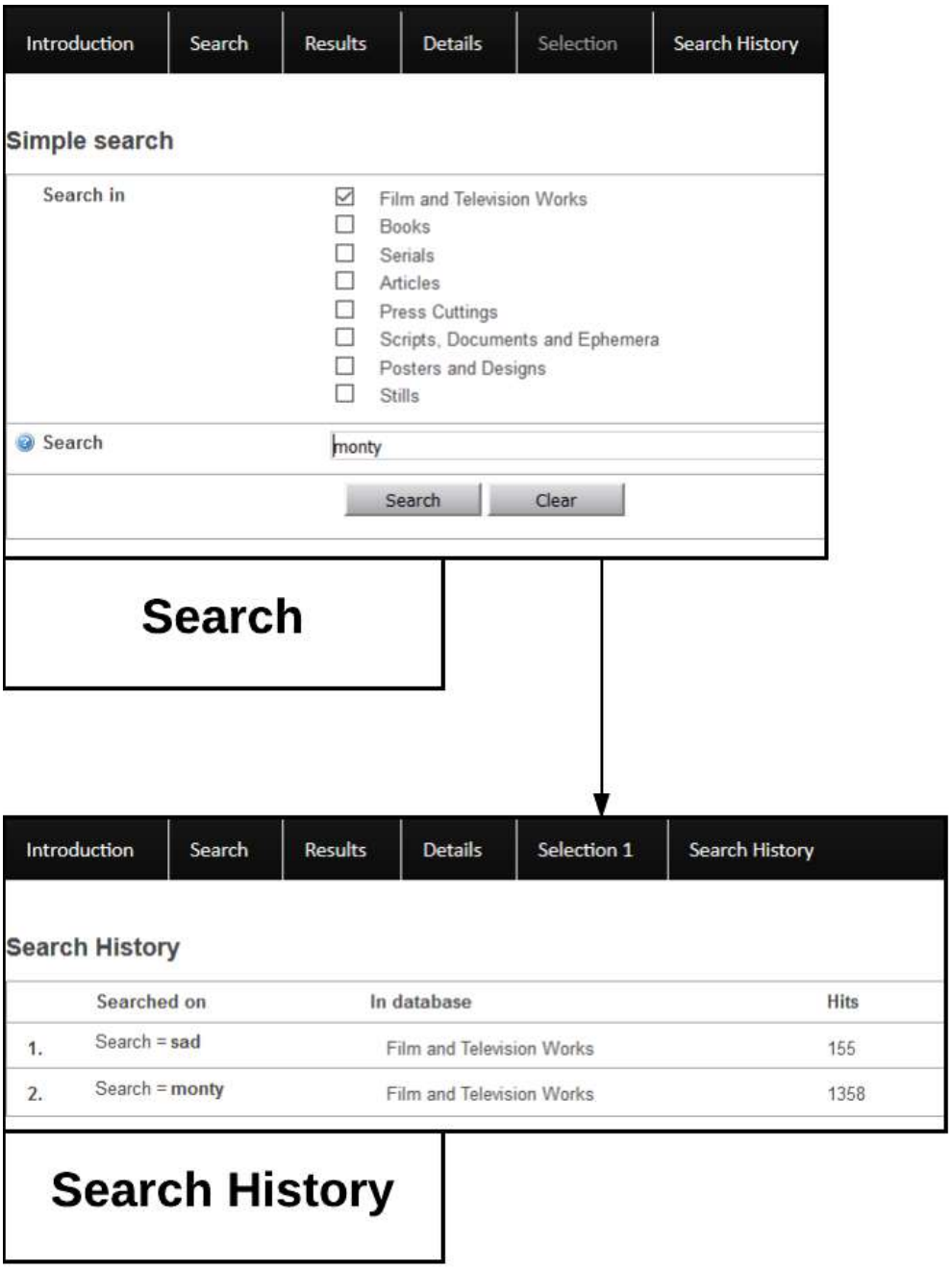


Figure 34. BFI Search history screenshots

The concept of search history is explained in the flowchart below (Figure 35) . The flowchart explains how the menu item reacts to the state of search page and in what order actions can occur.

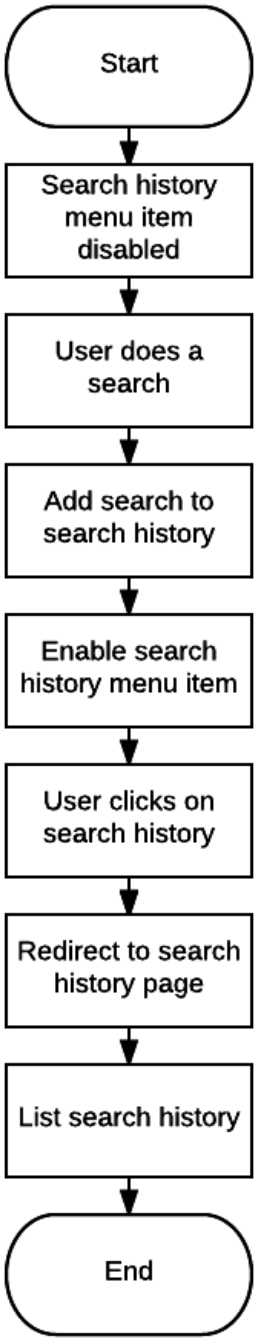


Figure 35. BFI search history flowchart

The sequence diagram in Figure 36 goes more into detail of the functions of search history in British Film Archives Collection Search page.

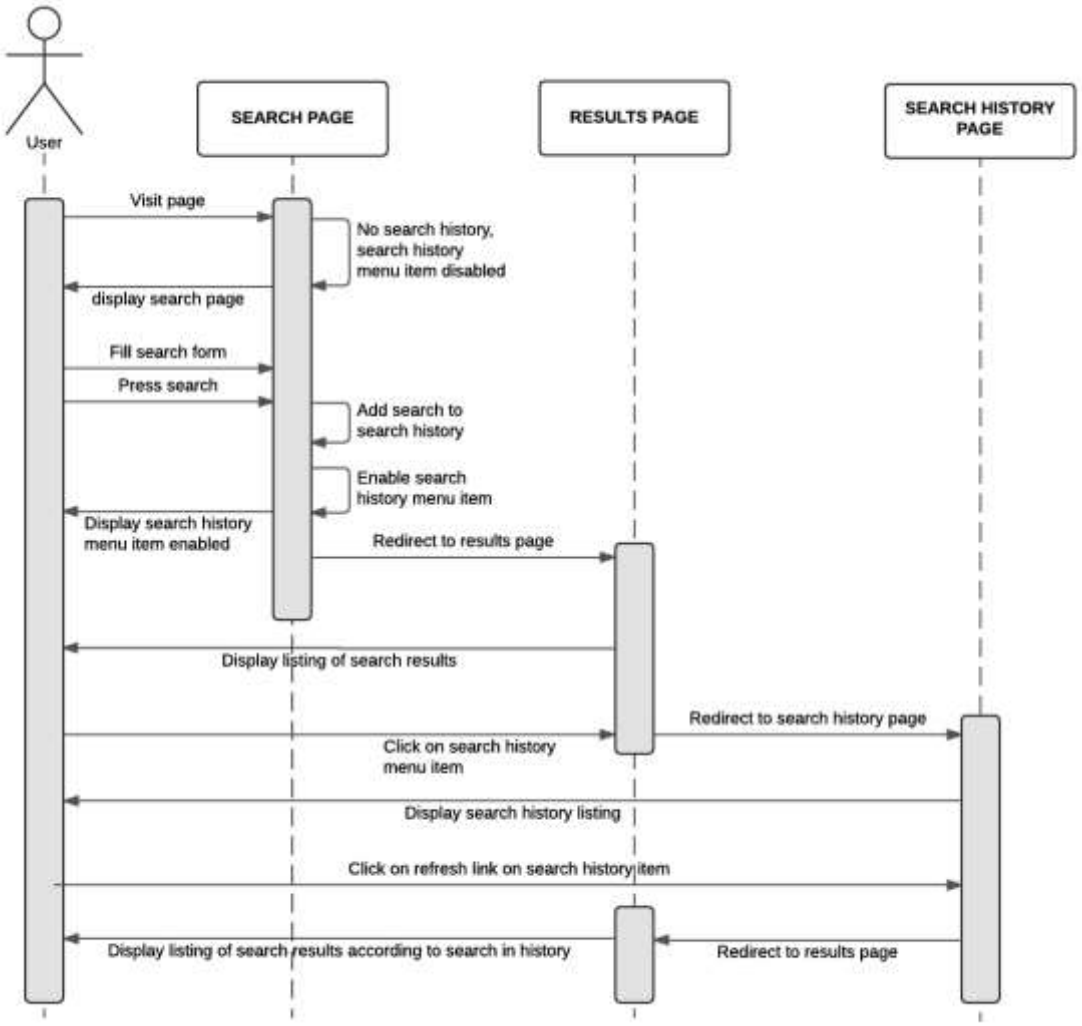


Figure 36. BFI Search history sequence diagram

Summary

The BFI collection search website remembers everything a user has done. The user can select an item from multiple searches and go later to the selection page and everything is there. It also remembers the user's search history and lists it to the user on the search history page. The structure is like on the other sites. There is search, results and detailed page where all information on the item is available. Films contain basic information that is displayed on top of the details page and on search results page. On Details page there is additional metadata presented below basic information of the film. This concludes the techno-logical research in this report. In in next section the research is summarized as a proposed system.

4 Results

The results chapter discusses the results of the reverse-engineering of the previous chapter.

4.1 Outcome of Research

By researching how film archives and such organizations have concluded similar projects in the past can help to find useful technologies, readymade solutions; it also helps to learn and understand how to succeed in the project.

By researching different standards an overall look at the structure of standards is gained, which makes it easier to compare it to the requirements gathered from FAPOT and see which standards would work best.

Internet Movie Database

IMDB has made a lengthy advanced search page, which can be used to get ideas on what and how to search from FAPOT collection. IMDB also offers search results in a basic list or with basic film information. The user can already have a good view on what the movie is about, who was involved from just looking at search result listing. The same basic information and format are also used at the beginning of the movie page. On the movie page IMDB presents additional movie information below the basic information.

Collection Access

Collection Access is a functional collection managing solution, free and open-source. Administration side of Collection Access was not available in the demonstration website. If Collection Access is chosen, it should be installed and investigated further.

British Film Institute Collection Search

The search in the British Film Institute Collection Search was well thought out and works. Finding the wanted items was easy and the website was easy to understand. Administration and back-end were not available for research naturally; however, a similar front-end could be easily developed to support the back-end after the requirements are locked and the development is starting.

In BFI Collection Search the user was offered search to find wanted items. The website saved user's session impressively. Changing to browse page showed the old results after changing the website, and the search history was found useful.

4.2 Front-end Features

For front-end most usable features was found in the British Film Institution. British Film Institution's website was well thought out and easy to use. Some features found in IMDB website were also found useful.

Browsing items and search results

Most collections offer browsing feature where the user is able to look through the items in collection usually in alphabetical order. The lists were usually also available to sort in numerous ways, and it would be beneficial to develop filtering function to the list since collections can be large. The lists were divided into separate pages using pagination. Listing usually shows basic information about the item.

The same kind of listing feature can also be used to present the search results. The listing features should be designed bearing this in mind to minimize doing same thing twice.

Search History

In the British Film Institution collection, the search webpage had a search history page. It can be useful when the user is doing more complicated research and might want to go back to the previous results. The search history page could also be extended to include film or item pages that the user has visited.

Introduction Page

The introduction page can be used as landing page for a user that gives information about functionality and practices in the collection site.

Search

Every collection and movie database website visited was built heavily around search. Search feature is usually the fastest way to find the desired item. Most sited had search bar on top of the website or a separate search page. The search page usually

was quite simple. The most common search parameter offered was title. Also, genre or category were offered often. More complicated search pages were separated from the search page and were usually called as advanced search.

In search for the page the developer should consider checking user's input; the minimum checking should be conducted to see if the search form is empty or not. Regular expression checks can be beneficial.

Item page

An item or film page is usually where the user is heading to. It was common practice that the same basic information from the search results or browsing list was repeated on an item's page and below the basic information there was additional information about the item. This was similar to how EN 15744 and 19507 standards work together.

Selection

In the British Film Institute collection search there was a feature called selection. The user could check the checkbox next to an item on the browse page. After checking the checkbox, the item was added to the selection page. The selection also had a counter to count the number of selected items. The counter was updated in real time next to the selection menu item in menu bar. Inside the selection page the user could find a listing of selected items.

4.3 Administration Features

For the back-end most usable features were found on Collective Access demonstration website. The back-end can benefit from the same features than the front-end. Administrators need a way to access the data, and the same search and listing features work for this task as well.

Login

Login pages are used to keep unauthorized users away from webpages. In Collective Access webpage the login page was simple. It contained textboxes for username and password. After clicking the login button, a form was sent and if credentials did not

match a simple error message was displayed. If the credentials were correct the user was redirected to the main page.

The login page could use the “help” and “create/request new account features”.

Adding collections and items

Collective access had features to add new collections. It is used through a form with the necessary information. Everything else in the form was optional except the title. The title was also only information that was checked after pressing the create button, and if there was already an existing collection or item with same title name, an error message was shown.

Editing collections and items

Collective access had also a feature to edit items, which worked similarly to adding. The form was same with pre-filled existing information for editing.

Deleting collections and items

Collective access also had a feature to delete edit collections when finding the wanted item in the listing and pressing delete. The item would be deleted and user would be redirected to a refreshed page where the item was missing. This concludes the technological research and the design of the system will be discussed in the next chapter.

5 Design

The design chapter goes more into detail about the actual system design, features and technologies.

5.1 System overview

The system was planned to contain two parts: the front-end and back-end. The front-end contains the user interface and user experience elements for normal users and admin users. It provides an interface for users to login, logout, view, edit, add, list, search, filter and remove data. The front-end communicates through HTTP or HTTPS protocol with the back-end. The back-end contains the business logic including database, queries for database, architecture for communicate between user interface and database and regular expression checks for input. Minimum viable requirements for browsing film archive are listed below:

- display
- search
- search results
- history
- browse
- item page.

Figure 38 illustrates a system diagram with the planned technologies for the system. Bootstrap front-end framework will be used for faster user interface deployment. Laravel framework will be used for template and database interaction.

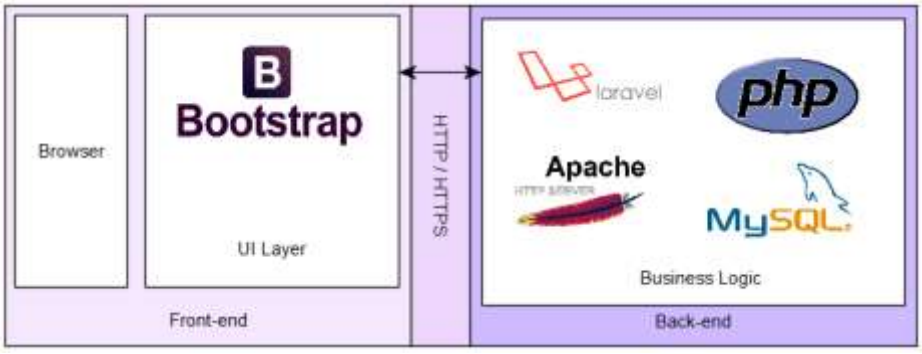


Figure 38. System diagram with planned technologies.

The system can be developed with any or the current web development language, e.g. PHP, Python, Ruby, C#, Visual Basic or Java. It is currently unknown where the project will be deployed and who will take care of the development. It is also unknown where the data will be placed. It is assumed that the system will use a relational database. In this research the technologies are in the background, since many questions are open and the technologies are changing fast. The system is based on research and intends to show what technologies enable to develop the system. The PHP has a good performance on compute-intensive sites when compared to Node.js and Python. The Node.js has better performance in high concurrency situations (Lei, K., Ma, Y. & Tan, Z. 2014).

The development of the system can be achieved with all the technologies that were researched. The Technologies and the hosting need to be compatible. The technologies can be chosen based on what hosting supports or host service can be selected based on the chosen technologies. For the sake of reducing the scope, PHP was chosen, mainly for strong knowledge and popularity. The usage of a web framework is well justified for a system of this size. A framework will reduce the workload greatly by having a Model View Controller architecture in place, the system will require less code and file organization to help teamwork (Sem, J. 2013). Developers are required to learn the conventions of framework; however, it will be beneficial in the end.

The Laravel framework has gained much popularity in the past years and is recommended for the system (The Best PHP Framework for 2015. 2015). Recommendation was given due RESTful routing system that helps to connect resources easily, blade template engine, Eloquent object-relation mapping implementation for interacting with the database, migration engine and well done documentation (7 Best PHP Frameworks for 2015. 2015). If performance becomes an issue, a custom-made controller for database use is advised.

5.2 Front-end

Introduction

The user interface should be created using ready-to-use frameworks to minimize work, for the ease of creating the layout, use of a front-end framework is recommended. One such framework is Bootstrap.

Overview

Figure 39 below illustrates a sitemap for the proposed FAPOT film collection page, the main goal of which is to help users to find the wanted film item.

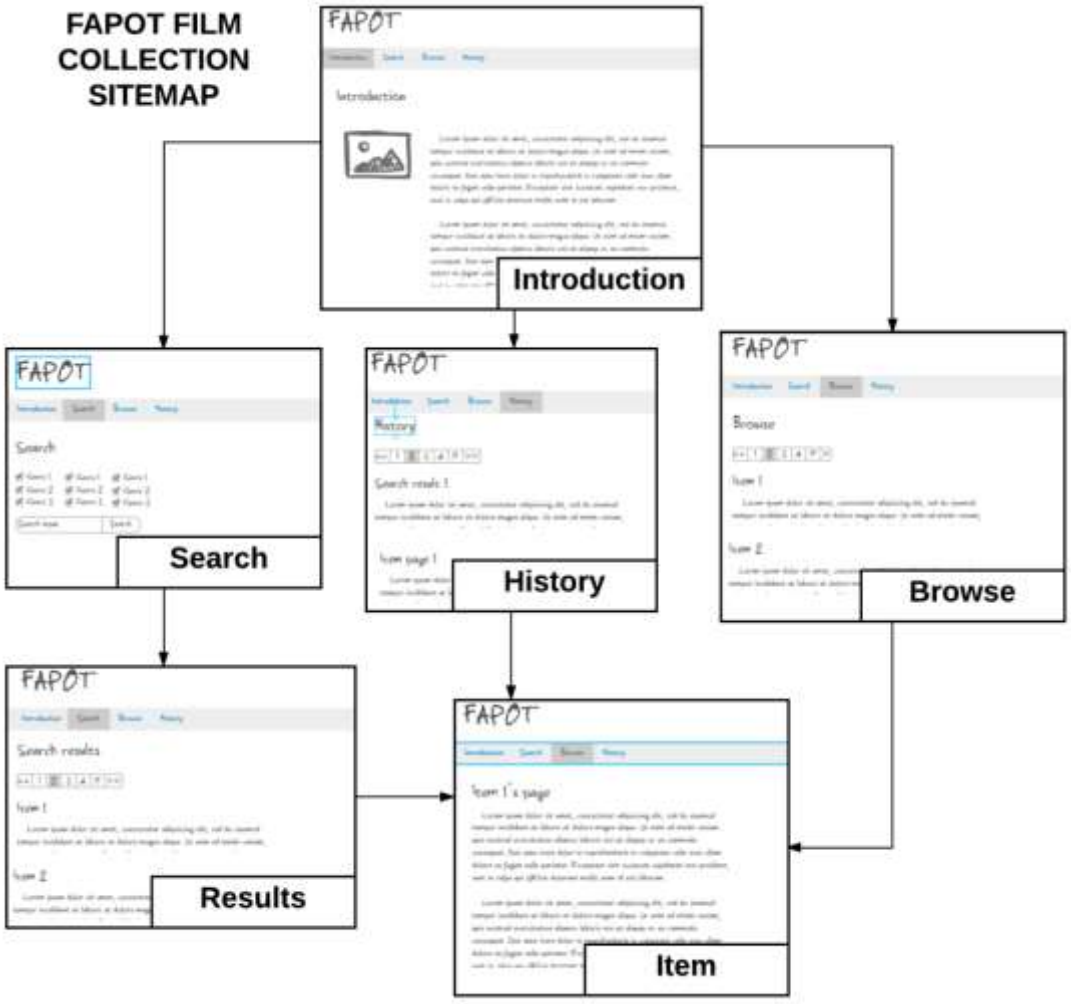


Figure 39. FAPOT Film Collection front-end sitemap

The user can find the wanted film item using three ways. Using search, browse or history page are described in Figure 40 flowchart.

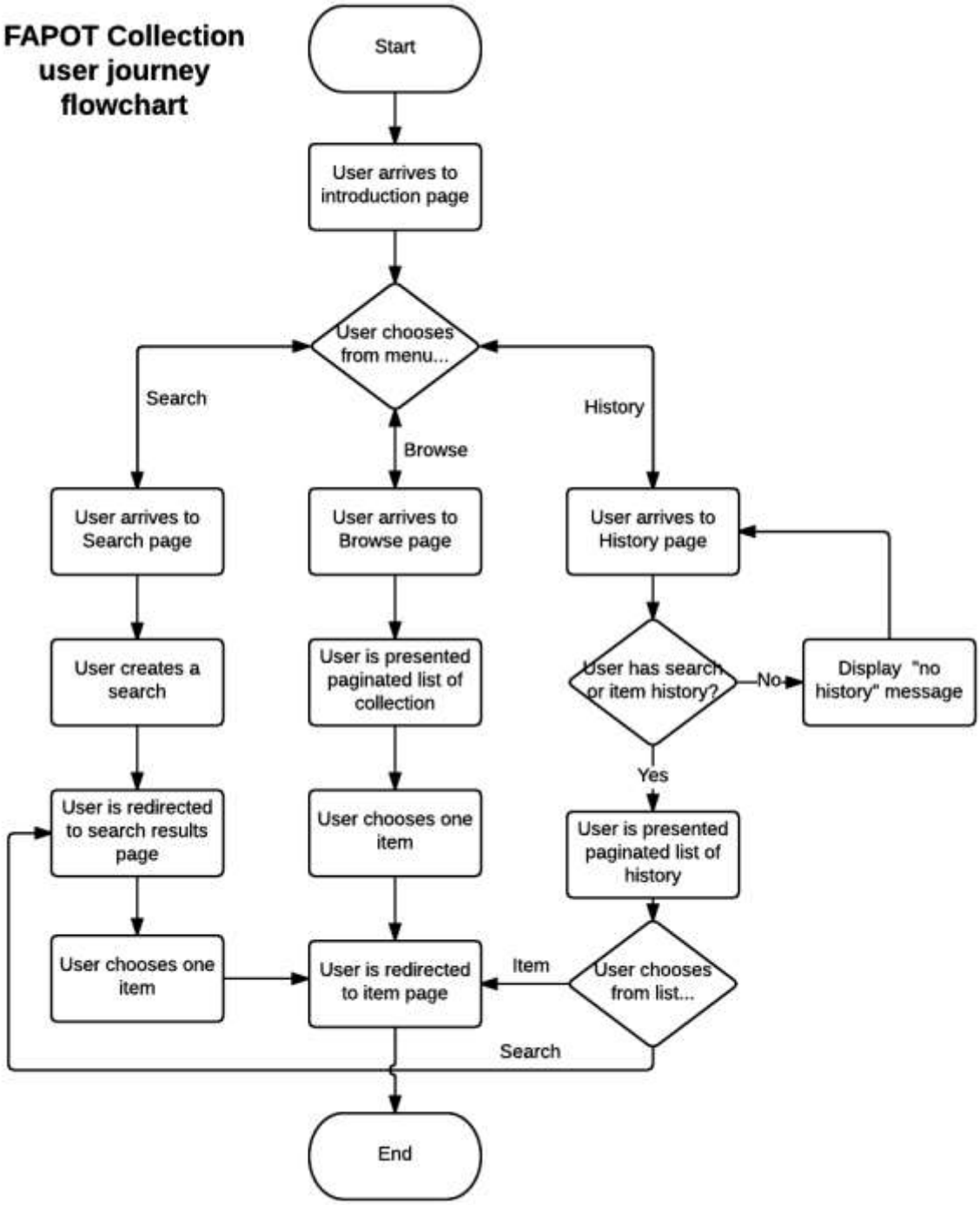


Figure 40. FAPOT Film Collection user journey flowchart

Introduction page

The introduction page is where the user will arrive and the item page is where users are trying to get. The site offers three ways to find what the user is searching: by searching, browsing or if the user has previously visited website through the history page. The search results page will display basic information of the item. The item page will display the first basic information of the item and then additional information. Figure 41 is a mockup of the introduction page.

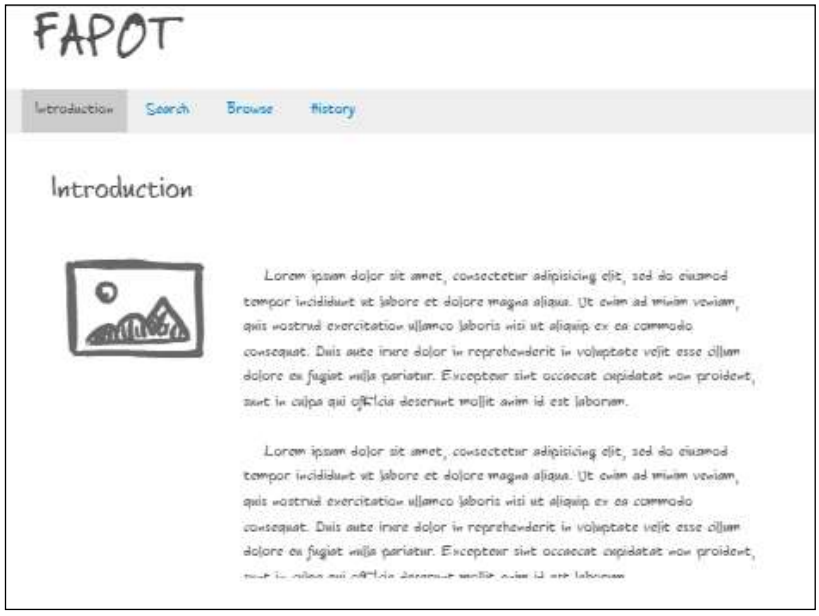


Figure 41. FAPOT Film Collection introduction page mockup

The introduction page is used to give general guidance and information about the website. It is also the landing page for users and offers menu bar items that direct users to search, browse and history pages.

Search page

The search page is used to give the user tools to search the collection. The user can search for title, keyword, genre or crew member. The search page should be expanded according to the client's needs. If the search gets complicated, it should be divided into two pages: one with a simple search and another page with advanced search. The mockup of the search page is presented in Figure 42. The search criteria of the search page should be discussed with the client.

Figure 42. FAPOT Film Collection search page mockup

Search Results

The search results page is used to show search results after the search button has been pressed on the search page. The results are listed in alphabetical order. The results have been divided into multiple pages using pagination. The mockup of the search results page can be seen in Figure 43.

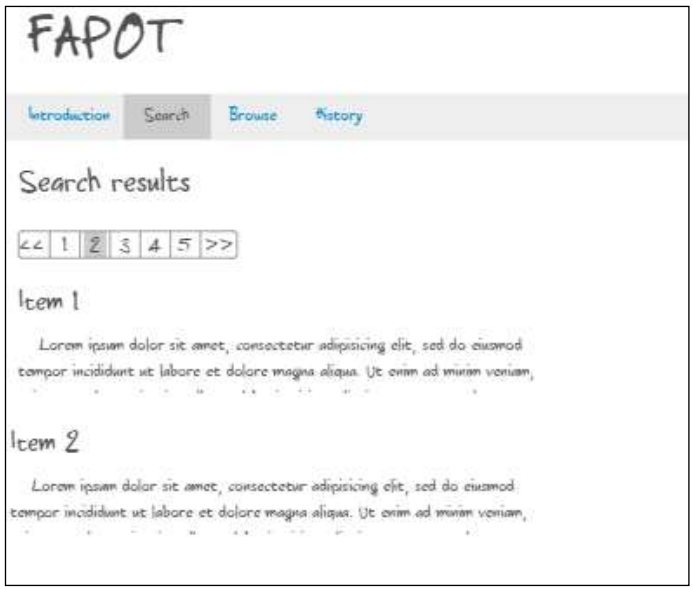


Figure 43. FAPOT Film Collection search results page mockup

When the required dataset is known there could be a more compact table view or more detailed basic information view like IMDB is using as shown in Figure 44.

Figure 44. IMDB Compact and Detailed search results

Search History

The search history page includes the history of the visited items and the user's past searches. History will be a list with pagination. One possibility is to add filters or search to the history. The mockup of the history page can be seen in Figure 45.

Figure 45. FAPOT Film Collection search history page mockup

Browse Page

The browse page is used to give users a list of all contents in collection. This example site is showing a list of items with basic information. The items have been divided into numbered pages with pagination. The browse page could include filtering and order by methods. The mockup of the browse page can be seen in Figure 46.

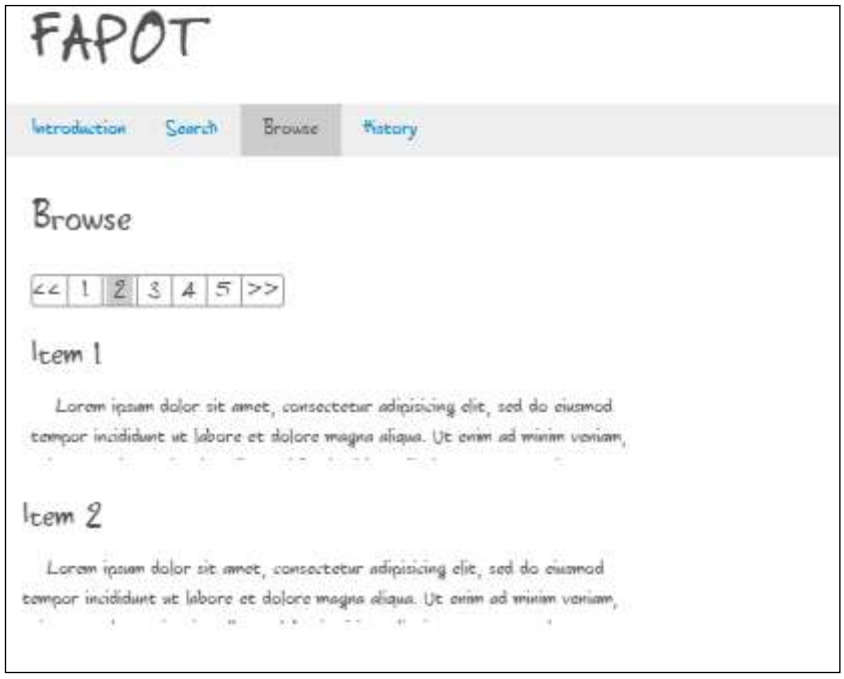


Figure 46. FAPOT Film Collection browse page mockup

Item Page

The item page will display the whole dataset of the film item. At top there is basic information and additional information under that. Hyperlinks could be used for searches. For example, clicking a certain film genre would bring up search results for that genre. The same could be applied to crew members. The mockup of the item page is shown in Figure 47.

Figure 47. FAPOT Collection item page mockup

5.3 Administration

For administration of the website there are two solutions: a separate application to manipulate database or extending the existing website. The minimum viable administration application or website should contain the following functionality:

- Display cinematographic work
- List cinematographic work
- Search cinematographic work
- Add cinematographic work
- Edit cinematographic work
- Delete cinematographic work

This concludes the part of the proposed system. Mockups and further planning for the administrative side of the system, back-end and hardware should be expanded further when the requirements specifications are available. The front-end design is usable as-is and can be modified to better accommodate the client's requirements. This concludes the discussion of the design and in the next chapter the project will be summarized.

6 Conclusion

The goals of the project were to research and design film archive management system for FAPOT. This included researching film archives, site structures, standards, technologies, software, functions and features. The design part included a user-interface design, functions and features for the film archive website.

There are many existing film archives in the world and similar systems have been developed before. Film archives and databases can be divided into two groups: private and public.

The private archives are commercial company hosted websites that get profit from advertisement and from the content. Websites like these are more user oriented, offering users entertainment, photos, information, reviews, rankings, heavy social media integration, user integration and users promoting their content. Private companies are not open about their technologies, and access to their databases are limited. For non-private use, it is estimated that getting an API access to the most popular movie databases starts from $15,000 annually (Does IMDB provide an API? 2016).

The historically and nationally important public film archives are often hosted by non-profit organizations or public sector. The main purpose on these sites is to share information and secure national film history.

Creating a new standard when multiple well recognized standards are readily available is usually a way to create more of the same. Instead, it is recommended to use pre-existing standard and add additional modules if additional metadata is required. The European metadata standards for cinematographic work, EN 15744 and EN 15907 were studied and they work in a same way as many systems that were studied, where the first visible metadata was used for identification of the item and the second set of metadata was used to give additional information of the item.

Rough enhanced entity–relationship model database schematics were designed based on available documentation on filmstandards.org website (EN 15907 preliminary version of an XML schema). However, the work on diagrams was halted until the decision on which standard will be used was made. Diagrams are presented in Appendix 4 and Appendix 5.

The technologies and the hosting need to be compatible. The technologies can be chosen based on what hosting supports or host service can be selected based on chosen technologies. For sake of reducing the scope PHP was chosen, mainly for strong knowledge and popularity. The usage of a web framework is well justified for a system of this size. A framework will reduce the workload greatly, and with a Model View Controller architecture in place, the system will require less code and file organization to help teamwork (Sem, J. 2013). Developers are required to learn the conventions of the framework; however, it will be beneficial in the end.

The Laravel framework has gained much popularity in the past years and is recommended for the system (The Best PHP Framework for 2015. 2015). The recommendation was given due RESTful routing system that helps to connect resources easily, blade template engine and Eloquent object-relation mapping implementation for interact with database (7 Best PHP Frameworks for 2015. 2015).

From researching the features on a similar system, the most important point from the user interface standpoint was how the users find what they are looking for. The user should be able to find what they are looking for as easily as possible. Traditional search was a feature in every system that was studied. Some systems included advanced search that let the user filter results with more precise criteria. Some of the researched systems featured search history, in case the user wanted to backtrack their past searches. Additionally, users could select results to add to their collections for later use.

In the designed system, a user would be given the option to search, go through items in alphabetical order if they are not exactly sure what they are looking for and through their history, in case they want to find item that they had searched for in the past. These three paths will ensure that a user will find the wanted item. Sitemap and useful features were designed for the system.

In the project, important take-away was not to reinvent the wheel but to see how much re-verse-engineering can help with project like this. By studying and drawing diagrams of the websites with similar features that the FAPOT project wants to achieve can give a good outlook on what kind of features are beneficial for the system.

The project succeeded to research different film archives and databases, their structures, features, functions and to document all process. Software, such as the Content Management Systems were studied and useful features were researched. Rough database schematics were designed based on the standards. The private databases like IMDB.com were not open to share information about their technical solutions and for this reason exploring database schematics was not possible. Due to limited time and missing requirements specifications work on algorithms, queries and administration site design were not completed.

The end result of this project can be used as part of the bigger system that FAPOT requires. The results will give good overview on film archives, useful features and standards. The diagrams can be used to further investigate what the client's needs are. For further development the work on algorithms and queries should be continued. The administration site design gives a starting point for a complete system.

References

7 Best PHP Frameworks for 2015. 2015. Article on LinkedIn.com website. Accessed on 11 November 2016. Retrieved from https://www.linkedin.com/pulse/7-best-php-frameworks-2015-winspire-web-solution

Alternative Interfaces. N.d. Page on Internet Movie Database's website. Accessed on 26 July 2016. Retrieved from http://www.imdb.com/interfaces

Beal, V. N.d. API - application program interface. Definition of the term API in Webopedia online tech dictionary website. Accessed on 3 December 2016. Retrieved from http://www.webopedia.com/TERM/A/API.html

Archive. N.d. Definition from Cambridge's dictionary. Accessed on 22 November 2016. Retrieved from http://dictionary.cambridge.org/dictionary/english/archive

Can I use IMDb data in my software? N.d. Page on Internet Movie Database's website. Accessed on 26 July 2016. Retrieved from http://www.imdb.com/help/show_leaf?usedatasoftware

Collective Access. N.d. Main page of Collective Access CMS, features page. Accessed on 26 November 2016. http://www.collectiveaccess.org/#specs

Collective Access Live Demo Login Page. N.d. Accessed on 26 November 2016. Retrieved from http://demo.collectiveaccess.org/index.php/system/auth/login#

Collection Space. N.d. Collective Access website about information. Accessed on 26 August 2016. Retrieved from http://www.collectionspace.org/about/

Does IMDB Provide an API? 2016. Discussion at stackoverflow.com about using IMDB's API. Accessed on 11 November 2016. Retrieved from http://stackoverflow.com/questions/1966503/does-imdb-provide-an-api

EN 15907 metadata representation. N.d. XML encoding example from metadata Standards for cinematographic works website. Accessed on 16 August 2016. Retrieved from http://filmstandards.org/demo/xml/001/vitelloni_04_v.html

EN 15907 preliminary version of an XML schema. N.d. Accessed on 16 August 2016. Retrieved from http://filmstandards.org/schemas/EN15907-d1/

The JSON Data Interchange Format. 2013. Document that describes ECMA-404 standard. ECMA International. Accessed on 3 December 2016. Retrieved from http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf

Lei, K., Ma, Y. & Tan, Z. 2014. Performance Comparison and Evaluation of Web Development Technologies in PHP, Python, and Node.js. From publication of IEEE 17th International Conference on Computational Science and Engineering (CSE), Chengdu, December 2014, 661 - 668. DOI: 10.1109/CSE.2014.142. Accessed on 30 November 2016. Retrieved from https://janet.finna.fi/, IEEE Xplore.

Munthe-Kaas, E. 2008. The Movie Database. Teaching materials. University of Oslo. Accessed on 26 October 2016. Retrieved from http://www.uio.no/studier/emner/matnat/ifi/INF3100/v08/undervisningsmateriale/Filmdatabase-ORM-UML_eng.pdf

Relational database. 2006. Article on SearchSQLServer's website that tells the definition of relational database. Accessed on 22 November 2016. Retrieved from http://searchsqlserver.techtarget.com/definition/relational-database

Rotten Tomatoes API - API Overview. 2006. Article on official Rotten Tomatoes' API documentation website. Accessed on 26 August 2016. Retrieved from http://developer.rottentomatoes.com/docs

Sem, J. 2013. When & Why to Use PHP Framework? Article written by professional web-developer. Accessed on 21 November 2016. Retrieved from http://www.templatemonster.com/blog/php-frameworks-when-and-why-to-use-them/

The Best PHP Framework for 2015. 2015. User survey on sitepoint.com. Accessed on 8 December 2016. Retrieved from https://www.sitepoint.com/best-php-framework-2015-sitepoint-survey-results/

What is an European Standard (EN)? N.d. Article on European Standardization Organization's website. Accessed on 22 November 2016. Retrieved from http://www.cencenelec.eu/standards/DefEN/Pages/default.aspx

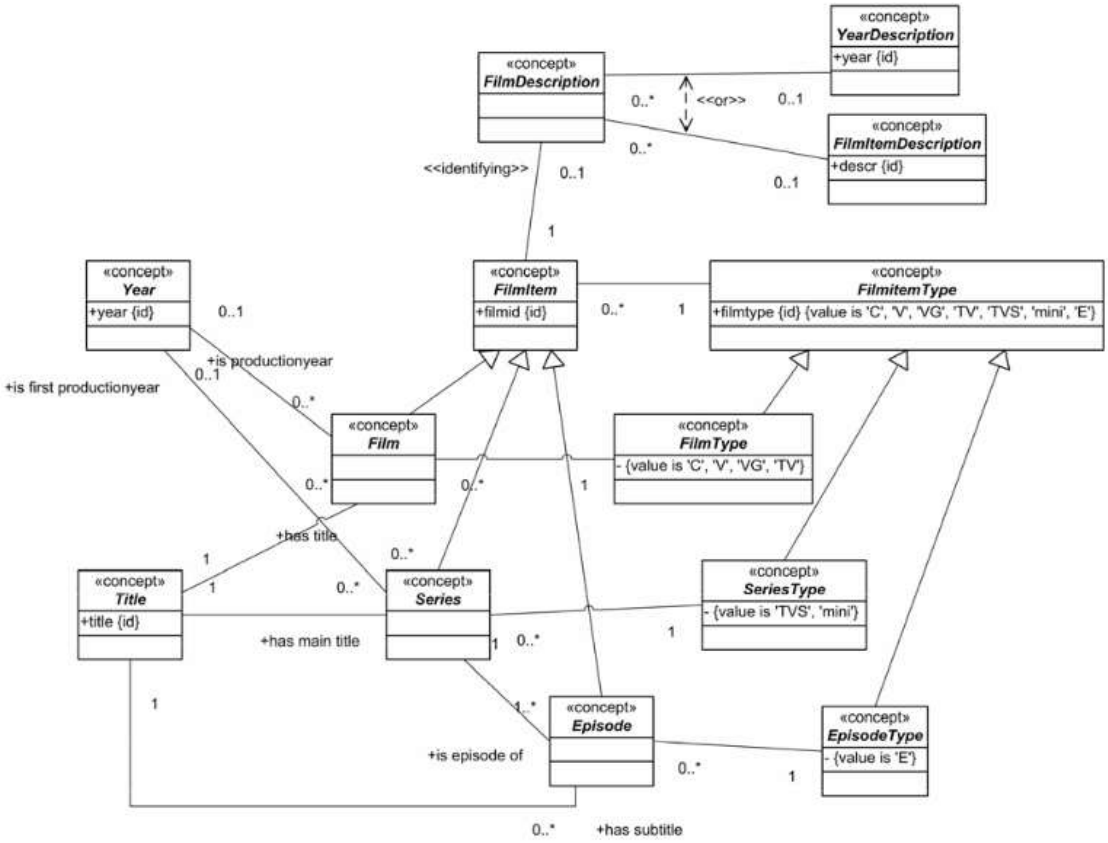
What is MySQL? N.d. Definition from MySQL's manual website. Accessed on 22 November 2016. Retrieved from https://dev.mysql.com/doc/refman/5.7/en/what-is-mysql.html

What is PHP? N.d. What is on PHP's manual. Accessed on 22 November 2016. Retrieved from http://php.net/manual/en/intro-whatis.php

Appendices

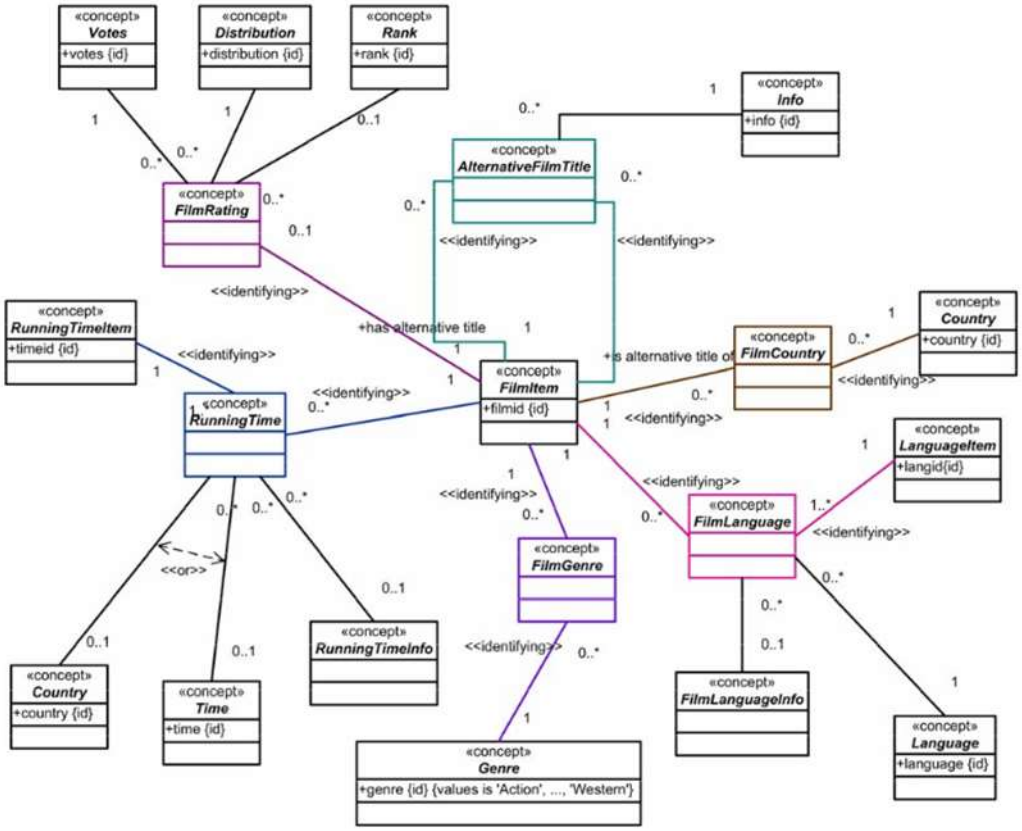
Appendix A

Intenet movie database ORM-UML diagram from 2007 (Munthe-Kaas, E. 2008).



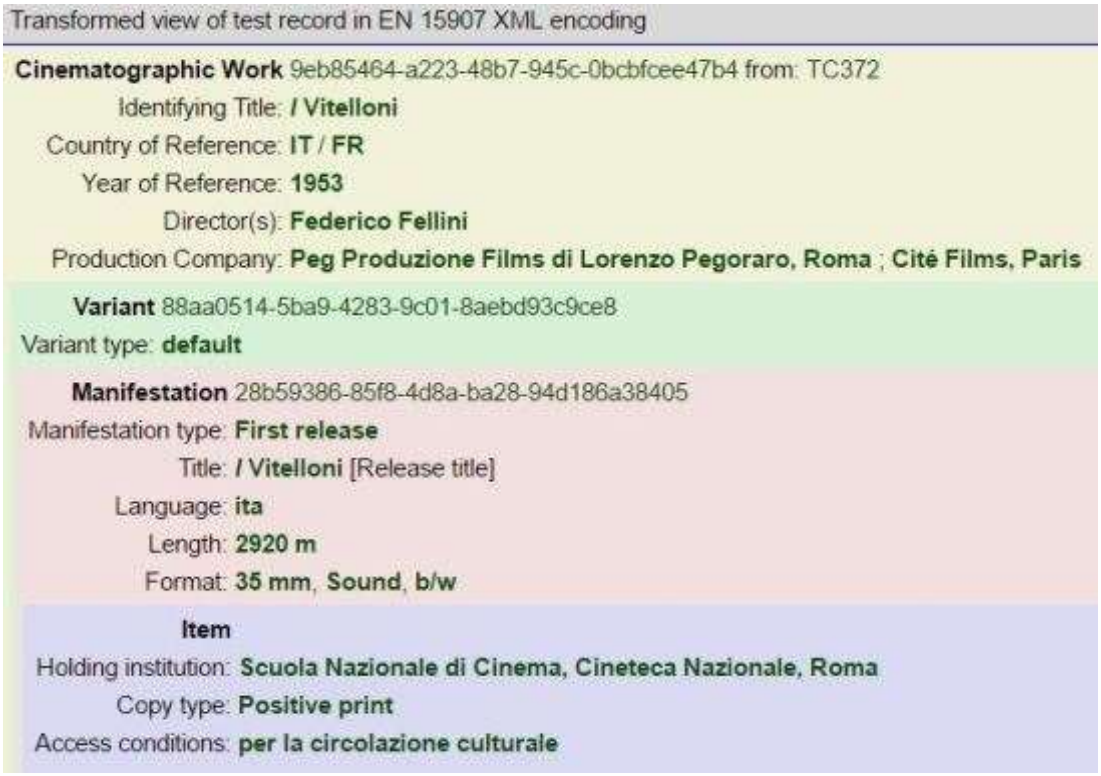
Appendix B

Intenet movie database, additional film information ORM-UML diagram from 2007 (Munthe-Kaas, E. 2008).



Appendix C

EN 15907 XML presenation (EN 15907 metadata representation)



Transformed view of test record in EN 15907 XML encoding

Cinematographic Work 9eb85464-a223-48b7-945c-0bcbfcee47b4 from: TC372
Identifying Title: **/ Vitelloni**
Country of Reference: **IT / FR**
Year of Reference: **1953**
Director(s): **Federico Fellini**
Production Company: **Peg Produzione Films di Lorenzo Pegoraro, Roma ; Cité Films, Paris**

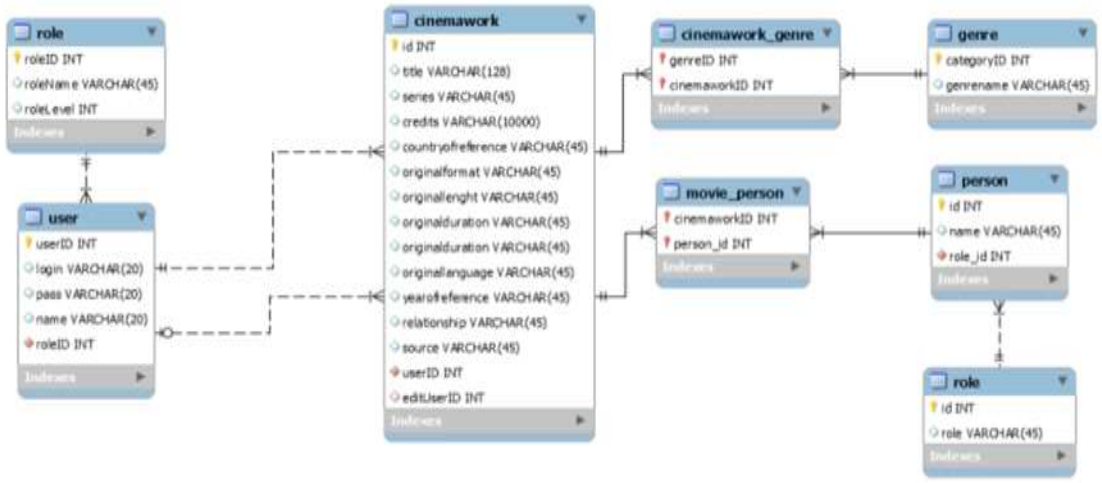
Variant 88aa0514-5ba9-4283-9c01-8aebd93c9ce8
Variant type: **default**

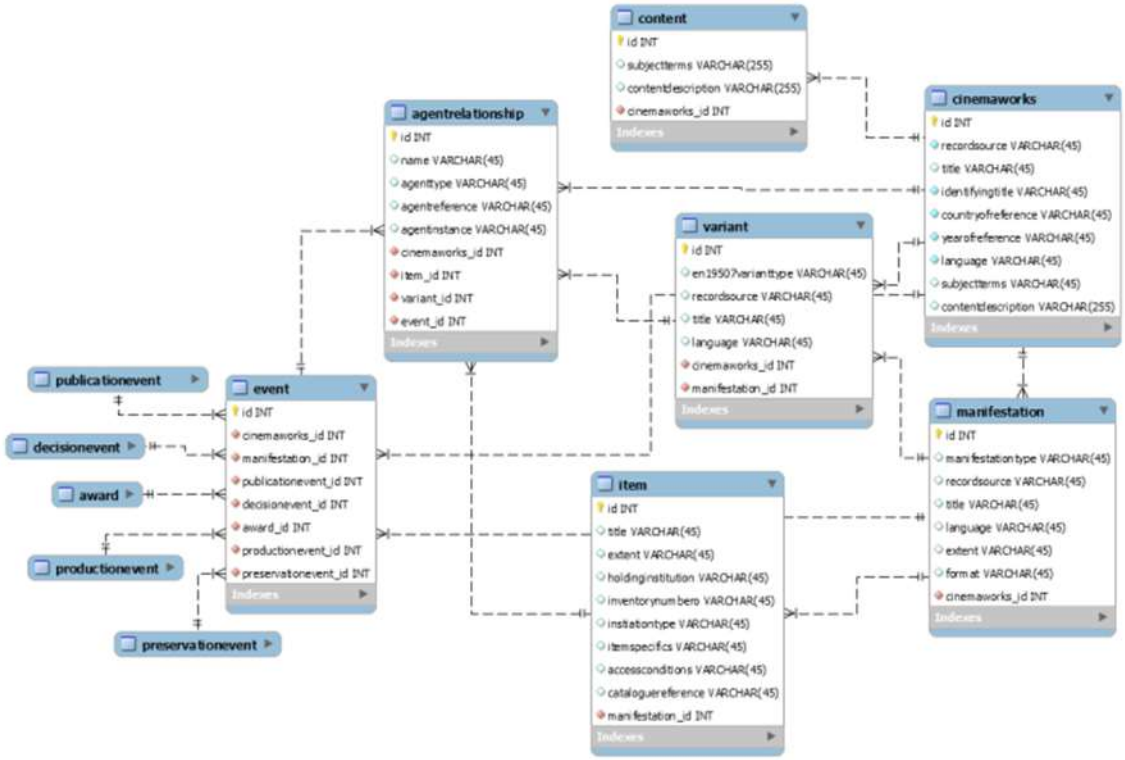
Manifestation 28b59386-85f8-4d8a-ba28-94d186a38405
Manifestation type: **First release**
Title: **/ Vitelloni** [Release title]
Language: **ita**
Length: **2920 m**
Format: **35 mm**, **Sound**, **b/w**

Item
Holding institution: **Scuola Nazionale di Cinema, Cineteca Nazionale, Roma**
Copy type: **Positive print**
Access conditions: **per la circolazione culturale**

Appendix D

EER model diagram based on EN 15744 data model – http://filmstandards.org/fsc/index.php/EN_15744



Appendix E

EER model diagram based on EN 15907 data model: http://filmstandards.org/fsc/index.php/EN_15907