Faculty of Computer Engineering, Cinema, and Mechatronics

MASTER'S DEGREE IN COMPUTER ENGINEERING

eeBook: Development of an Epub Reader in Rust with OCR System Integration

Academic Year 2021-2022

Developers: Claudio Di Maida
Giorgio Daniele Luppina

INDEX

INTRODUCTION

The following report discusses the process of developing an Epub Reader, highlighting key aspects of its development using the Rust programming language, which have led the software to provide a correct user experience. The various implementation choices that contributed to realizing an initial idea by Professor Alessandro Savino of the System Programming course, carried out by two university students, will be described and discussed.

The project development followed several main objectives, which will be analyzed in the subsequent chapters.

The first chapter will briefly introduce the application's GUI interface with a general description of the various tools present on the initial screen. The interface was developed with the support of the Rust framework, *Druid*[1].

The second chapter will focus on the process of opening a specific ".epub" book and the associated implementation choices. The entire operation process will be explained in detail.

The third chapter will address the "proofreading" tool, which allows real-time modification of the previously opened book by editing its HTML[2] source. The chapter will also discuss how the modified book is saved.

The fourth chapter will cover the additional features implemented to enhance the application's functional stability, providing the end user with mechanics that promote software usability.

---

[1] *Druid* is a data-driven declarative framework. The application model can be described in terms of data traits, creating a tree of widgets capable of displaying and modifying this data. https://docs.rs/druid/latest/druid/

[2] An ".epub" file has an internal structure divided into chapters written in HTML format.

The fifth chapter will discuss the integration of the OCR text recognition system, allowing conversion between paper and digital versions through relevant images. *Tesseract*[3] was chosen as the text recognition engine due to its open-source nature.

The *eeBook* software is free, executable on PCs running Windows® with Tesseract 4® installed, and can be downloaded by visiting the project's GitHub® page[4].

[3] For more information, visit: https://tesseract-ocr.github.io
[4] The GitHub® page link for *eeBook* is: https://github.com/TheFeNiXCode/eeBOOK.

1. USER INTERFACE

1.1. Project File Structure

For better organization of functionalities, the *eeBook* project consists of a *main* file and eight library files:

- chapter.rs
- command.rs
- constants.rs
- ebook.rs
- event.rs
- screen.rs
- search.rs
- toolbar.rs

1.2. The GUI

The *main* file generates a new window using the *WindowDesc::new(GUI)* function, initially set to full-screen. The GUI, created with the *Druid* library, starts with a main widget called with *Flex::column()*, to which two child objects are attached (Figure 1.1):

- The toolbar: contains a column and a row with seventeen child elements, such as buttons, sliders, and text bars.
- The screen: contains two columns, one for the cover and the other for the text.

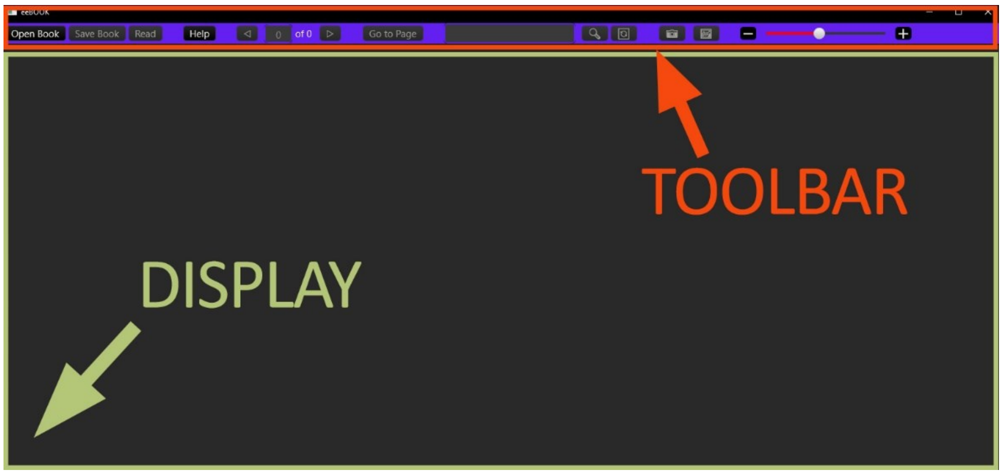


Figure 1.1. Graphic user interface of *eeBook*.

After generating all components, a *BookState* is instantiated, representing the main structure that holds information about the book to be opened. Finally, the application is launched.

As shown in Figure 1.1, the toolbar (defined in *toolbar.rs*) includes:

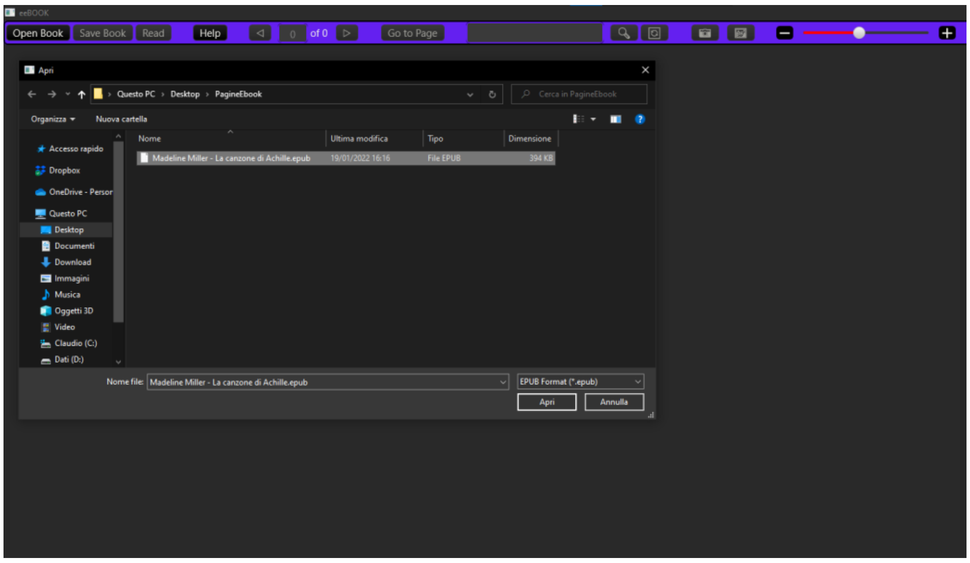
- ❖ *Open Book*: opens a new ".epub" book.
- ❖ *Save Button*: saves the book to a new destination.
- ❖ *Edit/Read*: switches between read mode and "proofreading" mode.
- ❖ *Help*: opens a screen explaining the features, goals, and authors of the application.
- ❖ ◁, ▷: buttons for navigating chapters or pages in *Edit mode* or *Read mode*, respectively.
- ❖ *Go to Page/Go to Chapter*: button to jump directly to a specific page or chapter specified in the respective *TextBox*.
- ❖ 🔍, ⟳: buttons for searching a word or phrase within the displayed page and refreshing the search results.
- ❖ *Slider per il font size*: using + and - buttons to enlarge or reduce text size.
- ❖ 📷, 🖼: buttons to switch between the paper and digital versions using OCR text recognition.

2. OPEN A BOOK

The first step, after addressing the interface, is to allow the user to open a selected book for reading. This section will describe the functionality of the *Open Book* button in the toolbar, its trigger mechanism, and the steps involved before displaying the book on the main screen.

2.1. *Open Book*

The *Open Book* button calls the only method implemented in *AppDelegate* (found in *command.rs*), which invokes the operating system to access the *File Explorer* (Figure 2.1), where the user selects the path of the ".epub" book.



Nome	Ultima modifica	Tipo	Dimensione
Madeline Miller - La canzone di Achille.epub	19/01/2022 16:16	File EPUB	394 KB

Nome file: Madeline Miller - La canzone di Achille.epub
EPUB Format (*.epub)

Figure 2.1. The *File Explorer* used to open a new book.

At this point, the resource is loaded via the static *loading()* method (in *ebook.rs*), which takes a *BookState* instance to populate.

Within *loading()*, the chapters that make up the book are created using the *Chapter* data structure. Once the chapters are reconstructed, the *HTML parsing* (from the *epub crate*[5]) and cover page parsing take place.

Then, the book is set up using the *setup()* function (in *ebook.rs*). This function generates a vector of strings representing the chapter's content, formatted as Title, Heading, or Paragraph. This design allows the *rendering* of the chapter to adapt to events like text search, page navigation, and text resizing, which alter some parameters of the *RichText*[6].

The implementation keeps both chapter scrolling (for easy chapter modification in *Edit mode*) and page scrolling (for convenient reading in *Read mode*).

Each digital page can hold a maximum of 10 items (Title, Headings, or Paragraphs). Creating a digital page involves concatenating the parts that make up a chapter. For example, Page 1 may consist of 1 Title, 1 Heading, and 8 Paragraphs; Page 2 may consist of 10 Paragraphs.

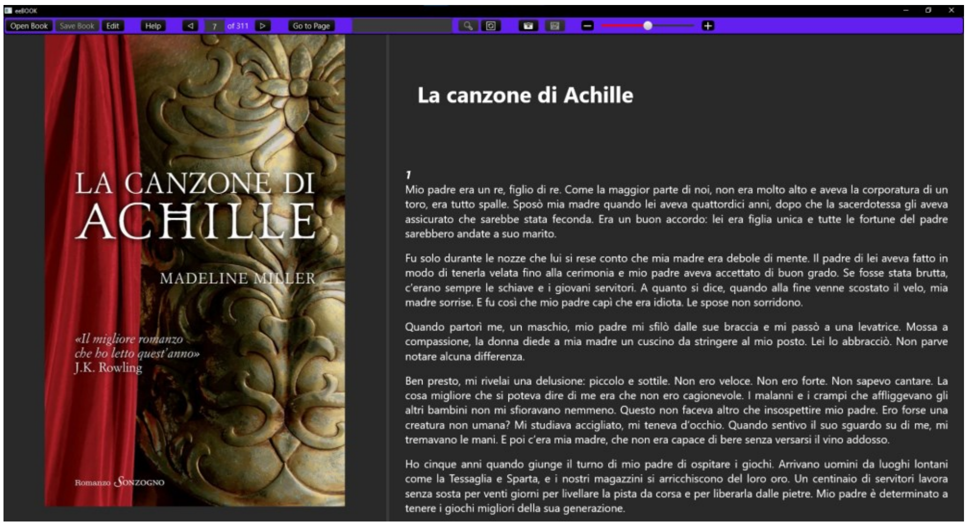
Two maps were created to number the pages (and chapters):

- ➢ The first map associates the chapter number (ID) with a vector of page numbers corresponding to that chapter.
- ➢ The second map associates the page number with the corresponding chapter.

In the *setup()* function, the HTML for *Edit mode* and the page for *Read mode* are initialized.

Finally, *ViewState::ReadMode* is set, which defines the application's states (*ReadMode*, *EditMode*, *HelpMode*, *Idle*), updating the display through *Druid*'s *ViewSwitcher*[7] abstraction.

[5] Web page of the *crate*: https://crates.io/crates/epub
[6] *Druid* data type for containing text in ".rtf" format.
[7] The *ViewSwitcher* is a widget defined in *Druid* that dynamically switches between two or more child widgets.



La canzone di Achille

1
Mio padre era un re, figlio di re. Come la maggior parte di noi, non era molto alto e aveva la corporatura di un toro, era tutto spalle. Sposò mia madre quando lei aveva quattordici anni, dopo che la sacerdotessa gli aveva assicurato che sarebbe stata feconda. Era un buon accordo: lei era figlia unica e tutte le fortune del padre sarebbero andate a suo marito.

Fu solo durante le nozze che lui si rese conto che mia madre era debole di mente. Il padre di lei aveva fatto in modo di tenerla velata fino alla cerimonia e mio padre aveva accettato di buon grado. Se fosse stata brutta, c'erano sempre le schiave e i giovani servitori. A quanto si dice, quando alla fine venne scostato il velo, mia madre sorrise. E fu così che mio padre capì che era idiota. Le spose non sorridono.

Quando partorì me, un maschio, mio padre mi sfilò dalle sue braccia e mi passò a una levatrice. Mossa a compassione, la donna diede a mia madre un cuscino da stringere al mio posto. Lei lo abbracciò. Non parve notare alcuna differenza.

Ben presto, mi rivelai una delusione: piccolo e sottile. Non ero veloce. Non ero forte. Non sapevo cantare. La cosa migliore che si poteva dire di me era che non ero cagionevole. I malanni e i crampi che affliggevano gli altri bambini non mi sfioravano nemmeno. Questo non faceva altro che insospettire mio padre. Ero forse una creatura non umana? Mi studiava accigliato, mi teneva d'occhio. Quando sentivo il suo sguardo su di me, mi tremavano le mani. E poi c'era mia madre, che non era capace di bere senza versarsi il vino addosso.

Ho cinque anni quando giunge il turno di mio padre di ospitare i giochi. Arrivano uomini da luoghi lontani come la Tessaglia e Sparta, e i nostri magazzini si arricchiscono del loro oro. Un centinaio di servitori lavora senza sosta per venti giorni per livellare la pista da corsa e per liberarla dalle pietre. Mio padre è determinato a tenere i giochi migliori della sua generazione.

Figure 2.2. Screen with a book open in *Read mode*.

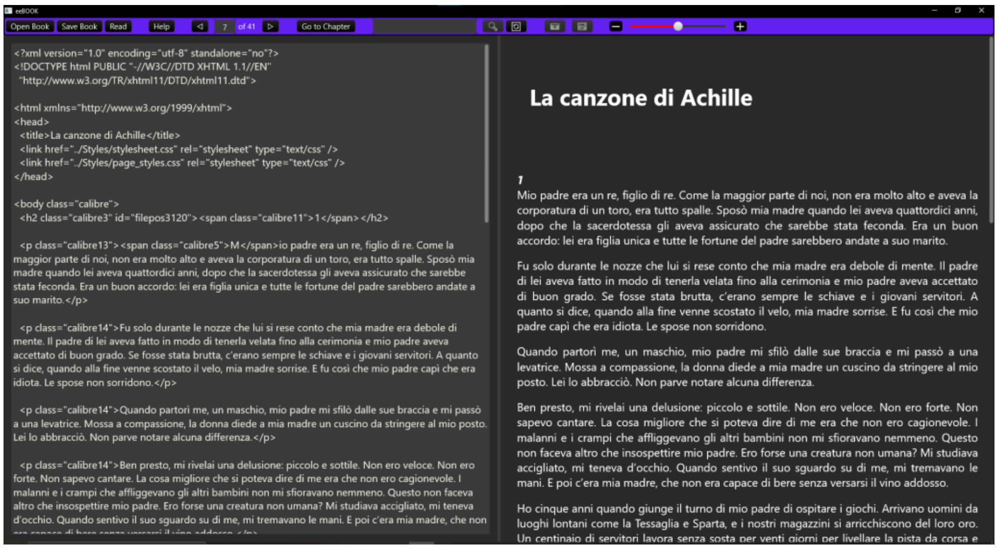
3. PROOFREADING TOOL

The implementation of the proofreading tool follows two main steps:

- ➢ Modifying the HTML file of the chapter being edited.
- ➢ Saving the modified book.

3.1. Modifying Text

The *Edit/Read* button in the toolbar switches between *Read* and *Edit* modes, triggering the *ViewSwitcher*.



```
<?xml version="1.0" encoding="utf-8" standalone="no"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
 "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">

<html xmlns="http://www.w3.org/1999/xhtml">
<head>
  <title>La canzone di Achille</title>
  <link href="../Styles/stylesheet.css" rel="stylesheet" type="text/css" />
  <link href="../Styles/page_styles.css" rel="stylesheet" type="text/css" />
</head>

<body class="calibre">
  <h2 class="calibre3" id="filepos3120"><span class="calibre11">1</span></h2>

  <p class="calibre13"><span class="calibre5">M</span>io padre era un re, figlio di re. Come la
maggior parte di noi, non era molto alto e aveva la corporatura di un toro, era tutto spalle. Sposò mia
madre quando lei aveva quattordici anni, dopo che la sacerdotessa gli aveva assicurato che sarebbe
stata feconda. Era un buon accordo: lei era figlia unica e tutte le fortune del padre sarebbero andate a
suo marito.</p>

  <p class="calibre14">Fu solo durante le nozze che lui si rese conto che mia madre era debole di
mente. Il padre di lei aveva fatto in modo di tenerla velata fino alla cerimonia e mio padre aveva
accettato di buon grado. Se fosse stata brutta, c'erano sempre le schiave e i giovani servitori. A quanto
si dice, quando alla fine venne scostato il velo, mia madre sorrise. E fu così che mio padre capì che era
idiota. Le spose non sorridono.</p>

  <p class="calibre14">Quando partorì me, un maschio, mio padre mi sfilò dalle sue braccia e mi passò
a una levatrice. Mossa a compassione, la donna diede a mia madre un cuscino da stringere al mio posto.
Lei lo abbracciò. Non parve notare alcuna differenza.</p>

  <p class="calibre14">Ben presto, mi rivelai una delusione: piccolo e sottile. Non ero veloce. Non ero
forte. Non sapevo cantare. La cosa migliore che si poteva dire di me era che non ero cagionevole. I
malanni e i crampi che affliggevano gli altri bambini non mi sfioravano nemmeno. Questo non faceva
altro che insospettire mio padre. Ero forse una creatura non umana? Mi studiava accigliato, mi teneva
d'occhio. Quando sentivo il suo sguardo su di me, mi tremavano le mani. E poi c'era mia madre, che non
era capace di bere senza versarsi il vino addosso.</p>
```

La canzone di Achille

1

Mio padre era un re, figlio di re. Come la maggior parte di noi, non era molto alto e aveva la corporatura di un toro, era tutto spalle. Sposò mia madre quando lei aveva quattordici anni, dopo che la sacerdotessa gli aveva assicurato che sarebbe stata feconda. Era un buon accordo: lei era figlia unica e tutte le fortune del padre sarebbero andate a suo marito.

Fu solo durante le nozze che lui si rese conto che mia madre era debole di mente. Il padre di lei aveva fatto in modo di tenerla velata fino alla cerimonia e mio padre aveva accettato di buon grado. Se fosse stata brutta, c'erano sempre le schiave e i giovani servitori. A quanto si dice, quando alla fine venne scostato il velo, mia madre sorrise. E fu così che mio padre capì che era idiota. Le spose non sorridono.

Quando partorì me, un maschio, mio padre mi sfilò dalle sue braccia e mi passò a una levatrice. Mossa a compassione, la donna diede a mia madre un cuscino da stringere al mio posto. Lei lo abbracciò. Non parve notare alcuna differenza.

Ben presto, mi rivelai una delusione: piccolo e sottile. Non ero veloce. Non ero forte. Non sapevo cantare. La cosa migliore che si poteva dire di me era che non ero cagionevole. I malanni e i crampi che affliggevano gli altri bambini non mi sfioravano nemmeno. Questo non faceva altro che insospettire mio padre. Ero forse una creatura non umana? Mi studiava accigliato, mi teneva d'occhio. Quando sentivo il suo sguardo su di me, mi tremavano le mani. E poi c'era mia madre, che non era capace di bere senza versarsi il vino addosso.

Ho cinque anni quando giunge il turno di mio padre di ospitare i giochi. Arrivano uomini da luoghi lontani come la Tessaglia e Sparta, e i nostri magazzini si arricchiscono del loro oro. Un centinaio di servitori lavora senza sosta per venti giorni per livellare la pista da corsa e

Figure 3.1. Screen with a book open in *Edit mode*.

As shown in Figure 3.1, the *ViewSwitcher* modifies the display so that the editable HTML text of the chapter appears on the left side of the screen, while the corresponding *RichText* appears on the right.

To see real-time text modifications, an event handler is added (in *screen.rs* for the *TextBox* widget containing the HTML) that returns a boolean depending on whether the event corresponds to a key press or release.

The *reactToHTMLModification()* function (in *event.rs*) acts as a wrapper for *updateHTML()*, which updates:

- The current HTML in the *BookState*'s HTML vector.
- The rendering of the text in ".rtf" format by invoking *update()*.

3.2. Saving the Book

Saving the book works similarly to opening it, as it calls the method in *AppDelegate* (found in *command.rs*), which invokes the operating system to access the *File Explorer* and select the path for saving the modified ".epub" file.

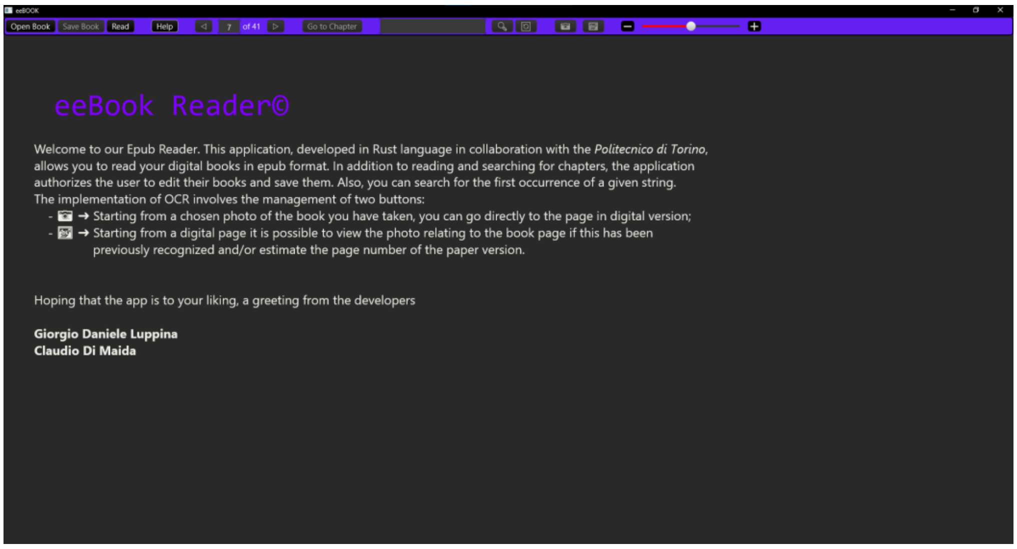
At this point, the file's extension is changed to ".zip", so the Windows operating system recognizes it as an extractable archive. The modified HTML files are then inserted into the archive, which is compressed and renamed with the original ".epub" extension.

4. ADDITIONAL FEATURES

This chapter describes all the additional features not covered in previous sections.

4.1. *Help* Screen

The Help button changes the *ViewSwitcher* value to *Help mode*. This displays a screen where the application's functionalities are described (Figure 4.1).



eeBook Reader©

Welcome to our Epub Reader. This application, developed in Rust language in collaboration with the *Politecnico di Torino*, allows you to read your digital books in epub format. In addition to reading and searching for chapters, the application authorizes the user to edit their books and save them. Also, you can search for the first occurrence of a given string.
The implementation of OCR involves the management of two buttons:

- → Starting from a chosen photo of the book you have taken, you can go directly to the page in digital version;
- → Starting from a digital page it is possible to view the photo relating to the book page if this has been previously recognized and/or estimate the page number of the paper version.

Hoping that the app is to your liking, a greeting from the developers

Giorgio Daniele Luppina
Claudio Di Maida

Figure 4.1. *Help Mode* Screen.

Using *Druid*'s *Lens* trait, access is gained to a field in *BookState*, where the *RichText* describing the application's functions is initialized.

4.2. Page Scrolling

The [◁] and [▷] buttons scroll through chapters in *Edit mode* or pages in *Read mode*. Both invoke the *scroll()* function (in *ebook.rs*), which, depending on the direction and display mode, saves the modified HTML and updates the view with the next (or previous) page by calling *update()*. This function:

- In *Read mode*: retrieves the page to display from *BookState*, formats it, and retrieves the corresponding chapter using the *pageBelongs* map (which returns the chapter index). The rendering is pre-calculated, and the current chapter's HTML is retrieved.
- In *Edit mode*: iterates through chapters instead of pages, retrieving from *BookState* the plaintext of the chapter, rendering it, and loading the first page of the chapter.

Finally, the *update()* method updates a "*cursor*" variable, containing information on the current page (or chapter), which is displayed in a dedicated *TextBox* (Figure 4.2).

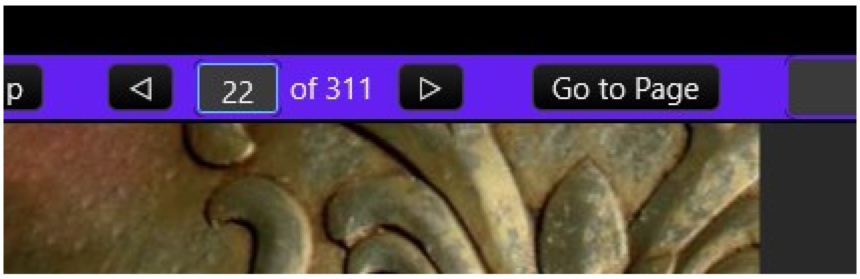


Figure 4.2. Detail of the *TextBox* displaying the current page number.

4.3. *Go to Page/Chapter*

The user can specify the page (in *Read mode*) or chapter (in *Edit mode*) by entering a value in the *TextBox* (Figure 4.2) and pressing the corresponding *Go to Page/Chapter* button.

This button acts as a wrapper for the static *jump()* method (in *ebook.rs*). Once it verifies that the input is valid, it calls *update()* on the page or chapter.

4.4. Text Search

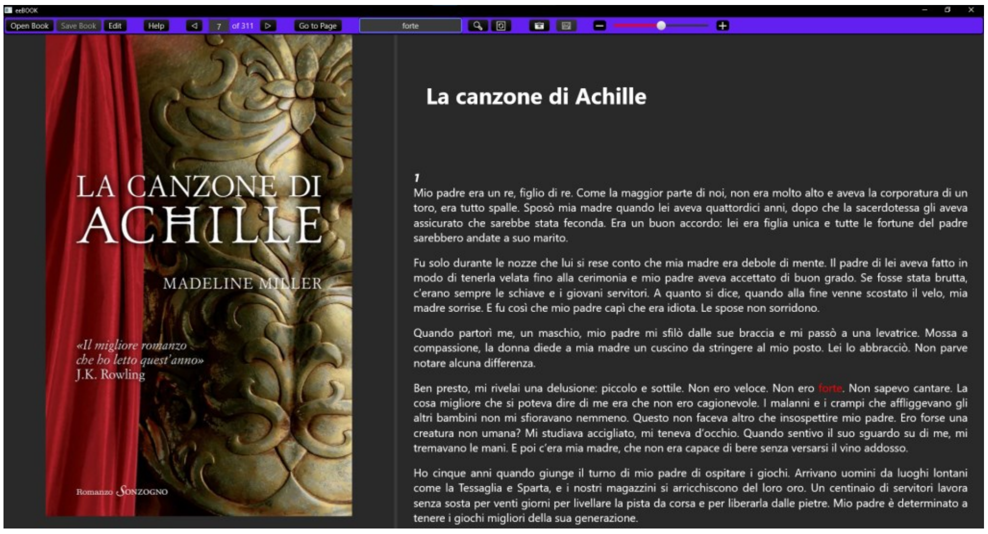
The text search feature consists of a search bar and two buttons:

- *SearchButton* [🔍]

- *refreshSearcButton* [🔄]

The *searchBar* functions as a front-end for regular expressions[8] (*Regex*). The *SearchButton* first determines whether the search is in Edit or Read mode and performs a search in the current chapter or page.

The *looksForMatchesInPage()* function (in *search.rs*) retrieves the string vector from *BookState* (10 strings per page, as described in Section 1.2) for the current page and applies the regular expression to each string.



La canzone di Achille

1
Mio padre era un re, figlio di re. Come la maggior parte di noi, non era molto alto e aveva la corporatura di un toro, era tutto spalle. Sposò mia madre quando lei aveva quattordici anni, dopo che la sacerdotessa gli aveva assicurato che sarebbe stata feconda. Era un buon accordo: lei era figlia unica e tutte le fortune del padre sarebbero andate a suo marito.

Fu solo durante le nozze che lui si rese conto che mia madre era debole di mente. Il padre di lei aveva fatto in modo di tenerla velata fino alla cerimonia e mio padre aveva accettato di buon grado. Se fosse stata brutta, c'erano sempre le schiave e i giovani servitori. A quanto si dice, quando alla fine venne scostato il velo, mia madre sorrise. E fu così che mio padre capì che era idiota. Le spose non sorridono.

Quando partorì me, un maschio, mio padre mi sfilò dalle sue braccia e mi passò a una levatrice. Mossa a compassione, la donna diede a mia madre un cuscino da stringere al mio posto. Lei lo abbracciò. Non parve notare alcuna differenza.

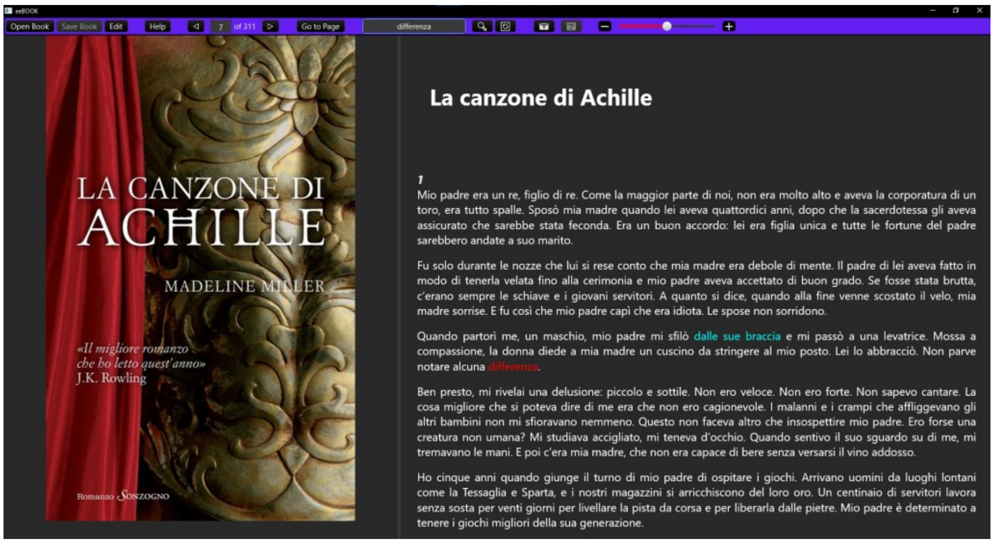
Ben presto, mi rivelai una delusione: piccolo e sottile. Non ero veloce. Non ero forte. Non sapevo cantare. La cosa migliore che si poteva dire di me era che non ero cagionevole. I malanni e i crampi che affliggevano gli altri bambini non mi sfioravano nemmeno. Questo non faceva altro che insospettire mio padre. Ero forse una creatura non umana? Mi studiava accigliato, mi teneva d'occhio. Quando sentivo il suo sguardo su di me, mi tremavano le mani. E poi c'era mia madre, che non era capace di bere senza versarsi il vino addosso.

Ho cinque anni quando giunge il turno di mio padre di ospitare i giochi. Arrivano uomini da luoghi lontani come la Tessaglia e Sparta, e i nostri magazzini si arricchiscono del loro oro. Un centinaio di servitori lavora senza sosta per venti giorni per livellare la pista da corsa e per liberarla dalle pietre. Mio padre è determinato a tenere i giochi migliori della sua generazione.

Figure 4.3. Search results for a word.

Each string may have multiple matches; thus, all match intervals are recorded using the *captures_iter()* function. These intervals are used to regenerate the ".rtf" page with highlighted matches via the *reactToSearch()* function, which assigns a different color to each matching portion of text (Figure 4.3).

[8] A regular expression is a sequence of symbols that identifies a set of strings.



La canzone di Achille

1
Mio padre era un re, figlio di re. Come la maggior parte di noi, non era molto alto e aveva la corporatura di un toro, era tutto spalle. Sposò mia madre quando lei aveva quattordici anni, dopo che la sacerdotessa gli aveva assicurato che sarebbe stata feconda. Era un buon accordo: lei era figlia unica e tutte le fortune del padre sarebbero andate a suo marito.

Fu solo durante le nozze che lui si rese conto che mia madre era debole di mente. Il padre di lei aveva fatto in modo di tenerla velata fino alla cerimonia e mio padre aveva accettato di buon grado. Se fosse stata brutta, c'erano sempre le schiave e i giovani servitori. A quanto si dice, quando alla fine venne scostato il velo, mia madre sorrise. E fu così che mio padre capì che era idiota. Le spose non sorridono.

Quando partorì me, un maschio, mio padre mi sfilò dalle sue braccia e mi passò a una levatrice. Mossa a compassione, la donna diede a mia madre un cuscino da stringere al mio posto. Lei lo abbracciò. Non parve notare alcuna differenza.

Ben presto, mi rivelai una delusione: piccolo e sottile. Non ero veloce. Non ero forte. Non sapevo cantare. La cosa migliore che si poteva dire di me era che non ero cagionevole. I malanni e i crampi che affliggevano gli altri bambini non mi sfioravano nemmeno. Questo non faceva altro che insospettire mio padre. Ero forse una creatura non umana? Mi studiava accigliato, mi teneva d'occhio. Quando sentivo il suo sguardo su di me, mi tremavano le mani. E poi c'era mia madre, che non era capace di bere senza versarsi il vino addosso.

Ho cinque anni quando giunge il turno di mio padre di ospitare i giochi. Arrivano uomini da luoghi lontani come la Tessaglia e Sparta, e i nostri magazzini si arricchiscono del loro oro. Un centinaio di servitori lavora senza sosta per venti giorni per livellare la pista da corsa e per liberarla dalle pietre. Mio padre è determinato a tenere i giochi migliori della sua generazione.

Figure 4.4. Result of a subsequent search, retaining previous highlights.

Additionally, the feature retains the previous search results, which remain highlighted in subsequent searches (Figure 4.4).

The *refreshSearchButton()* is a wrapper for the *update()* function, which re-renders the page without the highlighted text.

4.5. Font Size

The font size adjustment feature consists of a slider and two buttons [+] and [-]. Both the buttons and the slider perform the following steps:

1. Take the current font size as input.
2. Modify the value by a predetermined amount.

Once this is done, the display widget is updated via the controller in the *ZoomEvent* if the font size has changed.

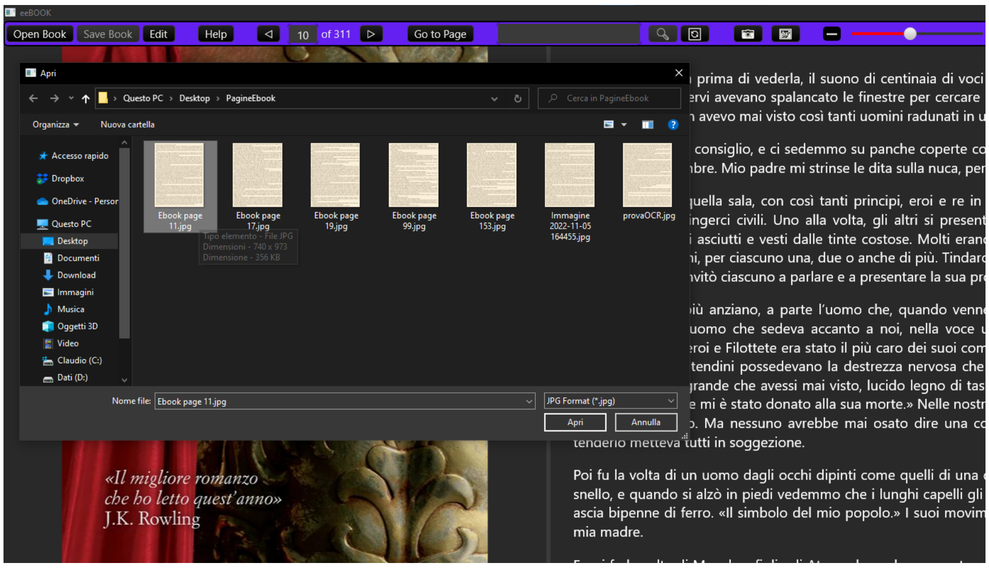
5. OCR RECOGNITION

In this final chapter, the implementations related to text recognition will be discussed. The goal is to integrate the camera system with OCR to link the physical version of the book with the digital one in the following ways:

- *OCR Direct*: in the digital version, jump to the point where you left off in the physical version by uploading a photo of a page from the book.
- *OCR Reverse*: from one or more recognized physical pages, indicate which page of the digital version corresponds to the physical one.

5.1. *OCR Direct*

Initially, the procedure for opening a photo (or image) follows the same flow as opening the book. The *AppDelegate* is used, but in this case, images are opened (Figure 5.1).



prima di vederla, il suono di centinaia di voci
ervi avevano spalancato le finestre per cercare
n avevo mai visto così tanti uomini radunati in u

consiglio, e ci sedemmo su panche coperte co
nbre. Mio padre mi strinse le dita sulla nuca, per

quella sala, con così tanti principi, eroi e re in
ingerci civili. Uno alla volta, gli altri si present
i asciutti e vesti dalle tinte costose. Molti eran
ni, per ciascuno una, due o anche di più. Tindar
nvitò ciascuno a parlare e a presentare la sua pr

più anziano, a parte l'uomo che, quando venn
uomo che sedeva accanto a noi, nella voce u
eroi e Filottete era stato il più caro dei suoi com
tendini possedevano la destrezza nervosa che
grande che avessi mai visto, lucido legno di tas
e mi è stato donato alla sua morte.» Nelle nostr
o. Ma nessuno avrebbe mai osato dire una co
tenderlo metteva tutti in soggezione.

Poi fu la volta di un uomo dagli occhi dipinti come quelli di una
snello, e quando si alzò in piedi vedemmo che i lunghi capelli gli
ascia bipenne di ferro. «Il simbolo del mio popolo.» I suoi movim
mia madre.

Figure 5.1. Opening a photo to search in the digital book.

The image opening is followed by the *Command* that saves the image path in a variable and executes the *loading()* function (in *ebook.rs*), which acts as a wrapper for the OCR text recognition function.

The OCR search begins by specifying the language to use for recognition, which is retrieved from the information contained within the ".epub" file. Next, an OCR process is executed via the command line that saves the result of the text analysis in a temporary file. This file is then read and saved in a string *tessOutput*, which is subsequently deleted.

The *tessOutput* string is passed to the *directOCR()* function (found in *search.rs*), which retrieves the vector containing the text of all the pages and initializes a vector of threads. A common tuple *bestHit* is used, protected by a *mutex*.

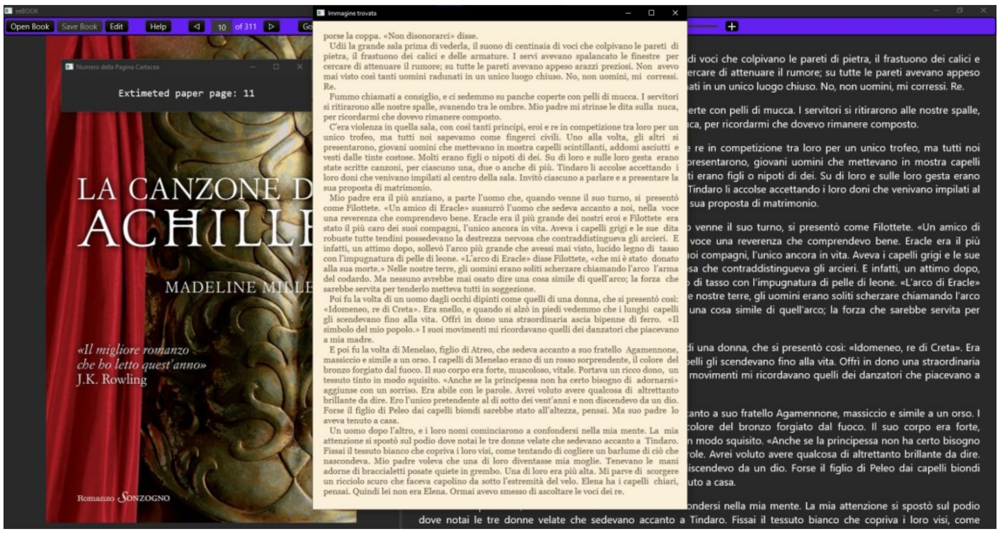
The process works as follows: each thread processes a specific chapter of the book (a subset of pages) and checks for the presence or absence of the strings acquired from the image. For each page, the number of matches is counted, and the value of *bestHit* is modified only if it is greater than the current value. The tuple *bestHit* contains the number of hits and the page on which they were found.

The threads are awaited by the main thread, and the algorithm identifies the page with the most matches and jumps to it using the *jump()* method, which simply wraps the *update()* function.

5.2. *OCR Reverse*

The various paths of previously acquired images form a small database for this part.

To estimate the page number of the physical version, all the images in the database are first opened, and the text is extracted using the *reverseOCR()* function. Each thread works on an image, extracting the text and comparing it with the current digital page. If the digital text finds a complete match with one of the images in the database, this image is considered the physical equivalent, using the same principle as *bestHit*, and its path is the output of the function. This image is then opened and displayed in a separate *Druid* window from the main one (Figure 5.2).



Extimeted paper page: 11

porse la coppa. «Non disonorarci» disse.
Udii la grande sala prima di vederla, il suono di centinaia di voci che colpivano le pareti di pietra, il frastuono dei calici e delle armature. I servi avevano spalancato le finestre per cercare di attenuare il rumore; su tutte le pareti avevano appeso arazzi preziosi. Non avevo mai visto così tanti uomini radunati in un unico luogo chiuso. No, non uomini, mi corressi. Re.
Fummo chiamati a consiglio, e ci sedemmo su panche coperte con pelli di mucca. I servitori si ritirarono alle nostre spalle, svanendo tra le ombre. Mio padre mi strinse le dita sulla nuca, per ricordarmi che dovevo rimanere composto.
C'era violenza in quella sala, con così tanti principi, eroi e re in competizione tra loro per un unico trofeo, ma tutti noi sapevamo come fingerci civili. Uno alla volta, gli altri si presentarono, giovani uomini che mettevano in mostra capelli scintillanti, addomi asciutti e vesti dalle tinte costose. Molti erano figli o nipoti di dei. Su di loro e sulle loro gesta erano state scritte canzoni, per ciascuno una, due o anche di più. Tindaro li accolse accettando i loro doni che venivano impilati al centro della sala. Invitò ciascuno a parlare e a presentare la sua proposta di matrimonio.
Mio padre era il più anziano, a parte l'uomo che, quando venne il suo turno, si presentò come Filottete. «Un amico di Eracle» sussurrò l'uomo che sedeva accanto a noi, nella voce una reverenza che comprendevo bene. Eracle era il più grande dei nostri eroi e Filottete era stato il più caro dei suoi compagni, l'unico ancora in vita. Aveva i capelli grigi e le sue dita robuste tutte tendini possedevano la destrezza nervosa che contraddistingueva gli arcieri. E infatti, un attimo dopo, sollevò l'arco più grande che avessi mai visto, lucido legno di tasso con l'impugnatura di pelle di leone. «L'arco di Eracle» disse Filottete, «che mi è stato donato alla sua morte.» Nelle nostre terre, gli uomini erano soliti scherzare chiamando l'arco l'arma del codardo. Ma nessuno avrebbe mai osato dire una cosa simile di quell'arco; la forza che sarebbe servita per tenderlo metteva tutti in soggezione.
Poi fu la volta di un uomo dagli occhi dipinti come quelli di una donna, che si presentò così: «Idomeneo, re di Creta». Era snello, e quando si alzò in piedi vedemmo che i lunghi capelli gli scendevano fino alla vita. Offrì in dono una straordinaria ascia bipenne di ferro. «Il simbolo del mio popolo.» I suoi movimenti mi ricordavano quelli dei danzatori che piacevano a mia madre.
E poi fu la volta di Menelao, figlio di Atreo, che sedeva accanto a suo fratello Agamennone, massiccio e simile a un orso. I capelli di Menelao erano di un rosso sorprendente, il colore del bronzo forgiato dal fuoco. Il suo corpo era forte, muscoloso, vitale. Portava un ricco dono, un tessuto tinto in modo squisito. «Anche se la principessa non ha certo bisogno di adornarsi» aggiunse con un sorriso. Era abile con le parole. Avrei voluto avere qualcosa di altrettanto brillante da dire. Ero l'unico pretendente al di sotto dei vent'anni e non discendevo da un dio. Forse il figlio di Peleo dai capelli biondi sarebbe stato all'altezza, pensai. Ma suo padre lo aveva tenuto a casa.
Un uomo dopo l'altro, e i loro nomi cominciarono a confondersi nella mia mente. La mia attenzione si spostò sul podio dove notai le tre donne velate che sedevano accanto a Tindaro. Fissai il tessuto bianco che copriva i loro visi, come tentando di cogliere un barlume di ciò che nascondeva. Mio padre voleva che una di loro diventasse mia moglie. Tenevano le mani adorne di braccialetti posate quiete in grembo. Una di loro era più alta. Mi parve di scorgere un ricciolo scuro che faceva capolino da sotto l'estremità del velo. Elena ha i capelli chiari, pensai. Quindi lei non era Elena. Ormai avevo smesso di ascoltare le voci dei re.

Figure 5.2. Result of the physical page estimation operation.

In any case, even if the corresponding physical image is not found in the database, a *Druid* window is still shown with an estimate of the actual page number of the book, according to an algorithm summarized in two steps:

1. Every photograph loaded into *eeBook* updates, in *directOCR()*, the *paperWordsEvaluation* variable contained in the *BookState*. It is initialized with the word count of the first image and modified with an arithmetic average of the word counts of the subsequent images.
2. Each time the *reverseOCR()* function is triggered, all the digital pages prior to the page for which the physical page number is to be estimated are considered, and a total word count is made for each chapter. The estimated number of pages for the chapter is derived by dividing this count by *paperWordsEvaluation*. If there is a remainder in the division, one additional page is estimated for the chapter. This process is repeated for all chapters until the target page is reached, providing an estimate for the page number, which is then set in the *digitalPageEvaluation* variable of the *BookState* and displayed on screen in a dedicated window.

INDEX OF FIGURES

BIBLIOGRAPHY

AA.VV., eBook source code available on the project's GitHub page: https://github.com/TheFeNiXCode/eeBOOK

AA.VV., *Crate "epub" latest documentation* available on the website: https://crates.io/crates/epub

AA.VV., *Druid latest documentation* available on the website: https://docs.rs/druid/latest/druid/

AA.VV., *Tesseract OCR documentation* available on the website: https://tesseract-ocr.github.io

S. Klabnik, C. Nichols*, The Rust Programming Language* available on the website: https://doc.rust-lang.org/book/