



Faculty of Computer Science and Information Technology

***Building Kenyah Badeng Corpora: An under-resourced
language in Sarawak***

HELDI BUNGA ANCHAU

**Bachelor of Computer Science with Honours
(Information System)**

2020

**BUILDING KENYAH BADENG CORPORA: AN UNDER-RESOURCED
LANGUAGE IN SARAWAK**

HELDI BUNGA ANCHAU

This project is submitted in partial fulfilment of the
requirements for the degree of
Bachelor of Computer Science and Information Technology
(Information System)

Faculty of Computer Science and Information Technology
UNIVERSITI MALAYSIA SARAWAK

2020

**MEMBINA KENYAH BADENG KORPORA: BAHASA YANG TIDAK
DIBEKALKAN DI SARAWAK**

HELDI BUNGA ANCHAU

Projek ini merupakan salah satu keperluan untuk
Ijazah Sarjana Muda Sains Komputer dan Teknologi Maklumat
(Sistem Maklumat)

Fakulti Sains Komputer dan Teknologi Maklumat
UNIVERSITI MALAYSIA SARAWAK

2020

UNIVERSITI MALAYSIA SARAWAK

THESIS STATUS ENDORSEMENT FORM

TITLE BUILDING KENYAH BADENG CORPORA: AN UNDER-RESOURCED
LANGUAGE IN SARAWAK

ACADEMIC SESSION: 2019/2020

HELDI BUNGA ANCHAU
(CAPITAL LETTERS)

hereby agree that this Thesis* shall be kept at the Centre for Academic Information Services, Universiti Malaysia Sarawak, subject to the following terms and conditions:

1. The Thesis is solely owned by Universiti Malaysia Sarawak
2. The Centre for Academic Information Services is given full rights to produce copies for educational purposes only
3. The Centre for Academic Information Services is given full rights to do digitization in order to develop local content database
4. The Centre for Academic Information Services is given full rights to produce copies of this Thesis as part of its exchange item program between Higher Learning Institutions [or for the purpose of interlibrary loan between HLI]
5. ** Please tick (✓)

- | | | |
|-------------------------------------|--------------|--|
| <input type="checkbox"/> | CONFIDENTIAL | (Contains classified information bounded by the OFFICIAL SECRETS ACT 1972) |
| <input type="checkbox"/> | RESTRICTED | (Contains restricted information as dictated by the body or organization where the research was conducted) |
| <input checked="" type="checkbox"/> | UNRESTRICTED | |



(AUTHOR'S SIGNATURE)

Permanent Address

No. 2 Blok M, Uma Badeng,
96900 Belaga, Sarawak.

Date: 12 August 2020

Validated by


Dr. S. S. S. S. S.
Lecturer (Automation Systems Programme)
Faculty of Computer Science and Information Technology
UNIVERSITI MALAYSIA SARAWAK

(SUPERVISOR'S SIGNATURE)

Date: 13 August 2020

Note * Thesis refers to PhD, Master, and Bachelor Degree

** For Confidential or Restricted materials, please attach relevant documents from relevant organizations / authorities

DECLARATION

I hereby declare that this project is my original work. I have not copied from any other student's work or any other sources except where due to reference or acknowledgment is not made explicitly in the text, nor has any part had been written for me by another person.



(HELDI BUNGA ANCHAU)

(12 August 2020)

ACKNOWLEDGEMENT

Firstly, I would like to state my sincere gratitude to my supervisor, Dr Suhaila binti Sae for giving me an opportunity to be under her supervision for my Final Year Project, assistance in stimulating suggestion and continuously guide me to complete my Final Year Project. Thank you for always give me guide and support, my progress for Final Year Project will not run smoothly without her.

Besides, I would like to thank you to my university which is University Malaysia Sarawak (UNIMAS) for give me a chance to gain my knowledge in UNIMAS. I also would like to express my thank to my faculty which is Faculty of Computer Science and Information Technology (FCSIT) for giving me a challenging experience to complete this project. This challenging experience has given me a lesson which is keep fighting until the end.

Not to forget, my thanks and appreciations go to my family who always supporting and understand me during the period of Final Year Project. Lastly, thank you to my colleague for helping me to complete the project.

ABSTRACT

Kenyah Badeng languages is a language that spoken by Kenyah community. Kenyah community are ethnic groups in Sarawak, Malaysia and Kalimantan, Indonesia. Most of Kenyah Badeng community are located at Belaga, Sarawak. The population of Kenyah Badeng community are still small. Nowadays, the language is less practiced by the people themselves. This is because the community less speak their language with the new generations and Kenyah Badeng also less of resourced in library and social media platforms. In other case, mixed marriage which is leads to less spoken of Kenyah Badeng language because the couples usually use other language instead of use Kenyah Badeng language to communicate. Hence, to prevent Kenyah Badeng language being extinct, building Kenyah Badeng corpora is needed.

Building Kenyah Badeng corpora is a way to enrich knowledge for Kenyah Badeng language and preserve the Kenyah Badeng language. This is because Kenyah Badeng language is one of an under-resourced language in Sarawak. Under-resourced language means the language lack of a writing system, limited presence on the web and lack of electronic resources for speech. Through the building Kenyah Badeng corpora, Kenyah Badeng language can be preserve and enrich the knowledge about Kenyah Badeng language. This is to make the corpora of Kenyah Badeng more interactive and avoid the corpora of Kenyah Badeng from extinction. Kenyah Badeng language resource must be saved into a digital format which can be store for a long decade. The corpora can be display in digital library because digital library is one of the methods to protect language resources in the type of text, audio, video and so on. In the digital library for Kenyah Badeng, there will display the corpus of Kenyah Badeng whereby the corpus of Kenyah Badeng will be taken from available resourced which is from books and recording video of Kenyah Badeng community. Therefore, Kenyah Badeng language can be known by other community not only the Kenyah community.

ABSTRAK

Bahasa Kenyah Badeng adalah bahasa yang dituturkan oleh masyarakat Kenyah. Masyarakat Kenyah adalah kumpulan etnik di Sarawak, Malaysia dan Kalimantan, Indonesia. Sebilangan besar komuniti Kenyah Badeng terletak di Belaga, Sarawak. Penduduk masyarakat Kenyah Badeng masih sedikit. Pada masa kini, bahasa tersebut kurang diamalkan oleh masyarakat itu sendiri. Ini kerana masyarakat kurang bercakap bahasa mereka dengan generasi baru dan Kenyah Badeng juga kurang sumber dalam perpustakaan dan media sosial. Dalam kes lain, perkahwinan campur yang menyebabkan kurang percakapan bahasa Kenyah Badeng kerana pasangan biasanya menggunakan bahasa lain dan bukannya menggunakan bahasa Kenyah Badeng untuk berkomunikasi. Oleh itu, untuk mengelakkan bahasa Kenyah Badeng pupus, pembinaan korpora Kenyah Badeng sangat diperlukan.

Membina Kenyah Badeng korpora adalah cara untuk meningkatkan pengetahuan terhadap bahasa Kenyah Badeng dan memelihara bahasa Kenyah Badeng. Ini kerana bahasa Kenyah Badeng adalah salah satu bahasa yang kurang sumber di Sarawak. Bahasa yang kekurangan sumber bermaksud kekurangan sistem penulisan bahasa, kehadiran yang terhad di laman sesawang dan kekurangan sumber elektronik untuk pertuturan. Dengan pembinaan Kenyah Badeng korpora, bahasa Kenyah Badeng dapat dipelihara dan mempertingkatkan pengetahuan tentang bahasa Kenyah Badeng. Ini kerana untuk menjadikan korpora Kenyah Badeng lebih interaktif dan mengelakkan korpora Kenyah Badeng daripada pupus. Sumber bahasa Kenyah Badeng mesti disimpan ke dalam format digital yang boleh disimpan selama satu dekad yang panjang. Korpora boleh dipaparkan di perpustakaan digital kerana perpustakaan digital adalah salah satu kaedah untuk melindungi sumber bahasa dalam jenis teks, audio, video dan sebagainya. Di perpustakaan digital untuk Kenyah Badeng, akan dipaparkan korpus Kenyah Badeng di mana korpus Kenyah Badeng akan diambil dari sumber yang ada iaitu dari buku dan rakaman video masyarakat Kenyah Badeng. Oleh itu, bahasa Kenyah Badeng dapat dikenali oleh masyarakat lain bukan sahaja masyarakat Kenyah.

TABLE OF CONTENTS

DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
ABSTRAK	iv
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER 1: INTRODUCTION	1
1.1 Introduction.....	1
1.2 Problem Statements.....	2
1.3 Objectives.....	2
1.4 Scope.....	3
1.5 Brief Methodology.....	3
1.6 Significance of Project.....	6
1.7 Project Schedule.....	6
1.8 Expected Outcome.....	6
1.9 Project Outline.....	6
CHAPTER 2: LITERATURE REVIEW	8
2.1 Introduction.....	8
2.2 Review of existing works.....	8
2.2.1 The Royal Society Corpus: From Uncharted Data to Corpus.....	9
2.2.2 Building and evaluating the Romanian Medical Corpus.....	10
2.2.3 Building Corpora for Philosophers.....	12

2.2.4 Building a comprehensive syntactic and semantic corpus of Chinese clinical Texts.....	13
2.2.5 Corpus building for Mongolian language.....	16
2.2.6 Compilation of an Arabic Children’s Corpus.....	17
2.2.7 Morphological System for Under-Resourced Languages Using Hybrid Approach.....	17
2.3 Comparison of methods between existing corpus.....	18
2.4 The direction of proposed system.....	20
CHAPTER 3: METHODOLOGY.....	22
3.1 Introduction.....	22
3.2 Methodology.....	22
3.2.1 Workflow to building corpus of Kenyah Badeng language.....	23
3.2.1.1 Questionnaires Analysis.....	30
3.2.1.2 Questionnaires Summary.....	40
3.2.2 Project design.....	41
3.2.2.1 Context diagram.....	41
3.2.2.2 Level 0 Diagram.....	42
3.2.2.3 Level 1 Diagram.....	43
3.3 Database Design.....	47
3.4 Interface Design.....	50
CHAPTER 4: IMPLEMENTATION.....	56
4.1 Introduction	56
4.2 Installation and Configuration of System’s Components.....	56
4.2.1 XAMPP.....	56
4.2.2 PhpMyAdmin.....	58

4.2.3 Sublime Text 3.....	59
4.2.4 Express Scribe Transcription Software.....	59
4.2.5 CamScanner.....	60
4.3 Transcription of Video Recording.....	61
4.4 Digitization of the Kenyah Badeng Books.....	64
4.5 System Module.....	68
4.5.1 Login page for admin.....	68
4.5.2 Homepage for admin.....	68
4.5.3 About page for admin.....	69
4.5.4 Wordlist for admin.....	69
4.5.5 Book List for admin.....	70
4.5.6 Video List for admin.....	71
4.5.7 Homepage for user.....	71
4.6 Summary.....	75
CHAPTER 5: TESTING.....	76
5.1 Introduction.....	76
5.2 Functional Testing.....	76
5.2.1 Test Cases.....	76
5.2.1.1 Login Test Case.....	77
5.2.1.2 Add, Edit, Delete About Test Case.....	78
5.2.1.3 Add, Edit, Delete Wordlist Test Case.....	79
5.2.1.4 Add, Edit, Delete Video Test Case.....	80
5.2.1.5 Add, Edit, Delete Book Test Case.....	81
5.2.1.6 View About, Wordlist, Video, and Book Test Case.....	82

5.3 Non-functional Testing.....	84
5.3.1 Usability and reliability testing.....	84
5.4 User Acceptance Testing.....	87
5.4.1 User Acceptance Analysis.....	88
5.5 Summary.....	93
CHAPTER 6: CONCLUSION AND FUTURE WORK.....	94
6.1 Introduction.....	94
6.2 Objective Achievement.....	94
6.3 Limitation.....	95
6.4 Future Works.....	95
6.5 Conclusion.....	95
REFERENCES.....	96
APPENDICES.....	98
APPENDIX A: PROJECT SCHEDULE	98
APPENDIX B: QUESTIONNAIRE FORM	100

LIST OF TABLES

Table 2.1: Comparison of methods between six articles.....	19
Table 3.1: Demographic profile of the survey.....	26
Table 3.2: Example of tokenize.....	29
Table 3.3: Data dictionary for user.....	48
Table 3.4: Data dictionary for corpus.....	48
Table 3.5: Data dictionary for video.....	48
Table 3.6: Data dictionary for book.....	49
Table 3.7: Data dictionary for about.....	49
Table 4.1: Total of post editing.....	67
Table 4.2: Word Error Rate.....	67
Table 4.3: Example of normalization.....	67
Table 5.1: Login Test Case.....	77
Table 5.2: Add, Edit, Delete About Test Case.....	78
Table 5.3: Add, Edit, Delete Wordlist Test Case.....	79
Table 5.4: Add, Edit, Delete Video Test Case.....	80
Table 5.5: Add, Edit, Delete Book Test Case.....	81
Table 5.6: View About, Wordlist, Video, and Book Test Case.....	82
Table 5.7: Test case for usability.....	85
Table 5.8: Test case for reliability.....	86
Table 5.9: Summary of the acceptance testing.....	87
Table 6.1: Objectives and Achievements.....	94

LIST OF FIGURES

Figure 1.1: An architecture of data collection for Kenyah Badeng	3
Figure 1.2: The process of collecting speech data.....	4
Figure 1.3: The process of collecting text data.....	4
Figure 1.4: The process of pre-processing the text data.....	5
Figure 2.1: Corpus-building steps; interaction with annotation and analysis.....	9
Figure 2.2: General statistics over the corpus.....	11
Figure 2.3: Main interface page.....	12
Figure 2.4: Original page image and corrected text output.....	13
Figure 2.5: Semi-structured sections in Chinese discharge summaries and progress notes....	14
Figure 2.6: An example of the annotation in a sentence from a progress note.....	15
Figure 2.7: Iterative annotation method for guideline and corpus construction.....	15
Figure 2.8: Current and future states of building a Mongolian corpus.....	16
Figure 2.9: Schema of building a Mongolian corpus.....	16
Figure 3.1 An architecture of methods.....	22
Figure 3.2 Workflow of the project.....	23
Figure 3.3 Text and speech data collection steps.....	24
Figure 3.4 Example of stemming.....	28
Figure 3.5 Example of lemmatization.....	28
Figure 3.6 Number of genders that give responses.....	30
Figure 3.7 Age of the respondents.....	31
Figure 3.8 Ethnic of the respondents.....	31
Figure 3.9 Marital status of the respondents.....	32

Figure 3.10 Number of the respondents that from mixed family.....	32
Figure 3.11 Native language for respondents.....	33
Figure 3.12 Number of respondents that know Kenyah Badeng language.....	33
Figure 3.13 Responses on how the respondents know about Kenyah Badeng language.....	34
Figure 3.14 Number of respondents that know to speak in Kenyah Badeng language.....	34
Figure 3.15 Number of respondents that use Kenyah Badeng as the primary language.....	35
Figure 3.16 Number of respondents that want to learn about Kenyah Badeng language.....	35
Figure 3.17 Opinion of respondents that Kenyah Badeng language should be save.....	36
Figure 3.18 Number of respondents that agree to preserve Kenyah Badeng language.....	37
Figure 3.19 Strategies that respondents recommend to preserving Kenyah Badeng language.....	37
Figure 3.20 Number of respondents that have heard about digital library.....	38
Figure 3.21 The expectation of respondents to the digital library for Kenyah Badeng.....	39
Figure 3.22 Number of respondents that agree to have Kenyah Badeng language in digital Library.....	39
Figure 3.23 Context diagram Digital Library for Kenyah Badeng language.....	41
Figure 3.24 DFD Level 0.....	42
Figure 3.25 DFD Level 1 for Process 1.0.....	43
Figure 3.26 DFD Level 1 for Process 2.0.....	44
Figure 3.27 DFD Level 1 Process 3.0.....	44
Figure 3.28 DFD Level 1 Process 4.0.....	45
Figure 3.29 DFD Level 1 Process 5.0.....	45
Figure 3.30 DFD Level 1 Process 6.0.....	46
Figure 3.31 Entity Relationship Diagram for the proposed project.....	47

Figure 3.32 Homepage Digital Library for Kenyah Badeng language.....	50
Figure 3.33 About page.....	51
Figure 3.34 Folk Stories main page.....	52
Figure 3.35 Folk story page.....	53
Figure 3.36 Wordlist page.....	54
Figure 3.37 Corpus list page.....	55
Figure 4.1: Official website to download XAMPP.....	57
Figure 4.2: XAMPP Control Panel.....	57
Figure 4.3: Start Apache and MySQL in XAMPP.....	57
Figure 4.4: Homepage of XAMPP.....	58
Figure 4.5: Homepage of PhpMyAdmin.....	58
Figure 4.6: Sublime Text 3.....	59
Figure 4.7: Homepage of Express Scribe Transcription Software.....	60
Figure 4.8: Logo of CamScanner.....	60
Figure 4.9: Step 1 of transcription.....	61
Figure 4.10: Step 2 of transcription.....	62
Figure 4.11: Step 3 of transcription.....	62
Figure 4.12: Step 4 of transcription.....	63
Figure 4.13: Step 5 of transcription.....	63
Figure 4.14: Step 1 of digitization.....	64
Figure 4.15: Step 2 of digitization.....	65
Figure 4.16: Step 3 of digitization.....	66
Figure 4.18: Login page for admin.....	68
Figure 4.19: Homepage for admin.....	68

Figure 4.20: About page for admin.....	69
Figure 4.21: First page of the wordlist page for admin.....	69
Figure 4.22: Wordlist page of alphabet A for admin.....	70
Figure 4.23: Book list page for admin.....	70
Figure 4.24: Video list for admin.....	71
Figure 4.25: Homepage for public user.....	71
Figure 4.26: About page for public user.....	72
Figure 4.27: First page of the glossary page for public user.....	72
Figure 4.28: Example of alphabet A.....	73
Figure 4.29: Book page for public user.....	73
Figure 4.30: The display of book for public user.....	74
Figure 4.31: Video page for public user.....	74
Figure 4.32: Example of folk story.....	75
Figure 5.1: Testing the user friendly for the system.....	88
Figure 5.2: Testing for interface design.....	88
Figure 5.3: Testing for the wording design.....	89
Figure 5.4: Testing for clearly view the pictures.....	89
Figure 5.5: Testing for clearly watch the video.....	90
Figure 5.6: Testing for the response time.....	90
Figure 5.7: Testing for able to find the information.....	91
Figure 5.8: Testing for information was clear.....	91
Figure 5.9: Testing for the benefit of the information.....	92
Figure 5.10: Testing for all the functions and capabilities.....	92
Figure 5.11: Testing for the satisfied with the system.....	93

CHAPTER 1: INTRODUCTION

1.1 Introduction

Kenyah is the name given to one of many ethnic groups whereby this ethnics groups known as Orang Ulu. Kenyah community is mostly in Sarawak, Malaysia and Kalimantan, Indonesia (Ethnologue, 2019). The language they usually used is an Austronesian language which is Kenyah language. Kenyah community is divided into various dialects such as Badeng, Lepo' Tau, Uma' Jalan, Bakung, Lepo' Tepu, etc. The proposed project will more focus on Kenyah Badeng language. Nowadays, the language is less practiced by the people themselves. This has led to the loss of Kenyah Badeng language and not recognized by other communities.

From Natural Language Processing (NLP) perspective (the technology that deal with human language), Kenyah Badeng is considered as an under-resourced language. Under-resourced language means the language lack of a writing system, limited presence on the web and lack of electronic resources for speech (Besacier, 2014). Although, Kenyah Badeng have available resource in books. For example, “Layan Pengudip Kenyah Badeng”, “Adat Pengelan Kenyah Badeng” and “Asat Buau Kenyah Badeng”. However, the books still not in a digital content.

NLP is commonly used by people. NLP usually used in applications such as language translation, word processor and personal assistant application. Each NLP application requires specific language resources. NLP provides different methods for creating, recording, processing and reusing language resources. Language resources refer to a set of speech or language data (Elra, 2015). Examples of language resources are written and spoken corpora, speech collection, thesaurus, dictionary, lexicon, etc. In the case of Kenyah Badeng, there is a need to preserve the language by building the corpus. Types of corpus to be collected are speech and text data. The corpus to be used in NLP application.

1.2 Problem Statements

Nowadays, Kenyah Badeng language is less spoken by the new generation of the people themselves. In order to survive, the people have to find a job in a place of employment. Therefore, the people must move to the city. Then, these people used to speak a language familiar with and easy to understand. Mixed marriage also gives a huge effect of less spoken Kenyah Badeng language. In mixed married, the couple usually uses language that prefer by the family instead of using Kenyah Badeng language. Therefore, their generations do not know how to speak a Kenyah Badeng language. The possible solution to overcome the problem is by documenting the language in a systematic way through the corpus building. The documentation then would be beneficial to the younger generation to educate more about Kenyah Badeng.

Besides that, Kenyah people are lacking of knowledge about Kenyah Badeng folk stories. This happens because lack of resources on Kenyah Badeng folk stories available on the web and social media. There are still more Kenyah people do not know about Kenyah Badeng folk stories. A possible solution for this problem is to educate the people about Kenyah Badeng folk stories by using the digital library. The digital library will give benefit as it can be easily accessed and the people can gain more about Kenyah Badeng folk stories.

1.3 Objectives

There are three objectives of this research are as follows:

1. to build Kenyah Badeng language resource in specific corpora which, consists of speech and written text corpus
2. to transcribe the speech corpus by using an open source transcriber tool
3. to represent the corpora in a digital library based on Kenyah Badeng folk stories

1.4 Scope

There are three scope of this project:

i. Kenyah Badeng data

The data will be collected from Kenyah Badeng data. There will be two methods to collect the data by interviewing the Kenyah Badeng community and from available resources. Available resources for Kenyah Badeng currently are in books. For example, “Layan Pengudip Kenyah Badeng”, “Adat Pengelan Kenyah Badeng” and “Asat Buau Kenyah Badeng”. The books are telling about religion, migration and lifestyle of Kenyah Badeng. The books also written in Kenyah Badeng language.

ii. Domain expert

The domain expert will be spoken about the folk stories. The folk stories are stories in the oral tradition where the folk stories will be record by recording video. The folk stories are related to culture, history and enemy.

iii. Language resource

Types of language resource that will be used is building the corpus. The corpus will be collected into two types which is speech and text data. The speech will be collected from folk stories and text data will be collected from available resources.

1.5 Brief Methodology

a) Data Collection

Data collection is the process of gathering information. Data collection is important as it helps to identify the issues.

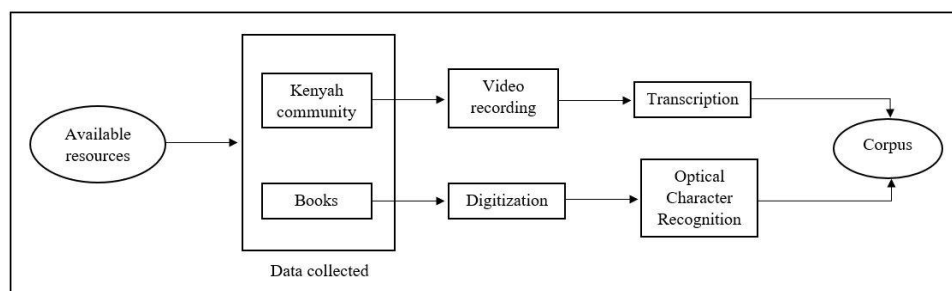


Figure 1.1: An architecture of data collection for Kenyah Badeng

In data collection, there are two different ways to collect the corpus which is speech and text data. The data will be collected from available resources whereby the available resources are in books and Kenyah community. The video recording will be obtained from Kenyah community. Then, transcribe the video into corpus by using transcriber tool. For the books, the written text is from Kenyah books. The books will be digitize and convert it into OCR. Then, the OCR process will get the corpus. The following below are the types of collecting the corpus:

i. Speech data

Speech data is one of the types of corpus that will be collected. The figure show below are the process of speech:

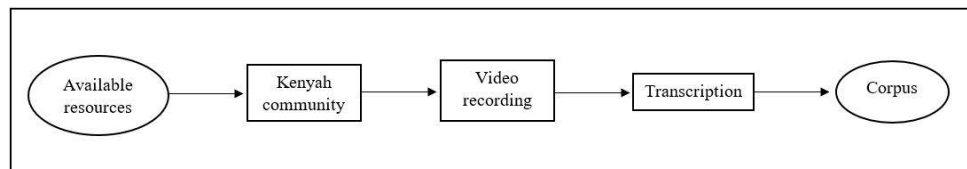


Figure 1.2: The process of collecting speech data

- Video Recording

The video recording will be obtained from Kenyah community. Video recordings is about folk stories. The video will be in Kenyah Badeng language. Then, transcribe the video into corpus by using transcriber tool. The purpose of the transcribe is to get the text.

ii. Text data

Text data is the types of corpus that will be collected. The figure show below are the process of text data:

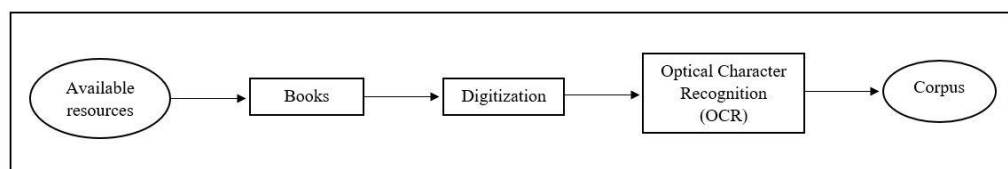


Figure 1.3: The process of collecting text data

- Scan document

The written text is from Kenyah books. The books will be digitize by scanning and then convert it into OCR where by OCR is a technology that recognizes text within a digital image. Then, the OCR process will get the corpus.

b) Pre-processing

Pre-processing is a technique that used to convert the raw data into a clean data. Pre-processing that includes is normalize and tokenize. The figure show below are the process of pre-processing:

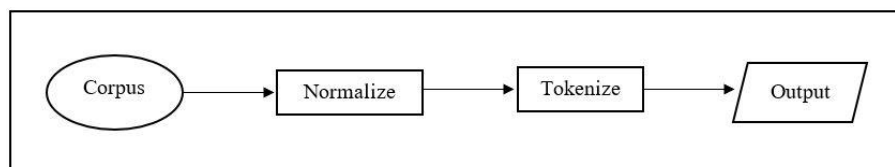


Figure 1.4: The process of pre-processing the text data

i. Normalize

Normalize the corpus. Normalize is a technique of organizing the data. The task of normalize is to handle a range of text issues. The purpose of normalize is eliminating redundant and useless data. Then, proceed to tokenize after the process of normalize.

ii. Tokenize

Tokenize the normalize corpus by using tokeniser. Tokenize is a process of creating token and replacing sensitive data into non sensitive data. Tokenize is a secure method of protecting data.

c) Output

The output will be represented the corpora in a digital library based on Kenyah Badeng folk stories.

1.6 Significance of Project

The project will give a positive impact to the people of ethnicity and the new generation. This project also can introduce Kenyah Badeng language to other community by using the corpus that has been collected. This because the corpus will be represented in a digital library. The digital library for Kenyah Badeng language will be published in social media platforms, to introduce Kenyah Badeng language to another community. Then, other community can learn about Kenyah Badeng language through the digital library. Therefore, the project will preserve the Kenyah Badeng language.

1.7 Project Schedule

Refer to Appendix A

1.8 Expected Outcome

The expected outcome of the project is a Kenyah Badeng language resource and will be represented in a digital library based on Kenyah Badeng folk stories. By using the digital library, this will achieve the aim of the project whereby to preserve the Kenyah Badeng language.

1.9 Project Outline

Chapter 1: Introduction

In chapter 1, there an introduction to the project, which includes the problem statement, objectives, scope, brief methodology of the project, the significant of the project, expected outcome and project outline.

Chapter 2: Literature Review

In chapter 2, there a literature review whereby summary, classification and comparison the previous research.

Chapter 3: Methodology

In chapter 3, explain the techniques that are used in research process to collect and process the data.

Chapter 4: Implementation

Chapter 4 will explain the implementation of the research.

Chapter 5: Conclusion and Future work

This chapter is the summarization of the research and make a suggestion for the future work.