POLITECNICO DI MILANO
Master of Science in Computer Engineering
Dipartimento di Elettronica e Informazione


# MULTILINGUAL EBOOK READER

Relatore: Prof. Licia Sbattella
Correlatore: Ing. Roberto Tedesco

Thesis by:
Guvener Gokce, 754837

Academic Year 2011-2012

*To my lovely grandmother...*

# Abstract

Nowadays, the world is becoming increasingly multilingual that the ability to speak a second language has never been as vital as it has been today. People can use many different ways to develop their proficiency in foreign languages, and reading is undoubtedly one of the best ways to improve their skills.

Reading a book written in original language is not easy task, however, and often leads to the reader giving-up the activity. From this point of view, reading multilingual books has many advantages for foreign-language learners; for example, they can read in the original language and use the translated text as a resource to verify their understanding.

Although multilingual books provide text in two languages, switching between them and finding the translations is not easy. Readers have to move their eyes back and forth from the original language text to the translation, and this will eventually cause breaks in reading activity. Moreover the reader, having the translated text too close to her/his visual field, tends to read it too often. Using two books, the one in original language and the other containing the translation, solves the latter problem, but accentuates the former.

As a solution, ebooks can be used to handle this valuable multilingual content representation efficiently without the need of moving reader's eyes.

In this thesis, a multilingual ebook reader application for iOS iPad is presented. The application exploits ebooks natural "dynamicity" to encounter difficulties in understanding the foreign language content by uniting the original text with the literary translation. Furthermore, it makes use of gestural interfaces in order to avoid distractions while presenting original-language content and provides engagement in reading.

# Sommario

Nella società di oggi, la conoscenza di una seconda lingua sta diventando sempre più importante. Esistono diversi approcci che possono essere utilizzati per sviluppare la propria conoscenza di una seconda lingua, e la lettura è senza dubbio uno dei metodi migliori.

Leggere un libro in lingua originale non è però semplice, e spesso lo studente si scoraggia e abbandona l'attività. Da questo punto di vista, i libri con testo originale a fronte offrono molti vantaggi: lo studente può leggere il testo nella lingua originale, e utilizzare la versione tradotta come una risorsa per verificare la comprensione del testo.

In un libro di questo tipo, però, il passaggio da una lingua all'altra, per trovare la traduzione corrispondente ad un certa espressione, non è facile: i lettori devono muovere gli occhi avanti e indietro dalla lingua originale alla traduzione, e ricercare il passo a cui si è interessati. Questo causa continue interruzioni nell'attività di lettura. Inoltre, avere costantemente il testo tradotto vicino al proprio campo visivo espone il lettore alla tentazione di guardarlo troppo spesso. Utilizzare due libri, il primo in lingua originale e il secondo contenente la traduzione, risolve quest'ultimo problema, ma acuisce il primo.

Come soluzione, gli ebook possono essere usati per gestire dinamicamente il contenuto multilingue, mantenendo costantemente la concentrazione sul testo.

In questa tesi, si presenta un'applicazione per la lettura di ebook multilingue, sviluppata per iOS su iPad, che sfrutta la naturale "dinamicità" degli ebook per unire il contenuto in lingua originale con la versione tradotta. L'applicazione sfrutta l'interfaccia gestuale per ridurre le fonti di distrazione e aumentare l'efficacia dell'attività di lettura.

# Acknowledgements

I would like to thank my project supervisor, Prof. Licia Sbattella for giving me the opportunity to work on this thesis project and my project advisor, Ing. Roberto Tedesco for his continuous guidance, support, and motivation throughout the project.

# Contents

**Abstract** **I**

**Acknowledgements** **V**

**1 Introduction** **3**
1.1 Motivation . . . . . 4
1.2 Purpose of Thesis . . . . . 6
1.3 Thesis Organization . . . . . 7

**2 Background Research** **9**
2.1 Multilingualism . . . . . 9
2.2 The Advantages of Multilingualism . . . . . 9
2.3 Methods of Learning a Second Language . . . . . 10
2.4 Multilingual Books . . . . . 11
2.5 Ebooks . . . . . 13
2.6 Multilingual Electronic Publications . . . . . 14

**3 Technologies** **17**
3.1 Choosing Technologies . . . . . 17
3.2 EPUB . . . . . 18
3.2.1 EPUB Structure . . . . . 19
3.2.2 Creating an EPUB file . . . . . 24
3.2.3 Advantages of the EPUB . . . . . 26
3.2.4 EPUB Reader . . . . . 27
3.3 Javascript . . . . . 27
3.4 CSS . . . . . 28
3.5 XML . . . . . 29
3.6 JSON . . . . . 30
3.7 Objective-C . . . . . 31

**4 Design of Multilingual Content Representation on EPUB 33**
4.1 EPUB Content Design . . . . . 33
4.2 Multilingual Content Representation . . . . . 34
4.3 Content Interaction . . . . . 35
4.4 Content Structure . . . . . 35
4.5 Multilingual Ebook Construction . . . . . 37
4.5.1 Sample Book and Translation Source . . . . . 37
4.5.2 Multilingual Content Organization . . . . . 37
4.5.3 Multilingual EPUB File Creation . . . . . 39

**5 User Interface and Gestures 41**
5.1 Gestural User Interfaces . . . . . 41
5.2 Home Screen . . . . . 45
5.3 Inline Translation . . . . . 46
5.4 Paragraph Translation . . . . . 47
5.5 Footnote Translation . . . . . 48
5.6 Font Size and Brightness Controls . . . . . 49
5.7 Importing New Ebook . . . . . 50

**6 System Architecture and Software Design 51**
6.1 Architecture Overview . . . . . 51
6.2 Presentation Layer . . . . . 52
6.3 Application Layer . . . . . 53
6.4 Data Access Layer . . . . . 53
6.5 Class Structures and Relationships . . . . . 54
6.6 Software Design . . . . . 55
6.6.1 Communication Bridge . . . . . 55
6.6.2 Script Initialization . . . . . 57
6.6.3 Script Data Flow . . . . . 58
6.6.4 Event Priorities . . . . . 60
6.7 System Requirements . . . . . 60

**7 Conclusion and Future Work 61**
7.1 Conclusion . . . . . 61
7.2 Future Work . . . . . 62

**References 65**

**A Documentation 69**

# List of Figures


2.1 Traditional bilingual text in English and Spanish . . . . . 12
2.2 Flip over multilingual book in English and French . . . . . 13
2.3 Electronic multilingual magazine in English and Turkish . . . 14

3.1 EPUB folder structure . . . . . 21
3.2 EPUB OPF file . . . . . 23
3.3 EPUB container file . . . . . 24
3.4 Adobe Digital Editions export view . . . . . 25
3.5 Adobe Digital Editions EPUB preview . . . . . 26

4.1 Multilingual content representation model . . . . . 34
4.2 Multilingual content structure . . . . . 36
4.3 Multilingual XHTML content sample . . . . . 38

5.1 BiText reader mockup for inline translation state . . . . . 42
5.2 BiText reader mockup for paragraph translation state . . . . 43
5.3 BiText reader gesture functionality flow . . . . . 44
5.4 BiText reader ebook listing . . . . . 45
5.5 BiText reader inline translation dialog . . . . . 46
5.6 BiText reader paragraph translation dialog, showing the "continuation" three-dot icon . . . . . 47
5.7 BiText reader EPUB auther footnote dialog . . . . . 48
5.8 BiText reader controls for adjusting fontsize and brightness . 49
5.9 BiText reader import ebook view . . . . . 50

6.1 Bitext reader architecture diagram . . . . . 52
6.2 BiText class diagram . . . . . 54
6.3 BiText class relationships . . . . . 55
6.4 BiText Reader - BiText Core communication bridge . . . . . 56
6.5 Multilingual EPUB content identification flow . . . . . 57
6.6 BiText core function for retrieving sentence data in JSON . . 58
6.7 BiText core function for event listener initialization. . . . . 58

6.8 BiText reader method for catching Javascript functions . . . 59

A.1 Bitext reader documentation . . . . . 69

# Chapter 1

# Introduction

*"One language sets you in a corridor for life. Two languages open every door along the way."*

Frank Smith, Psycholinguist

Since the first signs of modern human behavior, scientists began to search for ways to break the language barriers that prevented them from gaining access to the knowledge that other civilizations possessed. Their curiosity for the information that was being communicated in foreign languages constituted a "survival issue" in some cases and an "ethical issue" in others [10]. In today's world, language remains the main instrument of communication that is used to convey messages within and between different societies and cultures.

Along with a fascination with language, human desire to create lasting records of events and knowledge spawned the invention of writing and, consequently, books. History books note that during the time of Constantine the Great, there were 28 public libraries in Rome, and many smaller libraries in other cities [26]. Furthermore, the invention of printing techniques in the fifteenth century developed the field of translation and helped to spread written communications throughout different civilizations.

Today, information technology has changed the way that data is stored, disseminated and presented and the digital revolution has had a major impact on the world of publishing. Paperless or electronic publishing is increasingly gaining in popularity. Within this changing scenario, the standard and characteristics of books are also reshaping. Electronic books are more widely used and electronic publication (EPUB) is becoming the new standard of digital publishing. With the help of EPUB, ebooks can be created in different languages and readers can download and read them in special readers

that are able to access and display ebooks that are displayed different languages. As such, ebooks are beginning to offer quick and easy access to information that is stored in a wide range of different languages that can meet the needs of those who have an unstoppable thirst for knowledge.

## 1.1 Motivation

Mobile devices have evolved significantly over the last decade. They keep users connected to information on a variety of levels, for business, personal and educational purposes. In addition, the creation of software development kits (SDK), which give developers the opportunity to build creative mobile applications, has entailed that the multi-functional abilities of mobile devices has risen to a whole new level. These capabilities were not missed by Apple Inc. who emphasized and marketed the wide range of resources that mobile devices offer consumers through the use of their slogan "There's an app for that."

The transformation of mobile devices has also changed the way people access and read books. Online bookstores have gained in popularity in recent years and they continue to attract more and more users on a daily basis. Amazon.com, one of the world's largest booksellers, announced that for the last quarter of 2010, sales of ebooks that were designed for use with their reader, the Kindle, outnumbered sales of hardcover books within their online store [4]. This demonstrates the extent to which ebooks are becoming widely accepted in people's everyday lives.

Ebooks are economical, more convenient and more environmentally friendly than printed books. People prefer ebooks for many reasons. These include the ability for consumers to achieve the following:

- Carry thousands of books wherever they go in a single device that weighs less than a paperback book.
- Buy new books wherever they are and whatever time of day or night it is.
- Access over a million free out-of-copyright ebooks in different languages.
- Download free sample chapters of ebooks and review them before they buy.

Today the availability of ebook readers means that consumers can carry a whole library in their pocket. However, it isn't just this capacity to store

and access books that constitutes the only factor that motivates people to choose electronic publications. These devices extend and enhance the reading experience through offering many different features, such as the ability to adjust screen brightness according to the environment, the option to configure text size, the opportunity to add notes or highlight text and electronic bookmarking devices. These, together with many other capabilities, enhance the reading experience and make reading a book more comfortable and exciting for both avid readers and those with disabilities.

With the ease of access to multilanguage ebooks, users have gained increased access to resources that can aid their process of learning a second language (L2). With just a single tap, readers can access ebooks from all around the world in different languages. For the intermediate learner of a foreign language, reading is key to the continued improvement and development of language skills. With the help of electronic publishing, people can download an ebook that is published in different languages within minutes and start reading it immediately. However, if a reader is not fully fluent in the foreign language, reading in L2 can be frustrating and time consuming. The need to frequently look up the meanings of words in a dictionary may be necessary. Ebook readers, such as iBooks, let readers define a word in order to help them understand the content. However, even if the reader knows the meaning of every single word in a given sentence, the sentence as a whole may be difficult to understand. This thesis sets out to suggest a solution to this problem.

In order to solve the perception problem involved with learning a L2, previous studies that have examined language learning and multilingual book formats were researched and the results of this research will be presented in the background research section of this thesis. After background research, it is realized that knowing individual words in L2 is not necessarily sufficient to understand sentences or content perception is difficult for people with disabilities to read. In this scenario, we believe that ebooks that present the original text and allow users, upon request, the ability to view sentences, phrases, idioms, or entire paragraphs in the second language and in a more understandable way, would greatly reduce the effort required by the reader. The need to develop this functionality motivated this multilingual ebook reader project, which was codenamed BiText Reader.

BiText Reader aims to provide a new way of reading for people who want to improve their proficiency in a second language through reading books. It aims to decrease the time it takes to translate foreign language content from a second language to the reader's native language and thus aspires to enhance language-learning experiences by providing readers with

access to multilingual content through an ebook reader. Besides providing multilingual content, it is also intended that the BiText Reader will support multilingual footnotes so that translators can provide additional information next to author's footnote so that content can be perceived even more clearly.

## 1.2 Purpose of Thesis

This thesis focuses on the development of an ebook reader application that is capable of managing the display of content in multiple languages while maintaining mapping between the linguistic structures of different representations. For the representation of the content, following EPUB standards, a multilingual content format is created in a way that structures multilingual data so that it can be used to achieve the two ultimate goals of this thesis: improving information perception and increasing the ease of foreign language learning through an ebook reader.

BiText Reader offers several additional benefits pertaining to understanding content that other ebook readers are currently unable to replicate. Some of the major features of the reader entail that it becomes a distinguished tool for language learning through the following provisions:

**Instinctive:** Through the multilingual content structure, publishers can provide their content in different languages and readers can interact with the device and access the publisher-provided translation of a sentence or whole paragraph through the use of instinctive gestures.

**Accessibile:** BiText allows readers who would like to read a book in the author's language but do not have enough proficiency in that language to easily check the corresponding text in their mother language. As such, BiText decreases the gap between the foreign language and the reader's native language.

**Informative:** BiText supports footnoting across multiple languages so that readers can access not only the translated version of the author's footnotes, but also the footnotes provided by the translator. This technique aims to help readers understand the content, not only by providing direct translation of author footnotes, but also considering the culture of the reader and providing friendly footnotes that are relevant to this culture and their environment.

**Extendable:** The capabilities of BiText can be extended in many ways. For instance, as languages change, texts in an earlier version of a language-either original texts or old translations may be difficult for modern

readers to understand. When readers have difficulties understanding an article that was written in an old version of their mother language, BiText can be used to provide additional information or translation in modern language, so that they can easily compare the evolution of their language.

**Universal:** Multilingual content structure is developed using the official EPUB standard of the International Digital Publishing Forum (IDPF). As such, multilingual ebooks can be opened with popular ebook readers. Within the experiments, the BiText sample book was tested on major readers, including iBooks on iOS and Ehon reader on the Mac desktop, and ebook content is accessible without any breaks.

## 1.3 Thesis Organization

The rest of this document is organized into the following chapters:

**Chapter 2:** Background Research. This chapter discusses the research that has been going on in the field of multilingualism and digital publishing.

**Chapter 3:** Technologies. This chapter gives a brief overview about the technologies employed in development of multilingual ebook reader.

**Chapter 4:** Solution Design. Multilingual content structure and BiText reader communication is introduced in this chapter.

**Chapter 5:** User Interface and Gestures. This chapter provides the design of gestural user interfaces for BiText reader.

**Chapter 6:** System Architecture. Transferable abstraction of BiText reader is introduced in this chapter and supported with UML diagrams.

**Chapter 7:** Conclusion and Future Work. Conclusion and the possibility for future improvements are discussed under this chapter.

# Chapter 2

# Background Research

*"Research is to see what everybody else has seen, and to think what nobody else has thought"*

Albert Szent-Gyorgyi

This section introduces basic terminology and concepts relating to multilingualism and ebooks. The primary intention of this chapter is to outline the theoretical framework that supports this thesis. This research is concerned with multilingualism and the use of multilingual electronic books to aid language learning. Consequently, this chapter initially discusses arguments in favor of multilingualism and multilingual books and then considers the use of electronic multilingual books in language learning.

## 2.1 Multilingualism

Multilingualism is the ability of an individual speaker or a community of speakers to use multiple languages [20]. However, the issue of what level of language-skills one should acquire in order to be qualified as multilingual is one of the fundamental questions in defining multilingualism. For this reason there are many descriptions of multilingualism. Multilingualism is relative. Mackey [19] therefore concludes that multilingualism is simply the alternate use of two or more languages. Age, gender, intelligence, memory, language attitude and motivation are factors that influence multilingualism.

## 2.2 The Advantages of Multilingualism

Multilingualism is becoming a social phenomenon that is governed by the needs of globalization and cultural openness. According to Webb [30], there

are definite advantages to having multilingual people in a society. Firstly, knowledge of more than two languages gives an individual access to the extensive knowledge and skills that are available among the millions of speakers of those languages. Secondly, multilingual people possess a more developed capacity for managing cross-cultural contact. Thirdly, they have a more empathetic attitude towards speakers of languages other than their own. Furthermore, researchers have reported numerous multilingual cognitive advantages of speaking more than one language. These include concept formation, creativity, visual-spatial abilities and analogical reasoning [22]. However, in order to become multilingual through formal instruction, learners need to go through a process of second language acquisition. Methods of learning a second language will be discussed in the following section.

## 2.3 Methods of Learning a Second Language

The mother tongue or first language is very easy to learn. However, many individuals experience difficulties learning a second language (L2) and it can take people longer to learn a second language than it did to master their native tongue. Furthermore, there are several reasons why people learn a second language. In some cases it may be due to their need to communicate with foreigners, or in others to get better work opportunities based on language knowledge. However, the most effective method for learning a language is subjective and varies according to the individual. Some of the methods by which one can learn a second language are as following [7]:

**Movies and Television** watching TV programmes or films with the subtitles on. Individuals can explore the different language subtitles on the DVD they are watching, or rent a foreign movie and read the subtitles in their own language.

**Music** Individuals can listen to songs that are sung in another language or find the lyrics on the Internet and follow them as they listen to the song.

**Newspapers and magazines** Language learners can read news that is written in a foreign language and learn new vocabulary in subjects such as economics, business and politics. If there are words that the learners don't understand, they can write them down and look them up in a dictionary later.politics. If there are words you don't understand, write them down and look them up in a dictionary later.

**Phrase Books and dictionaries** can be useful in emergency situations when an individual needs to know how to communicate something.

**Taking a class** some people prefer to learn languages in a classroom because it gives them a place to focus and a physical building where they can go to learn.

Research performed by the Foreign Service institute (FSI) estimated that 700 to 1,320 hours of full-time instruction is needed to achieve high levels of fluency in a foreign language [28]. However, most university students on average spend only 150 hours per academic year actively studying a second language. As such, in order to achieve the desired level of proficiency, students need to read books.

## 2.4 Multilingual Books

A multilingual book is described as a book that balances the use of two different languages. Multilingual books help learners to engage with the book and develop biliteracy.

There are different types of multilingual books available:

- Two books, one in each language.
- A single book that contains the text written in two languages.
- Books that contain flip-over pages with one language contained on each side.

Multilingual books provide an ideal way to help bilingual children to learn about their language and culture. The mother language is contained side-by-side on the page with the second language and thereby provides students with a variety of literacy opportunities. It shows students that the written word comes in many different forms. Moreover, kids are also able to connect with the stories as they "see themselves" in the multicultural characters in the books. One of the examples of this is shown in Figure 2.1. Image courtesy of Gisela Ernst-Slavit & Margaret Mulhern [15] from "Bilingual Books: Promoting Literacy and Biliteracy in the Second-Language and Mainstream Classroom."

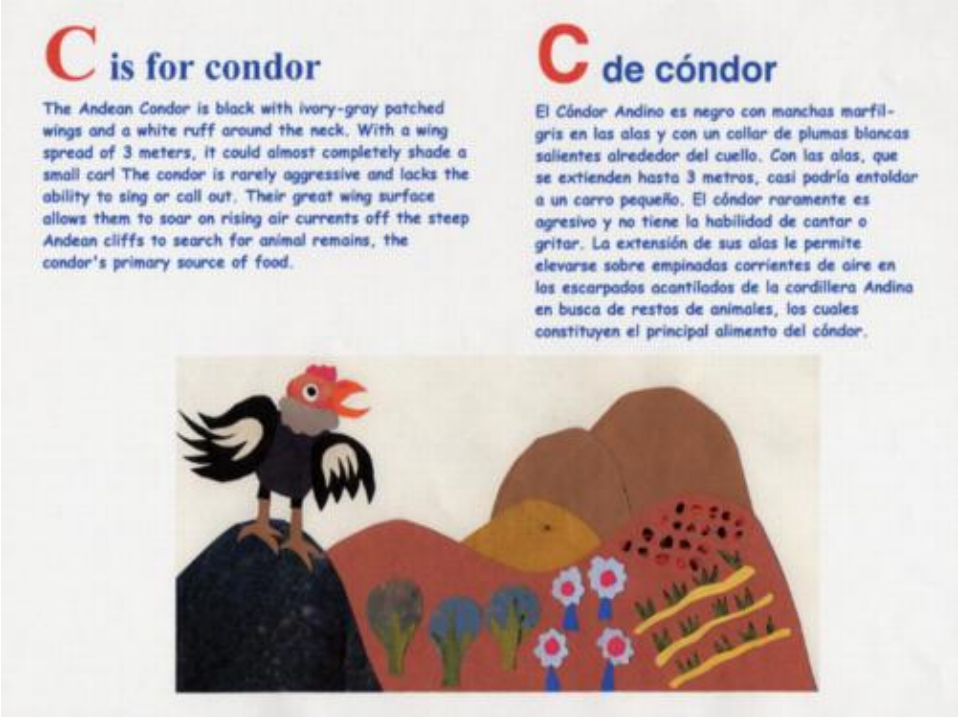
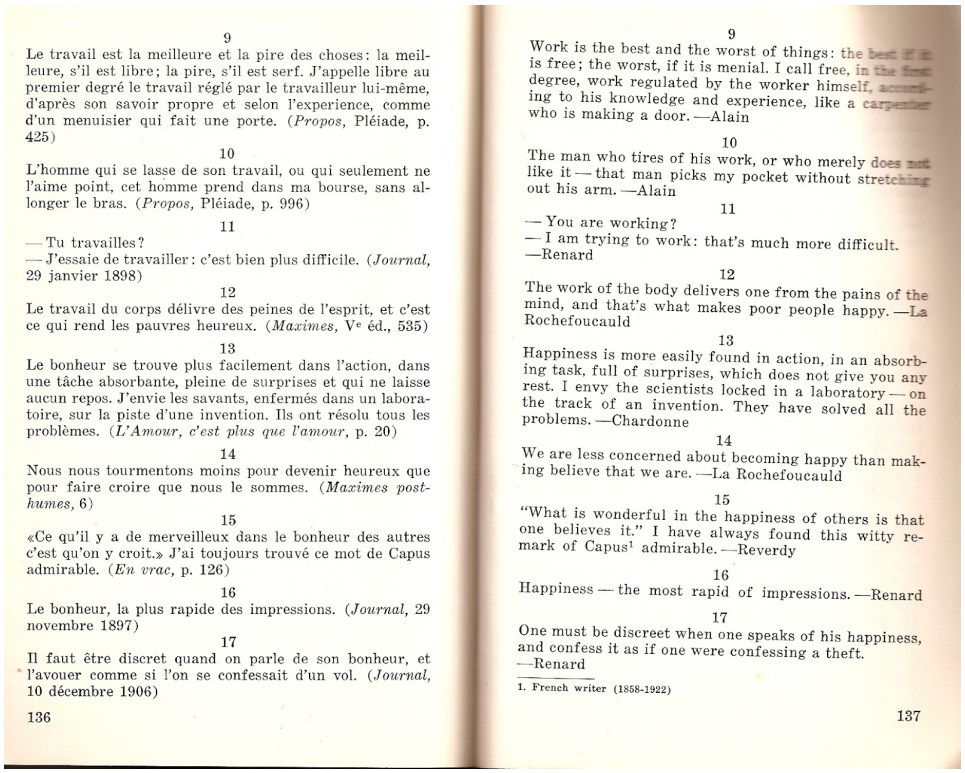



*Figure 2.1: Traditional bilingual text in English and Spanish*

Reading multilingual books has many advantages. Learners can read in the second language and use the first language text as a resource to verify their understanding. Additionally, they can compare original and contrast versions of a book that is written in two languages. Furthermore, multilingual reading activity encourages a good linguistic investigation so that readers can develop their proficiency in both languages.

Image courtesy of Joseph Palmeri [18] from "French Wit and Wisdom: A look at Life by Great French Writers".



9
Le travail est la meilleure et la pire des choses: la meilleure, s'il est libre; la pire, s'il est serf. J'appelle libre au premier degré le travail réglé par le travailleur lui-même, d'après son savoir propre et selon l'experience, comme d'un menuisier qui fait une porte. (*Propos*, Pléiade, p. 425)

10
L'homme qui se lasse de son travail, ou qui seulement ne l'aime point, cet homme prend dans ma bourse, sans allonger le bras. (*Propos*, Pléiade, p. 996)

11
— Tu travailles?
— J'essaie de travailler: c'est bien plus difficile. (*Journal*, 29 janvier 1898)

12
Le travail du corps délivre des peines de l'esprit, et c'est ce qui rend les pauvres heureux. (*Maximes*, Ve éd., 535)

13
Le bonheur se trouve plus facilement dans l'action, dans une tâche absorbante, pleine de surprises et qui ne laisse aucun repos. J'envie les savants, enfermés dans un laboratoire, sur la piste d'une invention. Ils ont résolu tous les problèmes. (*L'Amour, c'est plus que l'amour*, p. 20)

14
Nous nous tourmentons moins pour devenir heureux que pour faire croire que nous le sommes. (*Maximes posthumes*, 6)

15
«Ce qu'il y a de merveilleux dans le bonheur des autres c'est qu'on y croit.» J'ai toujours trouvé ce mot de Capus admirable. (*En vrac*, p. 126)

16
Le bonheur, la plus rapide des impressions. (*Journal*, 29 novembre 1897)

17
Il faut être discret quand on parle de son bonheur, et l'avouer comme si l'on se confessait d'un vol. (*Journal*, 10 décembre 1906)

136

9
Work is the best and the worst of things: the best if it is free; the worst, if it is menial. I call free, in the first degree, work regulated by the worker himself, according to his knowledge and experience, like a carpenter who is making a door. —Alain

10
The man who tires of his work, or who merely does not like it — that man picks my pocket without stretching out his arm. —Alain

11
— You are working?
— I am trying to work: that's much more difficult.
—Renard

12
The work of the body delivers one from the pains of the mind, and that's what makes poor people happy. —La Rochefoucauld

13
Happiness is more easily found in action, in an absorbing task, full of surprises, which does not give you any rest. I envy the scientists locked in a laboratory — on the track of an invention. They have solved all the problems. —Chardonne

14
We are less concerned about becoming happy than making believe that we are. —La Rochefoucauld

15
"What is wonderful in the happiness of others is that one believes it." I have always found this witty remark of Capus[1] admirable. —Reverdy

16
Happiness — the most rapid of impressions. —Renard

17
One must be discreet when one speaks of his happiness, and confess it as if one were confessing a theft.
—Renard

1. French writer (1858-1922)

137

*Figure 2.2: Flip over multilingual book in English and French*

A flip over multilingual book of Palmeri which is presented in the figure 2.2, contains texts both in French and English side by side. Although this type of representation provides text in two languages, switching between languages and finding the corresponding text in the L2 is not easy. Readers have to move their eyes back and forth from the L1 to L2 page, searching for the translation and this will cause breaks in reader's concentration. Ebooks can be used to handle this valuable multilingual content representation efficiently without the need of moving reader's eyes.

## 2.5 Ebooks

An ebook is an electronic version of a text that is composed in, or converted to, digital format for display on a computer screen or handheld device. Ebooks can use many different file formats; however, they all share certain attributes: they are portable, transferable, and searchable. Electronic media

can also incorporate other features, such as annotations, audio and video files, and hyperlinks. Most recently, publishers have added "interactivity" to ebooks as another set of enhanced features. Newer ebook readers, such as the Apple iPad, can render HTML5, CSS3 and Javascript. Publishers have used this opportunity to further engage readers with the book.

## 2.6 Multilingual Electronic Publications

Multilingualism is also used in electronic publications in order to reduce the effort taken to understand the content. For example: Skylife, Turkish Airlines' in-flight magazine displays multilingual content on their iPad App figure 2.3. The side-by-side presentation of English and Turkish content lets Turkish Airline passengers both read the article and also practice the language.


*Figure 2.3: Electronic multilingual magazine in English and Turkish*

Even though digital magazine in the figure 2.3 provides content in two languages, it does not suggest a solution for switching from L1 text to L2. Readers still need to search the text to find a paragraph or sentence in the proper language. This clearly shows that mimicking a paper book, is not a better way to exploit an ebook.

Multilingual books are also known as "diglots" and are gaining popularity in Learning L2 vocabulary, with the diglot weave technique now being used on a regular basis to aid language learning. The diglot weave technique involves inserting foreign words into sentences in a language reader.

Recently Disney Publishing released its first language learning iPad app that uses the diglot weave technique [8]. The diglot weave acts as a pedagogical approach to learning L2 vocabulary by inserting foreign words into sentences. Within the application, Toy Story 3 is narrated partially in the native language of the student and partially in the language being taught. As the learner's understanding progresses, an increasing proportion of the story is told in the foreign language until the reader experiences the story entirely in that foreign language.

# Chapter 3

# Technologies

*"Any sufficiently advanced technology is indistinguishable from magic."*

Arthur C. Clarke

## 3.1 Choosing Technologies

Choosing the right technologies is crucial to satisfy the two major development needs of the project: Multilingual digital content creation and multilingual ebook reader (BiText reader) development.

In order to create multilingual content, an ebook format is required. However there are many ebook formats available in the market for different purposes. On the other hand, each of them has both strengths and weakenesses. Since a universal ebook format do not exist, ebook format decision has to be made and it has a major role in the remaining development of the project. For that reason, an ebook format which has the most promising capabilities for the project needs, is chosen and it is EPUB.

EPUB which is also recommended as the standard by the American Association of Publishers [25], is the preferred format for multilingual digital content creation. EPUB standard internally uses XHTML to represent the document content and a subset of CSS to provide layout and formatting. Moreover XML is used to create the document metadata. Since EPUB is based on web standards, it benefits from other web technologies such as Javascript for interactive content creation. That clearly shows, EPUB provides a wide set of tools so as to create multilingual content.

After concluding ebook format decision, second important selection is to determine which device and programming language will be used for the de-

velopment of BiText reader. Choosing the right device is pivotal to demonstrate the characteristcs of multilingual digital content. For that reason, previous researches are investigated and BISG's Consumer Attitudes Toward ebook Reading report shows that Kindle, Nook, iPad and Android tablets are the major e-reading devices [21]. However Nook and Kindle ereaders are lack of integrated development environment(IDE), hence iPad and Android tablets could be utilized during the development. Between those two, iPad is selected as the target device due to its precise screen dimensions.

During the development of BiText reader, XCode IDE and naturally Objective-C programming language is used. Those are the official development tool and programming language for iOS development [16]. During this section each of the technologies will be explained and their roles in the project will be presented.

## 3.2 EPUB

EPUB is is a free and open ebook standard, an acronym of "Electronic publication" [12]. EPUB is developed and supported by the International Digital Publishing Forum (IDPF). EPUB defines a means of representing, packaging and encoding structured and semantically enhanced Web content — including XHTML, CSS, SVG, images, and other resources — for distribution in a single-file format.

It is an open standard improved by global contributors which makes it popular for structuring electronic books. "Open Standard" means that this standard can be used and modified without payment moreover anyone can create EPUB files and distribute them [27].

Today there are several tools available for automating the process of creating EPUB books. The abilities of these programs range from assisting in packaging the composed content into the EPUB format to providing a full editorial environment. Some of the popular EPUB editors are Adobe InDesign, Calibre, ePubIt and eCub. A use-case example showing how to export EPUB file in Adobe InDesign is provided during this section.

EPUB files can be used on a variety of operating systems and devices. It is one of the most supported ebook formats by mobile devices, since the text content is able to re-orient itself to fit to user's screen size. Some of the popular ebook reader applications are available for mobile devices such as Bluefire, Stanza or iBooks for iOS devices and Aldiko for Android devices. Moreover EPUB format is supported by many standalone ebook reading devices such as Sony Reader, iriver Story, Barnes & Noble Nook.

Key points about the EPUB standard can be listed as:

- EPUB is free and open standard
- EPUB uses re-flowable text.
- EPUB supports images, uses CSS and bases on web standards.
- EPUB can support DRM, but DRM is not required
- EPUB makes use of metadata

The EPUB file type uses ".epub" as the file extension which refers to two open standards, the Open Publication Structure (OPS) [11] and Open Container Format (OCF) [17], produced by the IDPF. EPUB's file extension allows publishers to produce and send a single file through distribution and offers consumers interoperability between software/hardware for unencrypted reflowable digital books.

Fundamentally EPUB file is a compressed folder whose structure is organized according to EPUB standards. Those standards defines a specicif flow to reach EPUB content and ebook properties. Following subsection expounds the files, folders and specifications inside the EPUB standards.

### 3.2.1 EPUB Structure

EPUB format is XHTML content wrapped in a .zip file package that contains XML files to describe the content and metadata. Additionally the zipped package consists of three main parts and EPUB file contains essentially the following files and folders:

- mimetype - provides file information about the package.
- META-INF folder - notifies the reader software about the content path of the electronic book.
- OEBPS folder - is the recommended location for the ebook contents and it contains:
  - images folder - holds ebook images.
  - Content.opf - lists the files inside the package.
  - toc.ncx - used for generating table of contents.
  - xhtml files - contains the book contents.
  - page-template.xpgt - is an optional file which is used by Adobe Digital Editions.

Each of these elements is discussed in detail below.

**MIME type** is an ASCII text file which tells to ebook reader how the ebook is formatted. Mimetype file contains "application/epub+zip" line. Ebook reader can interpret the filetype by using mimetype without the need of checking file extension. This file is required to be the first file in the zip file, and must be uncompressed.

**META-INF** is a folder that holds the XML container file. The basic META-INF consists of two components: Directory Folder and Container.xml. Directory folder is titled META-INF and it holds the container file. Container file points to the location of the Content.opf file. This is the first file a processor will search as it loads the book. Most container files for EPUB documents are similar. The purpose of the container document is to tell the device where to find the meta data information.

**OEBPS** is the folder that contains the content and navigational information for the ebook. A number of subsets make up the OEBPS folder. Each works to define and construct the content of the book. Similar to print books, EPUB works with small pieces that come together to bind the publication. Parts of the books such as the cover art, title, and chapters forms the parts of a book. This list covers the minimum subfolders necessary for a valid OEBPS.

**Images** is a folder for all the graphics in the book. EPUB requires images to be local so that they can be accessed offline. Most reading systems support a variety of images, but according to the OPF spec, only PNG must be supported by reading system.

**Content.opf** is an informational document that explains the OEBPS files. In more detail, this file provides a list of all files in the .epub container, defines the order of files, and stores meta data such as author, genre, publisher, etc. information. This file's name is independent meaning that any file name could be used as long as the container.xml file mentioned above points to the correct filename. It should be noted that filenames are case sensitive and container.xml should be case-sensitive formatted.

**Toc.ncx** is the table of contents for the book. An EPUB reader application uses this file to produce a digital list of book chapters. Additionally It helps to organize the menu system, so the person reading the ebook can navigate properly. However it should be noted that, each reading application will display the contents of toc.ncx differently. If publishers

needs to present a formatted table of contents to the reader, they can compose a XHTML file with the contents formatted in the way they prefer.

**Content XHTML files** are the actual pages of the book written in XHTML. Those files must conform to XHTML 1.1 syntax. Content files can be formed in any way publishers prefer. They can be all in one document with bookmarks at each chapter, or each chapter in a separate .xhtml file. However dividing content in to seperate .xhtml chapters gives a better performance since this way, ebook readers do not have to load whole content at once but they load part by part.

Following image 3.1 displays the folder structure of an unzipped EPUB file.


*Figure 3.1: EPUB folder structure*

### Content.opf

The Open Packaging Format file (OPF) acts as a container manifest and defines all of the files included in the EPUB zip file. OPF standard must provide a package document. This must be an XML document with a root element of `<package>` which includes elements called `<metadata>`, `<manifest>`, and `<spine>`. Figure 3.2. shows an overview of the package document for a sample epub ebook.

The OPF file is an XML file that consists of three elements:

### metadata

The `<metadata>` element of the package can contain wide ranging information about the publication. To keep OPF as open as possible, the metadata of an OPF package makes use of another open standard, namely the Dublin Core Metadata Intiative (DCMI). DCMI Metadata Terms provides a rich set of attributes that can be applied to electronic publications. Following three of them are mandatory:

**Title** represents the title of the book.

**Identifier** represents the unique identifier of the book. It is possible to publish an ebook with several identifiers such as internally generated identifier, a GUID, or print book's ISBN.

**Language** represents the language used in the book content. Language value must comply with RFC 3066.

### manifest

The package manifest identifies all of the resources that are needed to display the ebook. Each entry in the manifest consists of an `<item>` element for defining the pages and resources used in the `<spine>` element. Each item is also assigned an item ID that's used in the spine section of content.opf. This list is not necessarily in a particular order. Each `<item>` element has an 'id' attribute which identifies the unique resource and an 'href' attribute which points to the content document. For example in the Figure 3.2, an XML document identified with chapter-001 and linked to 'chapter-001.xml'.

### spine

The spine section lists the reading order of the contents. To prevent the confusion, note that the spine element do not necessarily list every file in the manifest, it only lists the reading order linearly such as: title page, followed by the author's note, followed by chapter and continues. Additionally `<spine>` element has a 'toc' attribute with the value 'ncx'. This id value is used to identify the table of contents. Moreover `<Spine>` element contains a list of `<itemref>` elements. Each `<itemref>` element has an optional attribute called 'linear' which is used to indicate whether the referenced document is primary or auxiliary. In the example, the values are all set to 'yes' which is the default.

```
<?xml version="1.0" encoding="UTF-8"?>
<package xmlns="http://www.idpf.org/2007/opf" unique-identifier="EPB-UUID" version="2.0">
   <metadata xmlns:opf="http://www.idpf.org/2007/opf" xmlns:dc="http://purl.org/dc/elements/1.1/">
      <dc:title>The Pit and the Pendulum</dc:title>
      <dc:creator opf:role="aut" opf:file-as="Edgar Allan Poe">Edgar Allan Poe</dc:creator>
      <dc:date opf:event="original-publication">1843</dc:date>
      <dc:publisher>Carey &amp;amp; Hart</dc:publisher>
      <dc:date opf:event="epub-publication">2012-08-15</dc:date>
      <dc:subject>Horror</dc:subject>
      <dc:subject>Short story</dc:subject>
      <dc:rights>Public Domain.</dc:rights>
      <dc:identifier id="EPB-UUID">urn:uuid:4B63B5CAFASCS</dc:identifier>
      <dc:language>en-gb</dc:language>
   </metadata>
   <manifest>
      <!-- NCX -->
      <item id="ncx" href="epb.ncx" media-type="application/x-dtbncx+xml" />
      <!-- Content Documents -->
      <item id="titlepage" href="title.xml" media-type="application/xhtml+xml" />
      <item id="book-01" href="book-01.xml" media-type="application/xhtml+xml" />
      <item id="section-001" href="section-001.xml" media-type="application/xhtml+xml" />
      <item id="chapter-001" href="chapter-001.xml" media-type="application/xhtml+xml" />
      <item id="end" href="end.xml" media-type="application/xhtml+xml" />
      <item id="main-css" href="css/book.css" media-type="text/css" />
      <!-- Images -->
      <item id="img-bookcover-jpeg" href="images/bookcover.jpg" media-type="image/jpeg" />
   </manifest>
   <spine toc="ncx">
      <itemref idref="titlepage" linear="yes" />
      <itemref idref="book-01" linear="yes" />
      <itemref idref="section-001" linear="yes" />
      <itemref idref="chapter-001" linear="yes" />
      <itemref idref="end" linear="yes" />
   </spine>
</package>
```

*Figure 3.2: EPUB OPF file*

### toc.ncx

NCX (Navigation Control file for XML) is an open standard which is maintained by the DAISY consortium. NCX contains the information about the structure of EPUB file in XML format. It has an important role in the hierarchical listing of 'navigation points' and presenting how the EPUB file is subdivided. This list generally used by the EPUB reading application to provide quick access to the EPUB chapters.

A `<navPoint>` element holds the text to be displayed to the reader in the `<navLabel>` element. EPUB reading applications read navigation point element to learn where to go for the content document. Content document file is held in the src attribute of the `<content>` element. Following code snippet presents a simple `<navPoint>` structure for NCX file.

```
<navPoint id="chapter01" playOrder="1">
     <navLable>
          <text>Chapter 1</text>
     </navLable>
```

```
    <content src="chap01.xhtml" />
</navPoint>
```

The <navMap> element contains an array of <navPoint> elements. Each navpoint represents a chapter listing, the text shows the chapter name, and the src links the source file. Moreover the <navPoint> element has attributes 'id' and 'playOrder'. 'id' stands for a unique identifier and 'playOrder' is a number, beginning from 1, that indicates the position of the navPoint in the sequence of content documents creting the EPUB.

The complete specification for NCX can be found in Section 8 of the Specifications for the Digital Talking Book.

#### container.xml

The container.xml file specifies the locations for the OEBPS folders that contain the content files and the OPF XML files. Container.xml file must be placed in the root folder's META-INF directory.

```
<?xml version="1.0" encoding="UTF-8"?>
<container xmlns="urn:oasis:names:tc:opendocument:xmlns:container" version="1.0">
   <rootfiles>
      <rootfile full-path="OPS/epb.opf" media-type="application/oebps-package+xml"/>
   </rootfiles>
</container>
```

*Figure 3.3: EPUB container file*

### 3.2.2 Creating an EPUB file

Publishers can export a document or book in EPUB format using the Adobe InDesign, an authoring editor frequently used to composit books. In order to export a document from InDesign, all an editor needs to do is clicking on "export for epub" from file menu, specify the desired options and save the epub file. Digital editions export window is displayed in the figure 3.4.

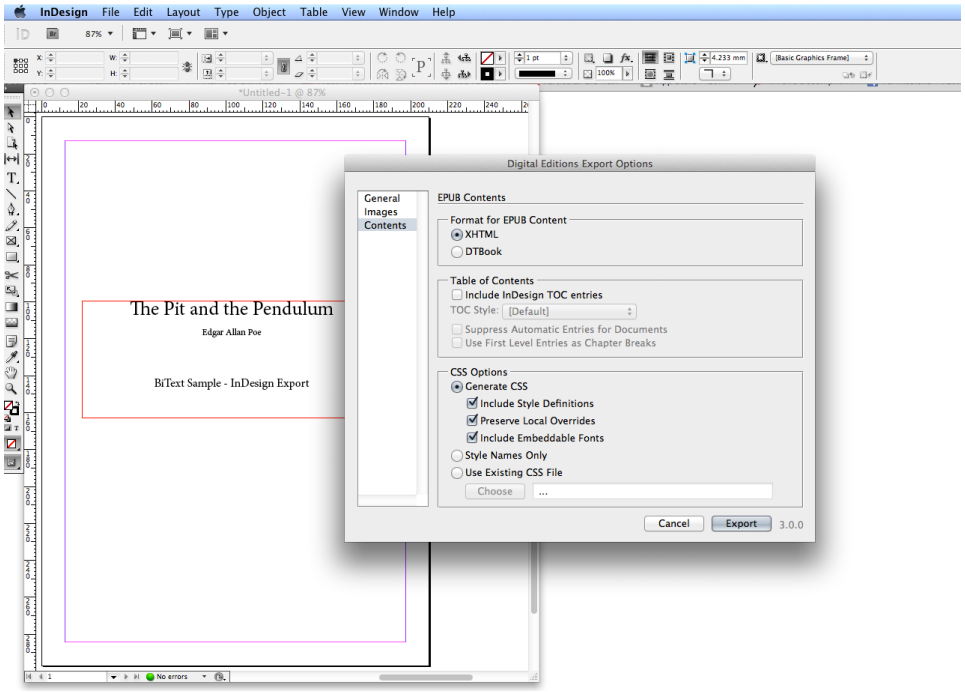



*Figure 3.4: Adobe Digital Editions export view*

Later further CSS editings can be done by changing the extension from .epub to .zip, and extracting the contents. Moreover it is possible to preview created EPUB file with Adobe Digital Editions which is a ebook reader application that comes with Adobe InDesign and it is also available as free download from Adobe site. Following Figure 3.4 shows preview of EPUB file on Adobe Digital Editions.

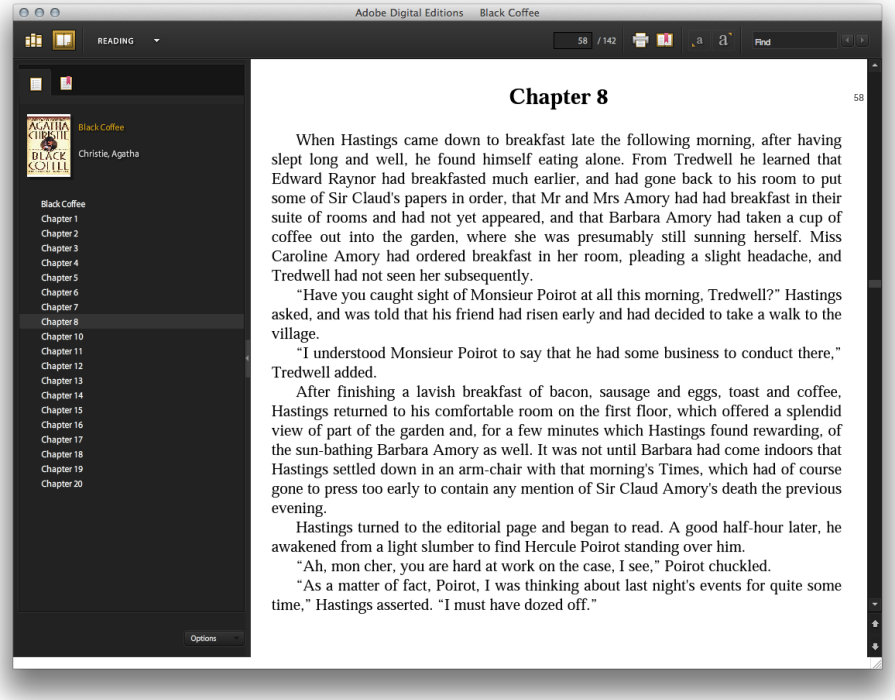



*Figure 3.5: Adobe Digital Editions EPUB preview*

Although Adobe InDesign is easy to compose EPUB files, due to its commercial distribution, it might not be suitable for amateur publishers or beginners. For that case there are also alternative open source applications that can be used to construct the EPUB file and compose the required metadata. Calibre and Sigil are two of the most popular open source multi-platform EPUB ebook editors.

### 3.2.3 Advantages of the EPUB

In May of 2008 the American Association of Publishers officially announced their support of the EPUB format as the standard ebook format. Additionally publishers supporting the EPUB format with content include Penguin Group USA, HarperCollins Publishers, John Wiley & Sons Inc., Hachette Book Group USA, Random House, and Simon & Schuster.

The IDPF describes the several advantages of the EPUB format on their Web site:

- EPUB files can be created by software most book designers already have such as InDesign and Quark.

- Publishers can reduce the cost of conversion by only creating EPUB files for multiple distribution channels.
- Content will be able to be sold from a multitude of outlets.

### 3.2.4 EPUB Reader

EPUB reader is an application which is used to open and display content of an EPUB file. It can be built by an application developer for running on various operating systems or devices such as mobile devices, desktop computers and even as an extension to browsers moreover it can optimize text for a particular display device. Functionality wise an EPUB reader opens the archived EPUB file reads the container xml and interprets the standard folder structure then renders the content on user's display. Beside displaying book content on the ePub, well structured standards provided by EPUB gives chance to build innovative readers. Those EPUB readers can be general purpose readers which simply works to open built in EPUB files, or they can be developed for wide range of purposes. Such as, Apple's iBooks application which is also an EPUB reader, supports reading many formats moreover it provides access to iTunes Bookstore to purchase new books. Additionally it gives users opportunity to bookmark pages, search the eletcronic book or change the font size to make reading activity more comfortable.

## 3.3 Javascript

Javascript is a lightweight, object-oriented scripting language used extensively for web pages however its capabilities are beyond web development and it is also used in many non-browser environments such as node.js or Couchbase [13]. It was jointly created by Netscape and Sun and today officially managed by Mozilla Foundation. Javascript is derived from java but it is not java, it is a platform Independent, interpreted language, works with integration in to HTML.

Javascript is designed for programming user events. With the help of Javascript, static HTML pages can be constructed event driven such interactive ways form a more user friendly user experience. Javascript is a client side programming language which is executed line by line on the client's browser. Since it runs on client's browser, it can manage certain type of operations without requiring the need of reloading from server.

The factors making Javascript a must use technology over many other scripting languages are: being open-source, globally acceptance and its sup-

port for both desktop and mobile browsers. Furthermore Javascript frameworks such as jQuery, makes Javascript even more convenient by encapsulating commonly used functions and providing design patterns. Moreover Javascript is simple to edit and run. A browser and a text editor is enough to start Javascript development. Hence, anyone can easily use Javascript as a script language for developing any web based application.

Some of the use cases of Javascript can be listed as:

- Javascript can be used to make HTML pages more dynamic.
- Javascript can react to various events which is not possible through HTML.
- Javascript can be embedded easily into an HTML document and JavaScript easily reads and writes HTML elements on the fly.
- Javascript can be used for the users' browser and device properties detection and can load the page according to the requirements of that browser.
- Javascript can generate Cookies.

It is possible to add numerous other features to that list. For instance, in the BiText reader development, Javascript is used for handling users' touch events on EPUB content and triggering Objective-C functions by sending document location change notifications to custom web view. Use of Javascript in the project will be explained in the solution design section.

## 3.4 CSS

CSS, acronym for "Cascading Style Sheets" [6], is a tool for styling and displaying the HTML (Hyper Text Markup Language) based pages. With CSS, styles such as margins, fonts, background images and colors can be defined. Main task of cascading style sheet is to tell interpreter how a web page should be displayed.

CSS styles can be saved in an external CSS file and can be used throughout the web site. Thus CSS provides a way to tidy style elements of a page in one place, which not only helps developers to modify the styles easily but also provides opportunity to display a content differently by using different CSS files for different user groups.

CSS separates the text content from presentation aspects, making the code extensible and easy to manage. It gives complete control over the

layout and design of HTML pages. This can be used to make web pages accessible from different sources by defining different CSS rules for different device such as tablets, smartphones or desktop computers.

Together with advantages of CSS, some of the reasons for using CSS can be listed as:

- Global and Local Formatting: Global page element styles can be defined to form the base of style sheet and local classes can be added inheriting from global styles that way styles can be formed both faster and more reliable.
- Readability: CSS behaves like a styling tier addition to presentation layer by gathering all styling rules in one place. This leads to faster, reliable and readable coding for team of developers.
- Browser Cache: when a web page is loaded the style sheet for the web page can be saved to the browser cache, and can consequently be used on multiple pages without being reloaded. This reduces data transfer over the network leading to increased bandwidth.

## 3.5 XML

Extensible Markup Language (XML) is a markup language that allows developers to describe the structure of a document's data entirely in text, using tags that can be defined arbitrarily. XML is defined in the XML 1.0 Specification produced by the W3C and the rules governing the structure are specified in a language schema such as DTD (Document Type Definition). A DTD uses a terse formal syntax that declares precisely which elements and references may appear where in the document of the particular type, and what the elements' contents and attributes are.

XML can be classified depending on markup purposes to following classes:

- Presentational Markup
- Descriptive Markup
- Procedural Markup

XML is commonly used in web services for serializing and sharing structured data between applications, systems, and even organizations using a standard format. Moreover Encoding objects and data into XML documents allows developers to easily view and modify objects and data in their serialized form. In addition, when receiving streams of serialized data over

the Internet, developers require to make sure that the document is valid before deserializing it. XML serialization provides DTD with facilities to accomplish this.

The role of XML is essential in the state of ebook design. Mike McNamara, managing director at Araman Consulting Ltd & Outsell-Gilbane UK Affiliate, states that "There are many benefits to be gained from implementing XML in a production workflow. However, it depends on what the publisher wants to do. For example, journal publishers probably can reuse their content in a number of different ways for differing products and specific target markets. XML can deliver this flexibility and reusability" [29].

## 3.6 JSON

JavaScript Object Notation (or JSON ) is a text-based open standard designed for human-readable data interchange [14]. It is derived from the JavaScript scripting language for representing simple data structures and associative arrays. The JSON format is often used for serializing and transmitting structured data over a network connection. It is used primarily to transmit data between a server and web application, serving as an alternative to XML. JSON is capable of representing:

- Number (double precision floating-point format in JavaScript, generally depends on implementation)
- String (double-quoted Unicode, with backslash escaping)
- Boolean (true or false)
- Array (an ordered sequence of values, comma-separated and enclosed in square brackets; the values do not need to be of the same type)
- Object (an unordered collection of key:value pairs with the ':' character separating the key and the value, comma-separated and enclosed in curly braces; the keys must be strings and should be distinct from each other)

JSON does not natively represent more complex data types like functions, regular expressions or dates. However values can be transformed when they are serialized, or prior to deserialization, to enable JSON to represent additional data types.

## 3.7 Objective-C

The Objective-C language is a high-level programming language designed to enable sophisticated object-oriented programming. Objective-C is defined as a small but powerful set of extensions to the standard ANSI C language [2]. It is the main programming language used by Apple for the OS X and iOS operating systems and their respective APIs, Cocoa and Cocoa Touch. Originally developed in the early 1980s, it was selected as the main language used by NeXT for its NeXTSTEP operating system, from which OS X and iOS are derived [1]. Objective-C is a layer designed on top of C, and moreover is a strict superset of C; it is possible to compile any C program with an Objective-C compiler, and to freely include C code within an Objective-C class.

# Chapter 4

# Design of Multilingual Content Representation on EPUB

*"Above all else show the data."*

Edward R. Tufte - Visual Display of Quantitative Information

In order to develop multilingual ebook reader, it is initially required to construct a definitive multilingual content representation on EPUB. This section presents how EPUB standard is used to provide a functional multi language book content without breaking EPUB standards. Moreover this section introduces the creation of a sample multilingual EPUB ebook that uses the proposed content structure.

## 4.1 EPUB Content Design

In the technologies section, it is mentioned that EPUB is chosen as the ebook format, due to its support for creating reflowable content based on web technologies such as XHTML, Javascript and CSS. Those technologies are used as the foundation of creating multilingual interactive content.

During content design, specifications of EPUB is followed so as to construct the multilingual content accessible not only by the BiText reader implemented in this thesis but also in other ebook readers.

## 4.2 Multilingual Content Representation

Multilingual representation of content with XHTML tags are required in order to render content and connect text fragments between languages. Figure 4.1 shows the representation model of multilingual content.

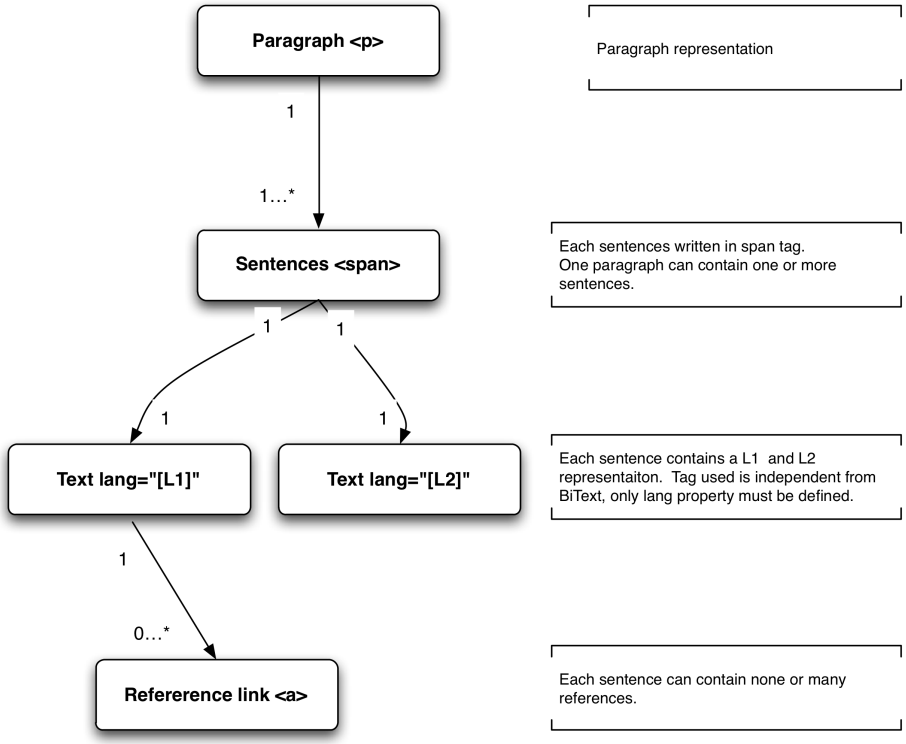



*Figure 4.1: Multilingual content representation model*

As it is shown in the figure, content is segmented at sentence level. Moreover a sentence can contain footnote(s) by the author or the translator. In order to clarify the model, a sample paragraph with single line of text is presented below.

```
<p>
        <span>
                <i lang="en">Sentence1 original text in L1.</i>
                <i lang="it">Sentence1 translated text in L2.</i>
        </span>
</p>
```

It is possible to add continuous lines of texts in this structure. More examples will be provided during the explanation of multilingual ebook construction in this section.

## 4.3 Content Interaction

After designing multilingual content represention in XHTML files, it is time to develop interaction functions and presentation styles. In order to achieve a tidy, reusable and extendable content structure, It is essential to encapsulate interactions and styles from the EPUB content. For that reason BiText Core is introduced.

BiText Core will be covered particularly in section 6 during software design however without going profound details, it is helpful to make a brief introduction so as to understand the interaction approach.

BiText Core consists of a Javascript and a stylesheet file. It is designed to let publishers integrate multilingual content effortlessly. Publishers can add BiText Core to their multilingual content with three lines of code and provide multilingual support without breaking EPUB standards. This allows other ebook readers to display the original content. Eventually those readers will not be able to exploit the multilingual features, but at least the original text is still readable without any break. Following three lines shows how to include BiText core in a XHTML content.

```
<meta name="BITEXT" content="CORE-20120902"/>
<link href="bitext/core.css" type="text/css"/>
<script src="bitext/core.js" type="text/javascript"/>
```

First line, BiText meta is used for notifying BiText reader that the rendered content supports multilingual data. Optionally if multilingual content is supported during the whole ebook, meta data can be added to content.opf file so that whole book content will be interpreted as multilingual by default. Following that, core.css file contains the css styles so as to manage which language's content will be displayed and which will be hidden. Finally, core.js file contains required functions in order to handle interaction events, communicate with the BiText reader and modify multilingual content.

## 4.4 Content Structure

Abstraction of the multilingual content structure can help to understand the major elements of multilingual EPUB more clearly. Figure 4.2 shows

the files and folders included in the final structure of the multilingual content. In order to use EPUB standards, mimetype file is built to provide package information, META-INF directory is created to hold the container file which links BiText reader application to OPS content folder. Content folder as specified in EPUB standards holds images that are used in the ebook, OPF file is used to list the files included in the package, NCX file is used to construct the structure of ebook index and multilingual contents are embedded in to XHTML book contents.

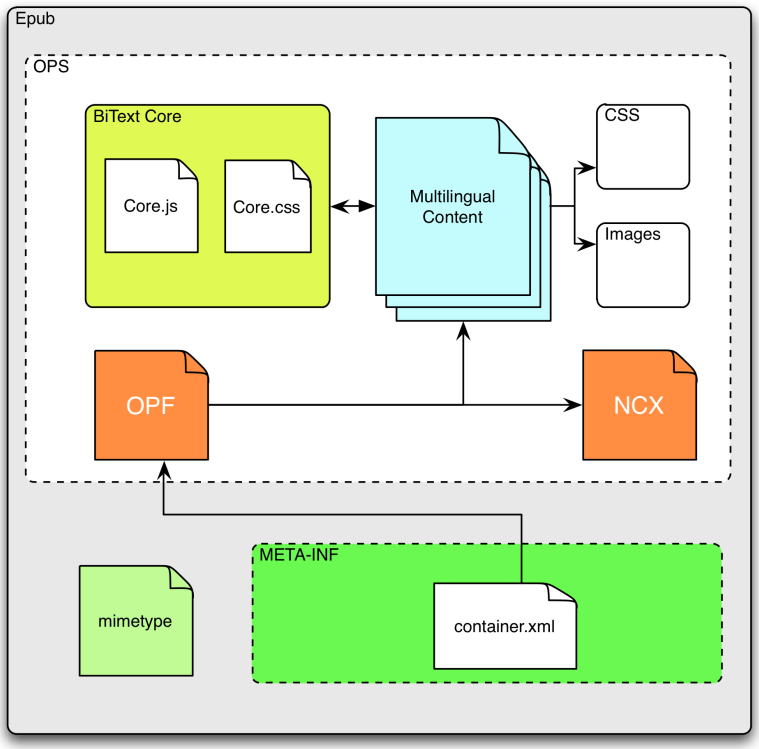



*Figure 4.2: Multilingual content structure*

As it is shown in the figure only XHTML files which are open to interactive content development, need BiText Core files. Those BiText Core files are utilized to support multilingual features and they will be covered in section 6.6. Furhtermore, as it is stated earlier, original text can be displayed by any ebook reader which supports EPUB format however they can not benefit from multilingual features.

## 4.5 Multilingual Ebook Construction

Before the further development of the BiText reader, it was essential to construct a functional multilingual ebook in suggested multilingual EPUB content structure. Fortunately there exists several sources which offer digitized classic books in public domain, such as Wikisource, Project Gutenberg, Oxford Text Archive, Berkeley Sunsite Classics and many others. In the following subsection, sample book selection and sources will be explained.

### 4.5.1 Sample Book and Translation Source

A short story written by Edgar Allan Poe, "The Pit and the Pendulum" is chosen for the development of the multilingual ebook sample. This story which is published before January 1, 1923, is in the public domain worldwide since the author passed away more than 100 years ago. The reason for choosing this short story not only its public domain and availability in many languages, but also its excellent use of literature and fluency. That way it can also be used in the future works while evaluating the effectiveness of the method proposed in this thesis.

Wikisource website was the source for obtaining this public domain [5] novel so as to use during the project. Wikisource identifies itsself as a repository of source texts in any language which are either in the public domain, or are released under the cc-by-sa 3.0 license [32]. Additionally literary translation in italian is obtained from edgarallanpoe.it [9] where the text was published courtesy of the provincial Library "Melchiorre Delfico" of Teramo (La Biblioteca provinciale "Melchiorre Delfico" di Teramo).

### 4.5.2 Multilingual Content Organization

After obtaining literary work and its translation in Italian. Creation of the BiText ebook started with plain text and the multilingual EPUB file is constructed from scratch. During the process, literary work and translation are aligned at sub-sentence level. So each paragraph and sentence in the original version is linked with the corresponding translation. That way, users can read the original version and consult the translation at any moment.

A real multilingual content representation from 'The Pit and the Pendulum" is presented in Figure 4.3. It also shows how BiText Core is included in the sample EPUB content.

```
<?xml version="1.0" encoding="UTF-8"?>
<!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" xml:lang="en">
   <head>
      <title>The Pit and The Pendulum</title>
      <link rel="stylesheet" href="css/book.css" type="text/css"/>
      <meta http-equiv="Content-Type" content="application/xhtml+xml; charset=utf-8"/>
      <meta name="EPB-UUID" content="4B63B5CA-6C01-1014-9C05-F192C088E586"/>
      <!-- BITEXT CORE EMBED -->
      <link href="bitext/core.css" rel="stylesheet" type="text/css"/>
      <script src="bitext/core.js" type="text/javascript"></script>
   </head>
   <body>
      <div class="body">
         <div class="section">
            <p>
               <span>I felt nothing; yet dreaded to move a step, lest I should be impeded by the
walls of a tomb.<i lang="it">Non sentivo niente, e però tremavo all'idea di muovere un passo, avevc
paura d'urtare contro i muri della mia tomba.</i>
               </span>
               <span>Perspiration burst from every pore, and stood in cold big beads upon my
forehead.<i lang="it">Il sudore mi usciva da tutti pori e si raccoglieva sulla fronte in grosse
gocce.</i>     </span>
               <span>The agony of suspense grew at length intolerable, and I cautiously moved
forward, with my arms extended, and my eyes straining from their sockets, in the hope of catching
some faint ray of light.<i lang="it">La sofferenza dell'incertezza mi divenne sempre più
insopportabile, e mi avanzai con prudenza, con le braccia in avanti e gli occhi fuori dell'orbita,
sperando di scorgere qualche debole raggio di luce.</i>
               </span>
            </p>
```

*Figure 4.3: Multilingual XHTML content sample*

Additionally it is possible to add footnote to any word or sentence by using footnote method "showRef" as following example:

```
<span>I was sick-sick unto death with that long <a href="#" rel="r1"
onclick="return showRef(this,'a');">agony<sup>A</sup></a>;
and when they at length unbound me, and I was permitted to sit,
I felt that my senses were leaving me.<i lang="it">Quella lunga agonia
mi aveva affranto, ero stremato di forze, e allorche finalmente mi
slegarono e potei sedermi, sentii che perdevo i sensi.</i>
</span>
```

First parameter for the showRef function is the object itsself and the second one is to identify footnote type, it can be either 'a' for author, or 't' for translator. Footnote corresponding to element can be positioned to any suitable place inside the same XHTML content. Reference classes are automatically hidden by BiText Core stylesheet and BiText Core scripts can find them using elements unique reference class tag. Following example shows a possible footnote definition.

```
<div class="reference r1 en">
       The Politecnico was founded on 29 November 1863.(N.d.A)
</div>
<div class="reference r1 it">
       Il Politecnico, fondato il 29 novembre 1863.(N.d.A)
</div>
```

### 4.5.3 Multilingual EPUB File Creation

After creating the required folder structure and contents of EPUB, next step is creating the zip container and generating .epub file so as to install the book on a reader. The rules for the ZIP Container in the Open Container Format Specification state: "The first file in the ZIP Container MUST be a file by the ASCII name of 'mimetype' which holds the MIME type for the ZIP Container (i.e., 'application/epub+zip' as an ASCII string; no padding, white-space or case change). The file MUST be neither compressed nor encrypted and there MUST NOT be an extra field in its ZIP header. If this is done, then the ZIP Container offers convenient 'magic number' support as described in RFC 2048 and the following will hold true" [17].

In order to complete compression operation correctly on Mac OS, it is required to omit DS_Store files, which are invisible files in Mac OS folders. Otherwise it will cause validation errors. Following code snippet is used to compress all EPUB files and folders excluding DS_Store files.

```
zip -Xr9Dq sample.epub * -x *.DS_Store
```

# Chapter 5

# User Interface and Gestures

> *"The location of visual elements in the UI has a huge impact on how the user interprets information."*
>
> Rick Oppedisan

This section introduces the design process of gestural user interfaces for multilingual ebook reader. The primary intention of this chapter is to demonstrate approach to user interfaces and interactions. According to Wigdor and his co-author the goal of designing natural user interfaces for touch and gesture is to build a user experience that is natural to the user, rather than somehow intrinsically natural [31]. In order to design a naturel user interface, following design principles suggested by Wigdor and they are commonly used in the design of BiText Reader user interfaces.

- Create an experience that user can feel like an extension of their body.
- Build a user interface that considers context, including the right metaphors, visual indications, feedback and input/output methods for the context.

## 5.1 Gestural User Interfaces

The design of BiText reader starts with defining the reading states that will be available to user with a gesture. There are three major reading states to display: inline translation state, paragraph translation state and footnote state.

Challenging part of the user interface, those reading states should all be triggered by interacting with the same text fragments in a way that application can recognize which function is requested by the user. Moreover at the same time interactions is required to be discoverable.

Dan Saffer defines the discoverability as one of the characteristics of good gestural interfaces[23] and adding that before an interaction starts with a gestural system, one has to know that, there exists an interactable element and how to begin to interact with it. As Saffer explained, being discoverable is a major issue in a gestural interface. In order to find out the right gestural interface, first mockups are prepared to show what we want to achieve. Figure 5.1 presents inline translation state of reading.

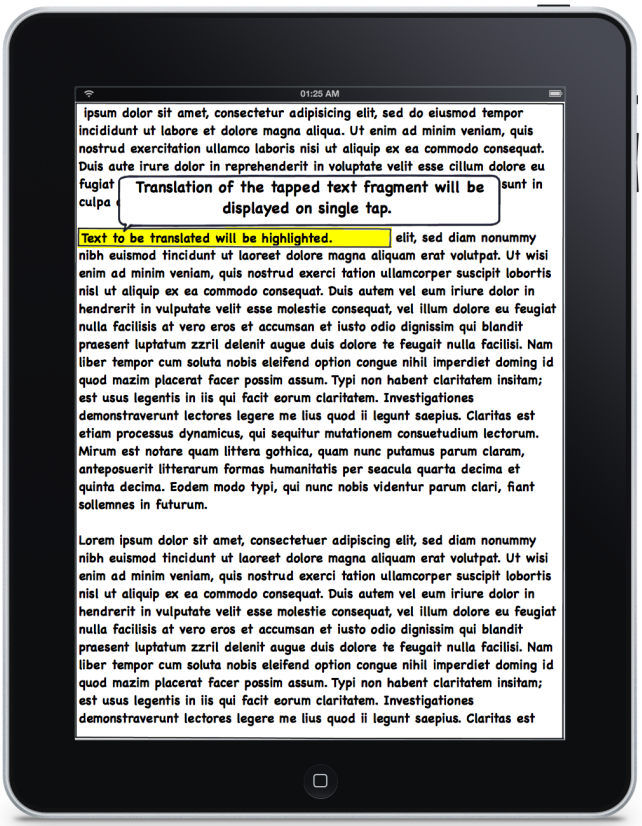



*Figure 5.1: BiText reader mockup for inline translation state*

It is clear that interaction for inline translation can be achieved with a tap gesture on the sentence. However following Figure 5.2 shows paragraph translation state which requires again a similar gesture. Moreover those gestures should be effortless and distinguishable from each other.

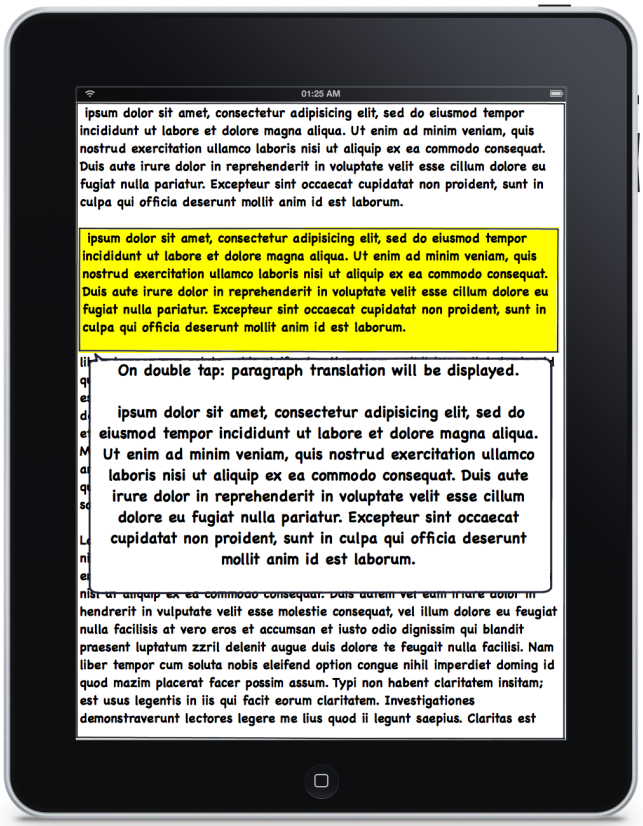



*Figure 5.2: BiText reader mockup for paragraph translation state*

Following guidelines found in "Designing Gestural Interfaces"[24], tapping an object with a single finger should select that object for manipulation. Moreover Tap to Select is usually the alternative to Tap to Open. Either one tap selects an item and a double tap to open selected item. In the design of paragraph selection, we used a similar gesture. Single tap gesture is reserved for selecting the inline translation and double tap is used to select the whole paragraph covering the touched content.

Following interaction diagram 5.3 shows the gesture functionality flow.

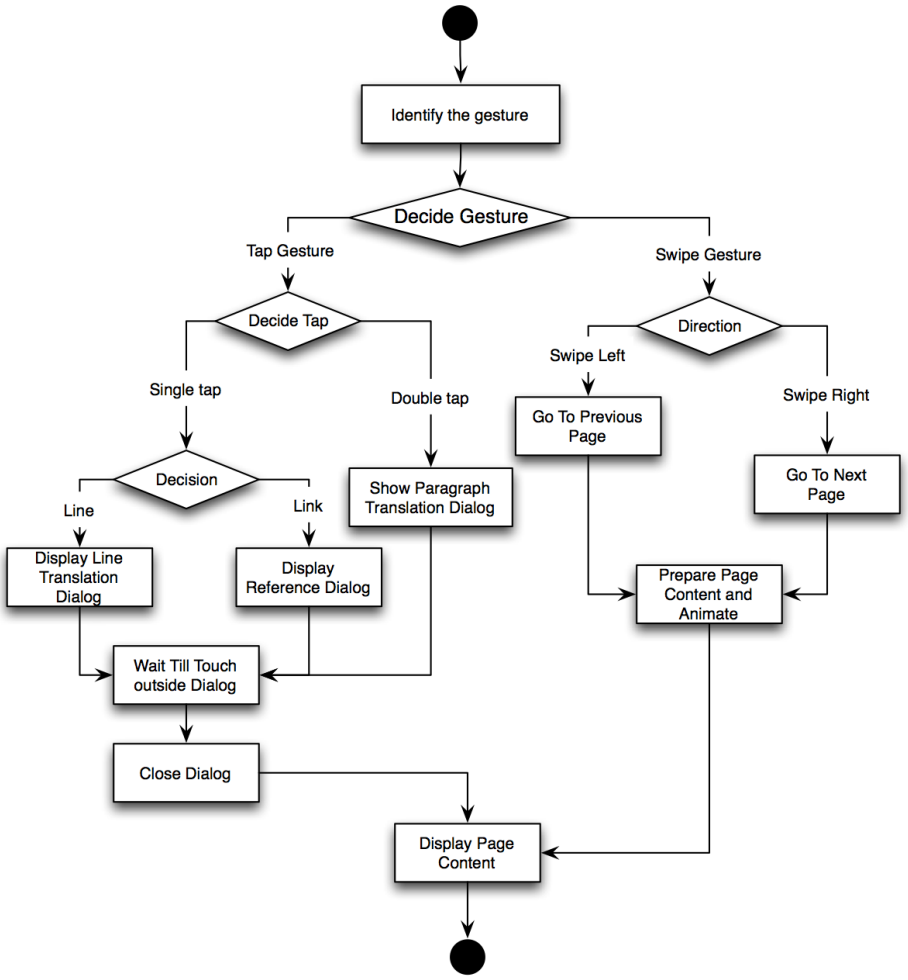



*Figure 5.3: BiText reader gesture functionality flow*

As it is shown in the flow there are four gestures on the EPUB content and each of them activate a function. Gestures and related functions are listed as below:

**Swipe Left:** Go to previous page with a page flip animation.

**Swipe right:** Go to following page with a page flip animation

**Single tap on Link:** Display footnote dialog with pop animation.

**Single tap on Text:** Display line translation dialog with pop animation.

**Single tap on Translation Dialog:** Close the translation dialog.

**Double tap:** Display paragraph translation dialog with pop animation.

## 5.2 Home Screen

Home screen UI of BiText reader focus on listing electronic books. A table view object displays ebooks in rows and users can flick or drag to scroll through ebooks. Additionally users can tap on an electronic book to open its content. Once the selected row is selected, it gets highlighted as the ebook content view slides into place.

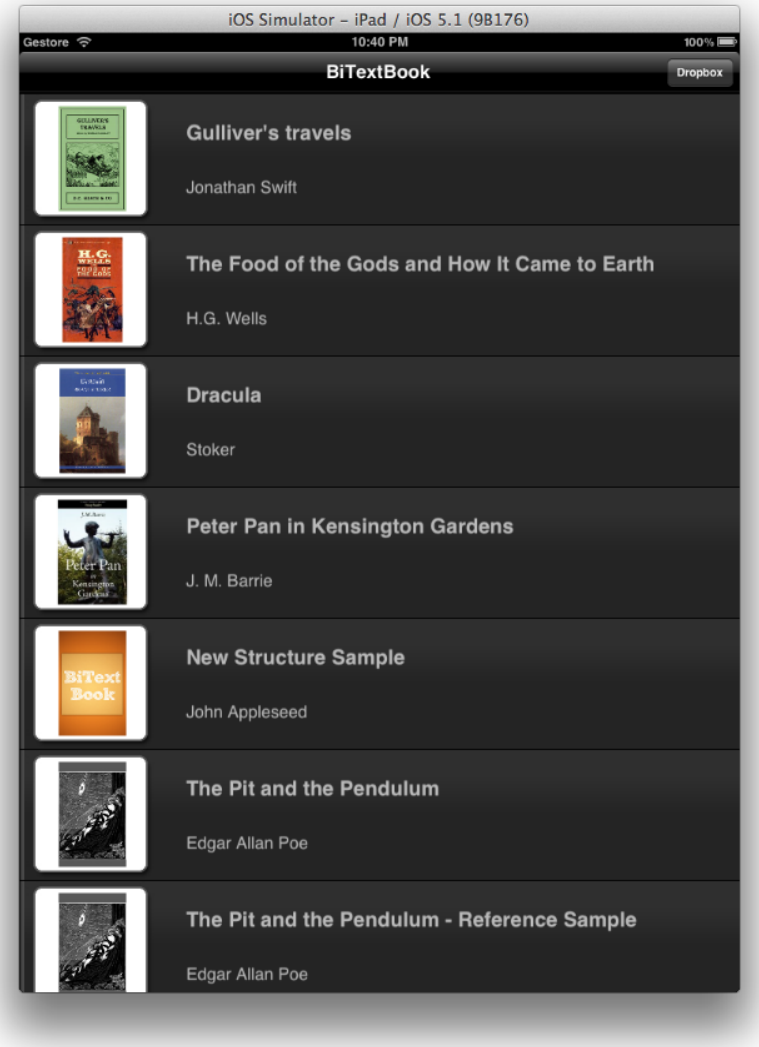



*Figure 5.4: BiText reader ebook listing*

## 5.3 Inline Translation

Single tap on the text is used to display line translation dialog. A popover object is used to display line translation and it hovers above the contents of a screen. Moreover it displays an arrow that indicates the point from which it emerged.

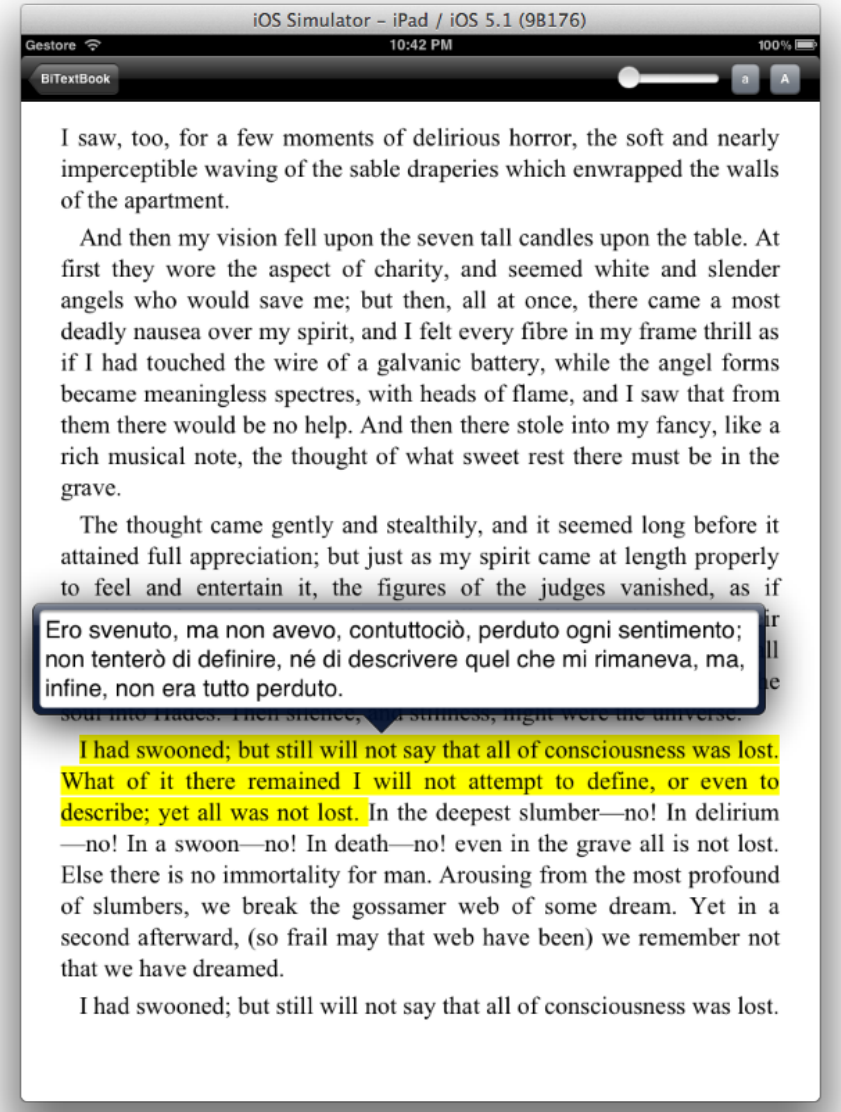



*Figure 5.5: BiText reader inline translation dialog*

## 5.4 Paragraph Translation

When users double tap on a text fragment, the literary translation appears above the original text. Similar to inline translation dialog, a popover object is used to display translation.

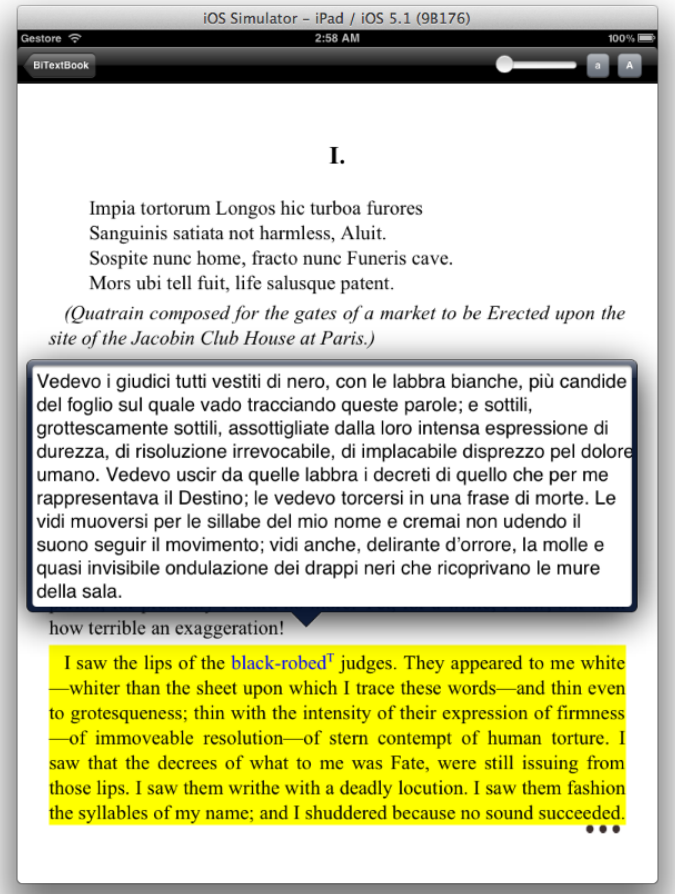



*Figure 5.6: BiText reader paragraph translation dialog, showing the "continuation" three-dot icon*

It is possible to encounter some paragraphs or sentences which starts in the current page and continue in the following page as it is shown in the figure above. In addition to that users may double tap on those large paragraphs to see their translation. In that case a three dot icon is displayed at the end of original text as a feedback that paragraph is partially displayed in the current page.

## 5.5 Footnote Translation

Users can touch on a footnote link and view additional information in the dialog. BiText reader supports two types of footnotes, one is added by the author and the other one is by the translator. Author's note can be idetified by the super letter A at the end of footnote link, on the other hand for the translator's note super letter T is utilized.

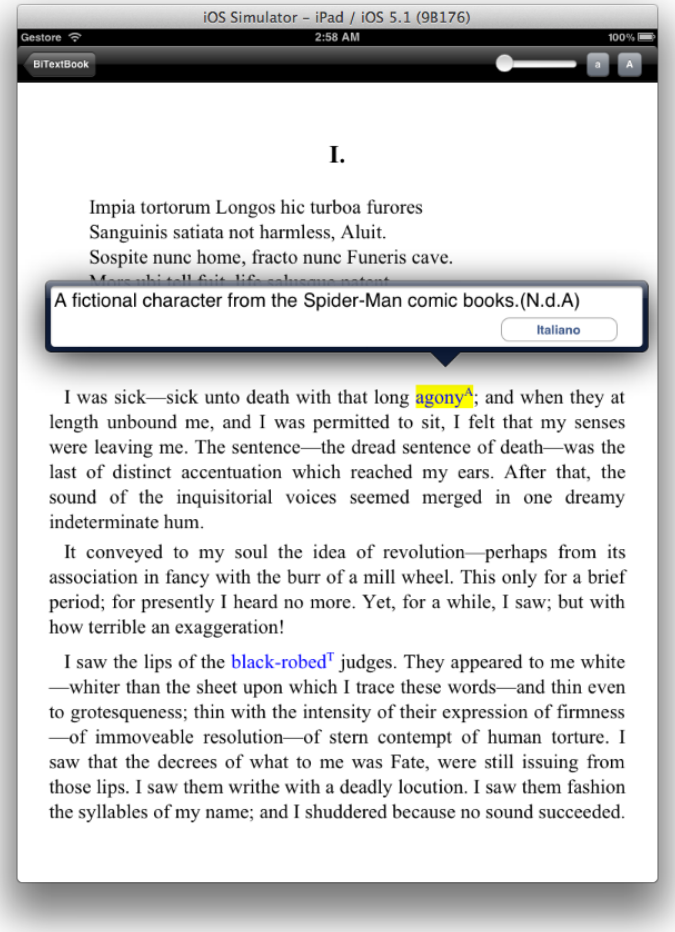



*Figure 5.7: BiText reader EPUB auther footnote dialog*

As shown in the figure 5.7, author's footnotes contain a translation button next to footnote content. Users can tap on this button and view the translation of the footnote. Naturally translator's footnote do not have a translation button since they are only available in translator's language.

## 5.6 Font Size and Brightness Controls

Users can change the font size of the text just right for their eyes. Moreover they can adjust the brightness of the display according to environment light. Those features are controlled by two buttons and a slider which are positioned on the right top side of the screen. Tapping Lower case 'a' button decrease the fontsize on the contrary tapping upper case 'A' button increase. Moreover brightness can be updated by the slider. Users can drag the thumb along the slider, and the display brightness is updated continuously.

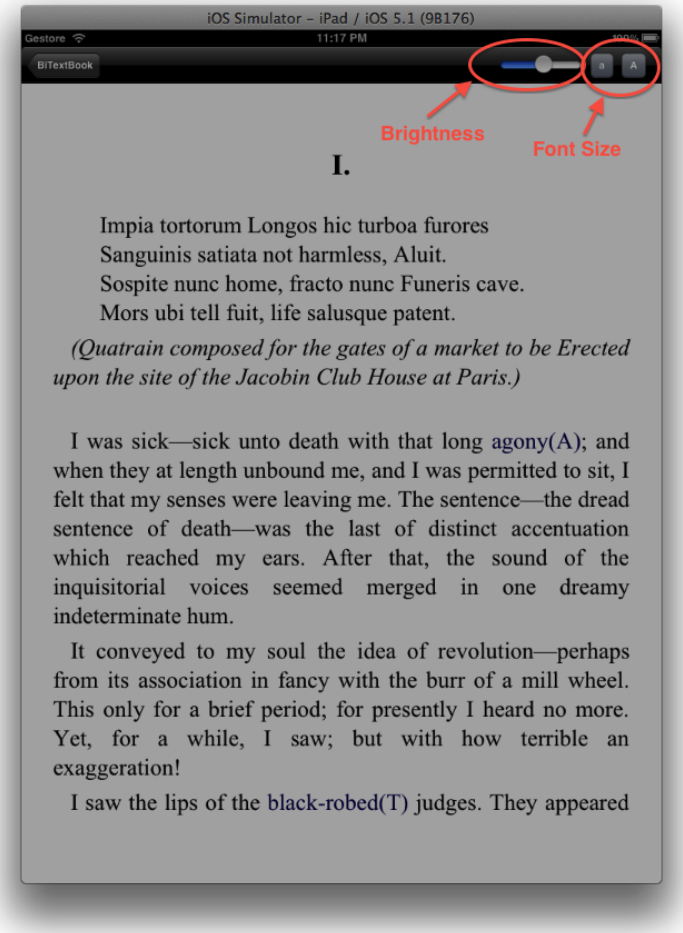



*Figure 5.8: BiText reader controls for adjusting fontsize and brightness*

## 5.7 Importing New Ebook

Importing new ebooks is essential for the users. In order to provide a simple way for importing ebooks, Bitext Reader utilize Dropbox service. Dropbox is a free service that lets people save their documents in the cloud storage. From Bitext reader, users can connect to their dropbox account so that they can browse and download their books. Dropbox login dialog is displayed in the figure 5.9.

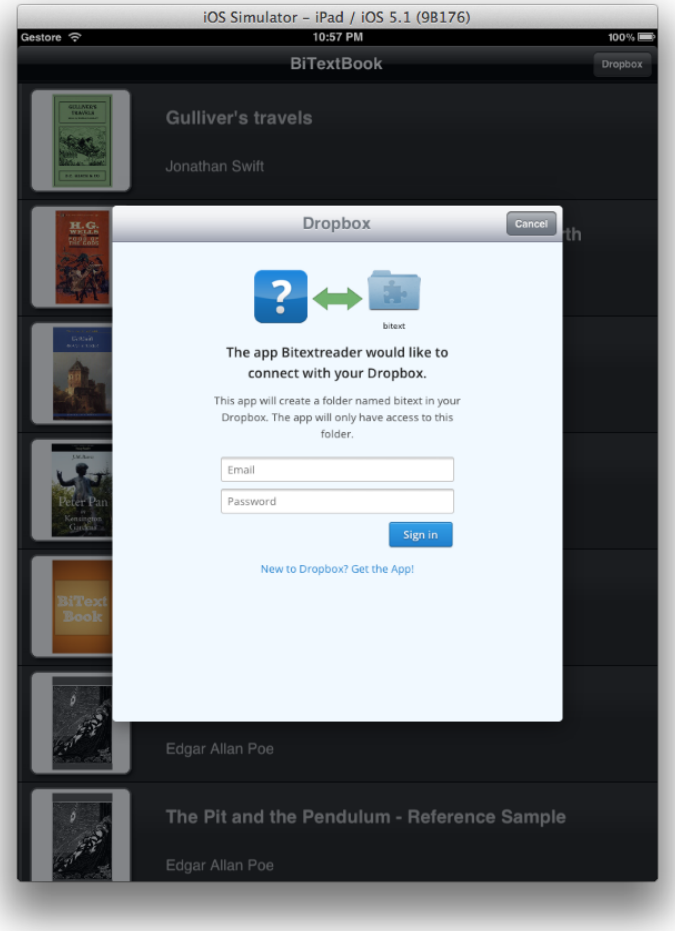



*Figure 5.9: BiText reader import ebook view*

# Chapter 6

# System Architecture and Software Design

*"If a project has not achieved a system architecture, including its rationale, the project should not proceed to full-scale system development. Specifying the architecture as a deliverable enables its use throughout the development and maintenance process."*

Barry Boehm

This chapter introduces transferable abstraction of BiText Reader application. At first, system architecture, conceptual layers and modules are introduced. Class structures and their relationships with other classes are explained with UML diagrams. Following that, software design of BiText Reader is presented and data communication flow is explained. Finally compilation environment and system requirements for further development is stated.

## 6.1 Architecture Overview

BiText Reader application is designed to be flexible and reusable. Segregated layers of BiText Reader application gives the option of modifying or adding a specific layer, instead of reworking the entire application. Architeture of BiText Reader is the derived combination of model view controller and n-tier architecture which are commonly used for web applications.

As it is shown in figure A.1, the high level abstraction of BiText Reader is composed of three layers. Each of the layers and components inside those layers will be explained during this section.

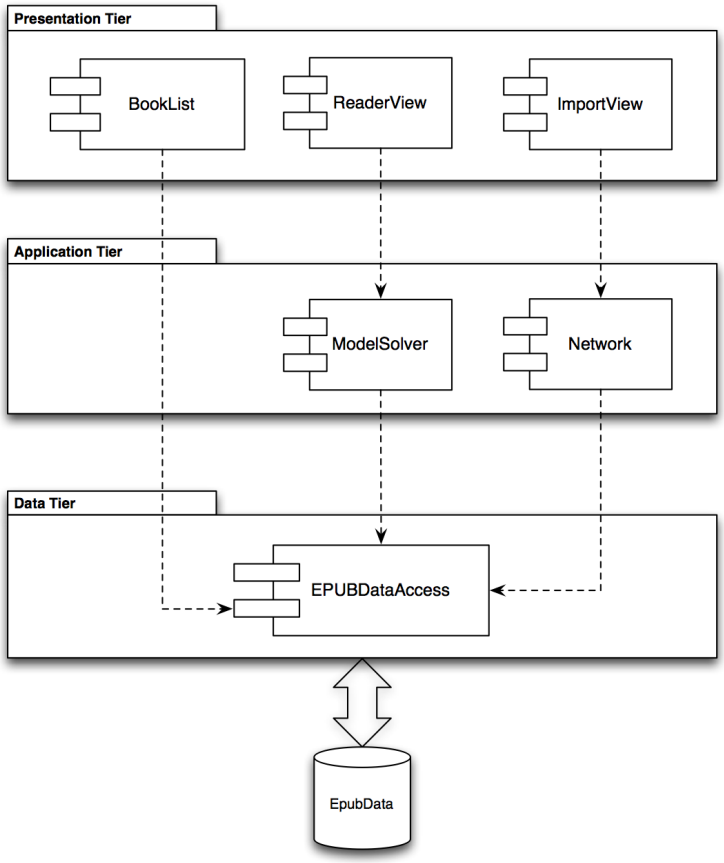



*Figure 6.1: Bitext reader architecture diagram*

## 6.2 Presentation Layer

This is the topmost level of the application. It consists of parts that are used to present data to the end user. Components on the presentation layer includes subclasses of UIView which comes with The UIKit framework.View objects are the main way BiText Reader interacts with the user, they have responsibilities such as drawing and animating, Layout and subview management moreover touch event handling. Presentation layer composed of three main view controllers:

### Book list

Book list controller is responsible for listing ebooks inside the application. It is the first view controller which gets loaded as soon as application is launched.

#### Content reader

Content reader controller is responsible for managing display of EPUB content and managing translation dialog views which are triggered through BiWebview which is a custom UIWebview class used to handle BiText script events.

#### Book import

Book import is completely managed by dropbox framework. Book import view is responsible for listing EPUB files contained in user's dropbox.

## 6.3 Application Layer

Application layer controls BiText Reader's functionality by performing detailed processing. There are two major components inside logic tier: model solver and network.

#### Model solver

Model solver contains modules for opening EPUB file and interpreting its contents and events. In order to achieve this goal it contains three modules and a model. ZipArchive module controls compressing/decompressing functionalities. XML module is used for managing XML queries and in a similar way JSON module is used in serialization and deserialization of JSON data. And Epub Model is reponsible for uncompressing and deserializing EPUB files.

#### Network

Network component holds two main modules, one of them is Dropbox module which is responsible for authorizing dropbox access and listing EPUB files, the other one is Asihttprequest module which is used for downloading EPUB files from dropbox.

## 6.4 Data Access Layer

This layer is repsonsible for storing and retrieving ebook information. This tier keeps data neutral and independent from the other layers. Data is stored in plist format and managed by Epub list controller.

## 6.5 Class Structures and Relationships

BiText Reader application is composed of 68 classes. Following diagram, Figure 6.2 shows the application's major class interfaces, their attributes, methods and their related module. Detailed documentation of the project can be found at the appendix section.

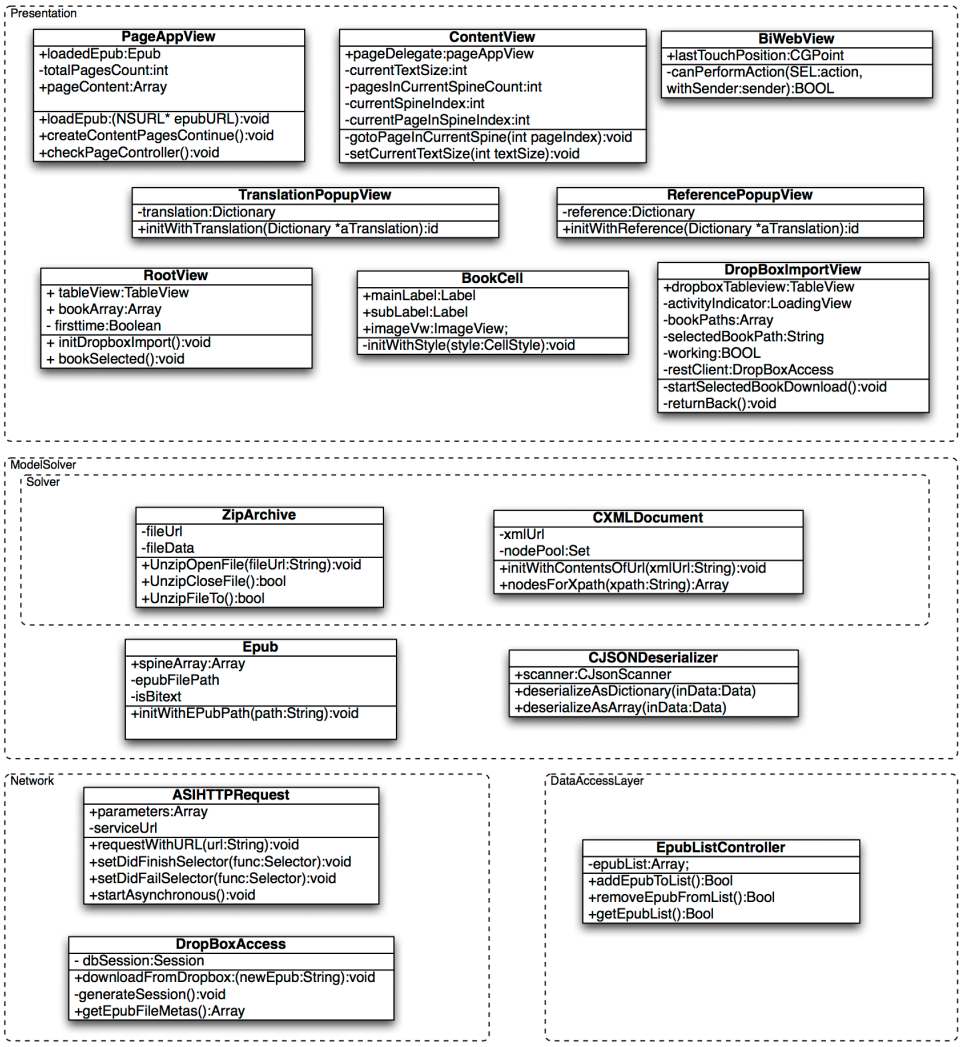



*Figure 6.2: BiText class diagram*

Relationships among classes provide the foundation for the structure of BiText Reader. In order to explore the associations between classes, Class relationship diagram is presented in figure 6.3.

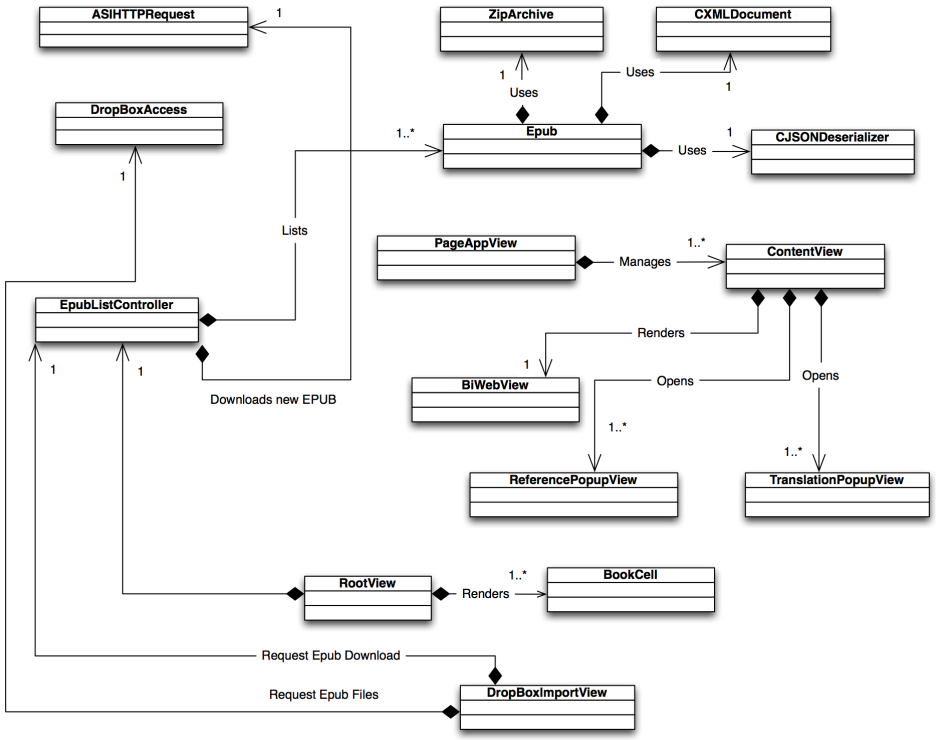



*Figure 6.3: BiText class relationships*

## 6.6 Software Design

BiText Reader is designed for rendering multilingual content. Additionally its goal is to support user interaction with the text fragments and the footnotes so as to display extra information in a second language. Initially a general purpose EPUB reader is designed to open .epub files, generate the index and present the content properly, then a two-way communication bridge is required between EPUB content and BiText Reader therefore user interactions can be managed asynchronously.

### 6.6.1 Communication Bridge

BiText Reader and BiText core functions embedded in the EPUB file are designed to notify each other on the interactive events. In order to achieve

this goal, BiText Reader calls Javascript function from Objective-C and the other way around, BiText Core Javascript functions are able to call Objective-C from Javascript. Figure 6.4 presents the communication bridge between BiText Reader and BiText Core.

The first way of communication, calling Javascript from Objective-C is done with stringByEvaluating Javascript FromString method of UIWebView instance. However calling Objective-C from Javascript method is not simple since Iphone SDK do not offer a native way for this purpose. So a work-around is implemented to solve this problem. UIWebViewDelegate is registered for the UIWebview object and its location change is caught immediately, url parameters are used as a functional reference and change request is canceled.

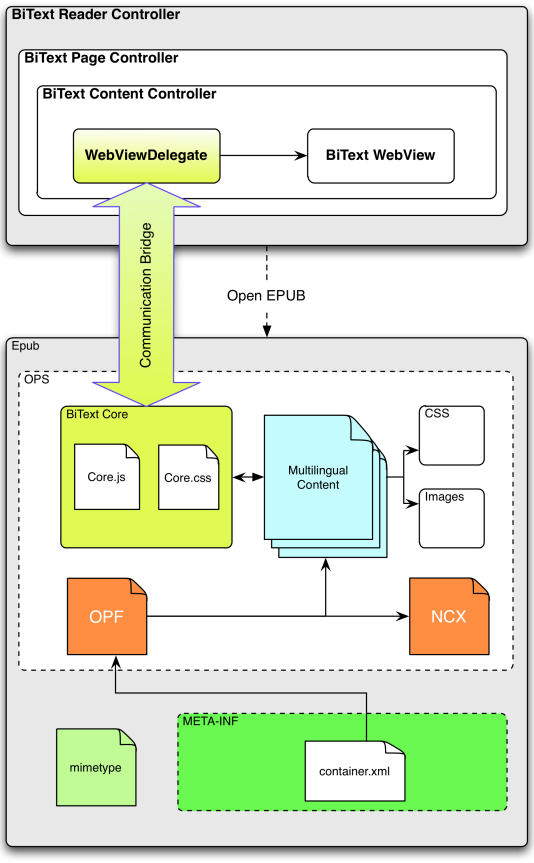



*Figure 6.4: BiText Reader - BiText Core communication bridge*

### 6.6.2 Script Initialization

BiText reader offers many features such as adjusting font sizes, highlighting text, scrolling page content; beside its unique multilingual format support. Thus it is important to centralize notification flow in order to keep all those features smoothly functioning. So as to achieve this goal, content interactions are managed from one place, from BiText Javascript core.

Once EPUB document is opened and ebook content is loaded, it is time to identify the document and start script initialization. In order to initialize the scripts, first EPUB file is checked weather it is supporting multilingual content or not. Second, initBitext method is called so as to activate text touch event listeners that way BiText communication bridge gets activated.

Figure 6.5 shows the flow for identifying whether EPUB file supports multilingual content and how events are managed.

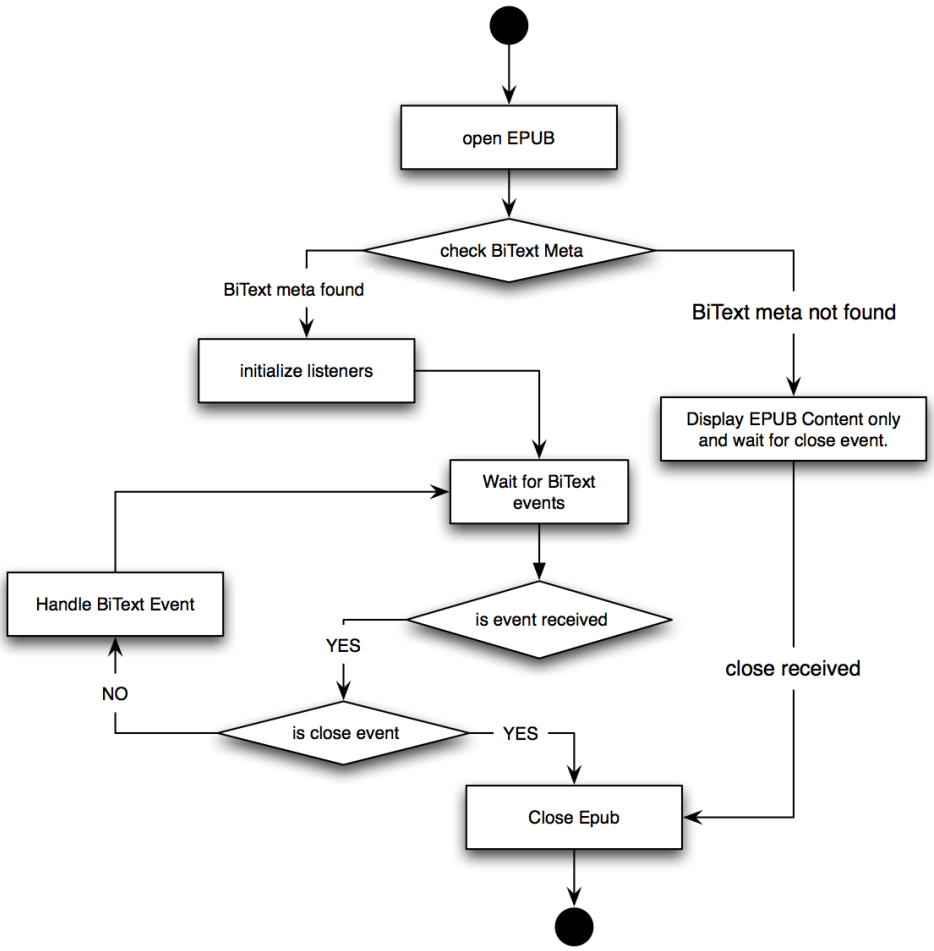



*Figure 6.5: Multilingual EPUB content identification flow*

### 6.6.3 Script Data Flow

Multilingual data transfer is essential for BiText Reader interactions. Data transfer parameters are converted into JSON format and passed to BiText Reader.

Figure 6.6 shows data construction for a sentence. As shown in the figure, first of all, text in second language is retrieved, secondly touched area is highlighted and position of the touch event is calculated then finally variables are gathered in a JSON object and returned.

```
function callBitext(object, type){
  /*...  only line handling displayed  */
        translated = object.find("[lang=it]").text();
        object.addClass("highlight");
        object.html("<div class='placer'>.</div>"+ object.html() );
        placerOffset = $(".placer").offset();
        $(".placer").remove();
        offset = object.offset();
        offsetTop = placerOffset.top;
        res = object;
   /*...  paragraphs are handled in similar manner  */
   result = {line: translated, cx:offset.left, cy:offsetTop, cw:res.width(), ch:res.height() };
   finalResult = $.toJSON( result );

   return finalResult;
}
```

*Figure 6.6: BiText core function for retrieving sentence data in JSON*

Touch event listeners have pivotal role in the BiText reader. They are used for identifying the touched object, retrieving multilingual data and passing parameters to url location so that event notification can be caught by the BiText Reader. Figure 6.7 shows the initBitext method executed in order to initialize touch event listeners.

```
function initBitext()
{
    $("span").doubletap(
        function(event){
            event.stopPropagation();
            result = callBitext($(event.target), 2);
            window.location = "bitextcall:-:jsbitext:-:" + result;
        },
        function(event){
            event.stopPropagation();
            result = callBitext($(event.target), 1);
            window.location = "bitextcall:-:jsbitext:-:" + result;
        },
        400
    );
}
```

*Figure 6.7: BiText core function for event listener initialization.*

The UIWebViewDelegate protocol defines methods that a delegate of a UIWebView object can optionally implement to intervene when web content is loaded [3]. UIWebViewDelegate has several tasks related to loading content and one of them is deciding wether requested url should be loaded or refused.

`webView:shouldStartLoadWithRequest:navigationType:`

Above instance method sent before a web view begins loading a frame. BiText Reader utilizes the method for function requests from Javascript. Following figure 6.8 shows how Javascript methods are extracted in Objective-C.

```
- (BOOL)webView:(UIWebView *)webView2  shouldStartLoadWithRequest:(NSURLRequest *)request  navigati
{
    NSString *requestString = [[request URL] absoluteString];

    if ([requestString hasPrefix:@"bitextcall:-:"]) {

        if( pageDelegate.isMovingToParentViewController || pageDelegate.isMovingFromParentViewContr
        {
            [webView stringByEvaluatingJavaScriptFromString:@"cleanHighlights()"];
            return NO;
        }

        NSArray *components = [requestString componentsSeparatedByString:@":-:"];

        NSString *function = (NSString*)[components objectAtIndex:1];
        NSString *param = [(NSString*)[components objectAtIndex:2]
                           stringByReplacingPercentEscapesUsingEncoding:NSUTF8StringEncoding];
        NSDictionary *paramDict = [[CJSONDeserializer deserializer] deserializeAsDictionary:[param

        if([function isEqualToString:@"jsbitext"])
            [self bitextcall:paramDict];
        else if([function isEqualToString:@"jsbiref"])
            [self birefcall:paramDict];

        return NO;
    }

    return YES;
}
```

*Figure 6.8: BiText reader method for catching Javascript functions*

At first, requested url value is checked if it is contains bitext method tag "bitextcall", if the value is not found url load permission is granted and normal execution continues. Otherwise if bitextcall request is received, url is divided in to fragments by unique parameter tag ":-:" which is reserved for bitext requests. With this break down operation function and parameters are extracted so that corresponding function in Objective-C can be called. At the same time parameters which are sent by Javascript in JSON format are deserialized so as to allow touched data interpretation.

### 6.6.4 Event Priorities

BiText Core events have different degrees of priority in order to prevent false positive touch recognitions. First degree is reserved for touch of footnotes, then second degree is for paragraph double touch events and third degree is for single touch sentence event. That way none of the events are overwhelming others and all events can work simultaneously without interruption.

## 6.7 System Requirements

This project is developed with Xcode which is the official IDE for iOS development. In order to run the project, Mac OS X 10.7+ operating system and Xcode IDE is required. Moreover Apple developer account is obligatory for installing the compiled application to test on iPad device however it is possible to use iPad simulator which comes built-in with Xcode. During the BiText reader development, 1st generation iPad with iOS 5.0 operating system is used for device tests.

# Chapter 7

# Conclusion and Future Work

*"If we all did the things we are capable of, we would astound ourselves."*

Thomas A. Edison

## 7.1 Conclusion

In this thesis, we introduced a multilingual content structure for EPUB formatted ebooks and applied this structure for creating a multilingual ebook of a story by Edgar Allan Poe. As well, an ebook reader iPad application is developed for the purpose of allowing users to access multilingual ebook content with intuitive gestures.

During this study, implementation of the project is covered step by step, from requirement analysis to executable solution. Moreover detailed explanation of the development progress shows that electronic books can be built for different purposes in order to provide a more comprehensible reading experience.

Our study is intended to present that electronic books have an important potential for making the content more accessible by supplying multilingual content. We hope that this study would encourage electronic book publishers to provide content in multiple languages therefore more readers can benefit from them.

## 7.2 Future Work

Multilingual ebook reader which is developed in this study, has high potential for further enhancements. Some of the future works and further features that can be adapted in to the project are listed as below:

**Text to speech:** It is important to make electronic books accessible to many blind and partially sighted people through assistive technologies. Text to speech output for book content helps people with sight disabilities to interpret content by listening. In addition to that, this feature can help foreign language learners to hear the pronunciation of a text. It can also be extended to support text dialogs in story books. Unique voices can be assigned to each character in a story book and dialogs can be generated in runtime. Moreover for dialogs different character voices can be used.

**Swipe dialog:** In the user interaction section, we mentioned that when user touches on a large text fragment such as a paragraph or a sentence which continues in the following page, reader needs to close L2 dialog and swipe to next page and then needs to touch the studied paragraph again. However this interaction could be designed in a more user-friendly way. For that reason swipe dialog is suggested: Swipe dialog should execute in three operations consequently when a swipe gesture is recognized on a translation dialog or on a translation text, the L2 dialog disappears, the page turns, and the L2 dialog reappears. In other words, the user can swipe everywhere in the page, as if the translation dialog was not present. That can help readers to compare original text and translated text easily, without performing many gestures.

**Dictionary:** A dictionary could be implemented so that readers tap-and-hold on a L2 sentence to see its dictionary entry in its original definition in L2. The dialog containing the dictionary entry could contain the usual button to switch to the "translation" in this case, it could show the corresponding entry definition in the L1 dictionary.

**Reader Features:** There are widely used features which are supported by the most of the ebook readers. Some of them are: the possibility to search content, add a bookmark, quick access to table of contents and links to chapters. During the project, we focused on the multilingual dialogs and skiped those handy features due to time constraints. Those features could be supported as a future work.

**Evaluation:** As a final future work, an evaluation could be made in order to measure the effectiveness of BiText reader in the perception of content. In the evaluation, one specific book could be delivered in different formats to group of people. For example: a traditional ebook reader with built in dictionary, a multilingual hardcopy of the book and multilingual structured ebook. After experimental group read the content, a comprehension test could be performed.

# References


[1] Apple Inc. Cocoa Fundamentals Guide: A Bit of History, 2010. Last accessed: 2nd Sep 2012 http://developer.apple.com/library/mac/#documentation/Cocoa/Conceptual/CocoaFundamentals/WhatIsCocoa/WhatIsCocoa.html%23//apple_ref/doc/uid/TP40002974-CH3-SW16.

[2] Apple Inc. The Objective-C Programming Language, 2011. Last accessed: 2nd Sep 2012 http://developer.apple.com/library/mac/#documentation/Cocoa/Conceptual/ObjectiveC/Introduction/introObjectiveC.html.

[3] Apple Inc. UIWebViewDelegate Protocol Reference, 2011. Last accessed: 1st Sep 2012 http://developer.apple.com/library/ios/#documentation/uikit/reference/UIWebViewDelegate_Protocol/Reference/Reference.html.

[4] Claire Cain Miller & Julie Bossman. E-Books Outsell Print Books at Amazon, 2011. Last accessed: 2nd Sep 2012 http://www.nytimes.com/2011/05/20/technology/20amazon.html.

[5] Creative Commons. Public Domain Certification, 2012. Last accessed: 2nd Sep 2012 http://creativecommons.org/licenses/publicdomain.

[6] The World Wide Web Consortium. Cascading Style Sheets, 2012. Last accessed: 4th Sep 2012 http://www.w3.org/Style/CSS/Overview.en.html.

[7] Melissa Dedina. 12 Ways to Improve Your Second Language with Popular Culture, 2011. Last accessed: 5th Sep 2012 http://www.multilingualliving.com/2011/04/08/12-ways-to-improve-your-second-language-with-popular-culture/.

[8] Disney. Disney Publishing Worldwide Launches Digital Language Learning Product, 2012. Last accessed: 2nd Sep 2012

http://d23.disney.go.com/news/2012/03/
disney-publishing-worldwide-launches-digital-language-learning-product/.

[9] Edgar Allan Poe. Il pozzo e il pendolo - racconto, 2012. Last accessed: 11th Sep 2012 http:
//www.edgarallanpoe.it/racconti/il-pozzo-e-il-pendolo-racconto/.

[10] Yousef Elbes. The importance of multilingualism, 2012. Last accessed: 1st Sep 2012 http://multilingualism.org/multilingualism/the-importance-of-multilingualism.

[11] International Digital Publishing Forum. Open Publication Structure (OPS) 2.0.1 v1.0.1, 2010. Last accessed: 4th Sep 2012 http://idpf.org/epub/20/spec/OPS_2.0.1_draft.htm.

[12] International Digital Publishing Forum. Epub specifications, 2012. Last accessed: 2nd Sep 2012 http://idpf.org/epub.

[13] Mozilla Foundation. Javascript, 2012. Last accessed: 4th Sep 2012 https://developer.mozilla.org/en-US/docs/JavaScript.

[14] Mozilla Foundation. JSON, 2012. Last accessed: 4th Sep 2012 https://developer.mozilla.org/en-US/docs/JSON.

[15] Gisela Ernst-Slavit & Margaret Mulhern. Bilingual Books: Promoting Literacy and Biliteracy in the Second-Language and Mainstream Classroom, 2011. Last accessed: 2nd Sep 2012 http://www.readingonline.org/articles/ernst-slavit/.

[16] Apple Inc. Developer Tools, 2012. Last accessed: 3rd Sep 2012 https://developer.apple.com/technologies/tools/.

[17] Learning Ally & Markus Gylling from DAISY Consortium
James Pritchett. EPUB Open Container Format (OCF) 3.0, 2010. Last accessed: 4th Sep 2012 http://idpf.org/epub/30/spec/epub30-ocf.html.

[18] Joseph Palmeri. French Wit and Wisdom: A look at Life by Great French Writers [Paperback], 2011. Last accessed: 2nd Sep 2012 http://www.amazon.com/gp/product/0966184300/.

[19] William F. Mackey. *Bilingualism and multilingualism. In: Ulrich Ammon; Norbert Dittmar; Klaus J. Mattheie. Sociolinguistics: An International Handbook of the Science of Language and Society*, pages 1483–1495. Walter de Gruyter, 2 edition, 2005.

[20] Richard Nordquist. Multilingualism, 2012. Last accessed: 8th Sep 2012 http://grammar.about.com/od/mo/g/Multilingualism.htm.

[21] Steve Paxhia and John Parsons. *Consumer attitudes toward ebook reading*. Book Industry Study Group, Inc, 2012.

[22] Cynthia Klingler Rafael M. Diaz. Towards an explanatory model of the interaction between bilingualism and cognitive development. In Ellen Bialystok, editor, *Language Processing in Bilingual Children*, chapter 8, pages 167–192. Cambridge University Press, 1991.

[23] Dan Saffer. *Designing Gestural Interfaces: Touchscreens and Interactive Devices*. O'Reilly Media, 1 edition, 2008.

[24] Dan Saffer. *Designing Gestural Interfaces: Touchscreens and Interactive Devices*. O'Reilly Media, 1 edition, 2008.

[25] John Tagler Sara Pinto, Kate Kolendo. A periodic alerting service leading you to information relevant to the professional and scholarly publishing industry, 2012. Last accessed: 4th Sep 2012 http://www.publishers.org/psp/psplinks91/.

[26] Jacalyn C. Spoon. Ancient Libraries of Greece and Rome, 1999. Last accessed: 3rd Sep 2012 http: //www.ithaca.edu/history/journal/papers/sp02ancientlibraries.html.

[27] Mikko Välimäki. Open Standards and Patents - A Start-up Perspective, 2010. Last accessed: 4th Sep 2012 http://cyber.law.harvard.edu/events/luncheon/2010/04/mikko.

[28] Wang, S. C., Jackson, F. H., Mana, M., Liau, R., Evans, B. Resource Guide to Developing Linguistic and Cultural Competency in the United States, 2010. Last accessed: 3rd Sep 2012 http://www.nflc.org/publications/.

[29] Jenn Webb. The agile upside of XML, 2011. Last accessed: 4th Sep 2012 http: //toc.oreilly.com/2011/10/xml-publisher-workflow-ebook-design.html.

[30] Victor Webb. *Language in South Africa*. John Benjamins Publishing Co., 2002.

[31] Wixon D. Wigdor D. *Brave NUI World:Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufmann, 1 edition, 2011.

[32] Wikimedia. About wikisource, 2012. Last accessed: 1st Sep 2012 http://wikisource.org/wiki/Wikisource:About_Wikisource.

# Appendix A

# Documentation

Documentation of BiText reader is published on the web via following link: http://polimi.guvenergokce.com/bitext/doc.pdf

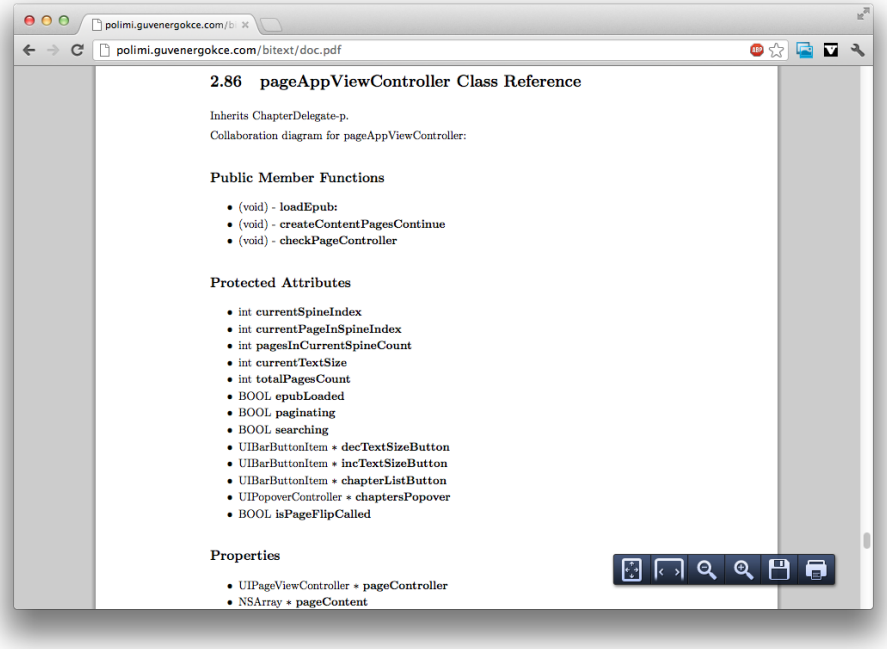



*Figure A.1: Bitext reader documentation*

News and further developments of the project can be followed through the project site: http://polimi.guvenergokce.com/bitext/