Final Report on

Bangla Spell Checker

Submitted to,

SPL I Evaluation Committee 2020-2021

Bachelor of Science in Software Engineering

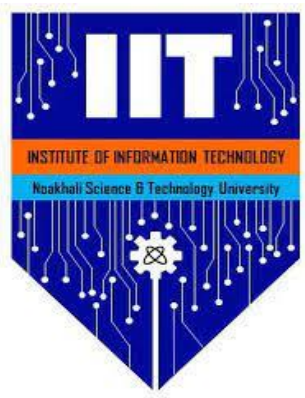
Institute of Information Technology

Noakhali Science and Technology University

Submitted By,

Sanjida Akter Samanta

samanta2515@student.nstu.edu.bd



Final Report on <***Bangla Spell Checker***>

Sanjida Akter Samanta

BFH2025010F

Session:2019-20

Year:2nd Term:1st

Email:samanta2515@student.nstu.edu.bd

Table of Contents

1. Introduction:

A software application or program feature that identifies Bangla misspelled words and notifies the user. Depending on the Bangla spell checker, the feature may automatically show the misspelled word or provide the user with a list of possible corrections. It gives alternative/correct spellings to the words you might have confused and misspelled. It scans each and every finds the words in it and compares each word with a well-known list of spelled words (dictionary). To find out the misspelled Bengali words use some algorithm like String Matching, Edit Distance which are work of intelligent mixed up. After implement, then its will be get the most probable suggestion for our misspelled Bengali words. If a file containing text, is given in this software, one can easily catch the mistakes and can correct them. For knowing about Bangla spelling, this software is useful.

2. Target Customers:

- Student
- Teacher
- Journalist who works in Bangla newspaper
- Content Writer
- Government officers
- Who writes document in Bangla

3. Features & Description:

In my proposed software,

a) **File Read and Detection**: The user can input any doc file. If any misspellings are found, then One window open and show the possible correct word for those misspelled word.
b) **Show Misspelled word**: The words of the file will be checked and misspelled or wrong words will be marked by red underline.
c) **Correcting Misspelled word with drop down box**: By clicking on the wrong or misspelled words, as suggestion, a bunch of correct and similar words will be shown.
d) If users want, they can add that word to my dictionary

3.1 Algorithm for spell checkers:

There are several types of spell checker algorithms. On this thesis described about two algorithms:

- String Matching
- Edit Distance

3.1.1 String Matching Algorithm:

String matching is an algorithm that attempts to locate one or more strings within a given text. It is an important class of string algorithms that searches a given position for a single string or multiple strings.

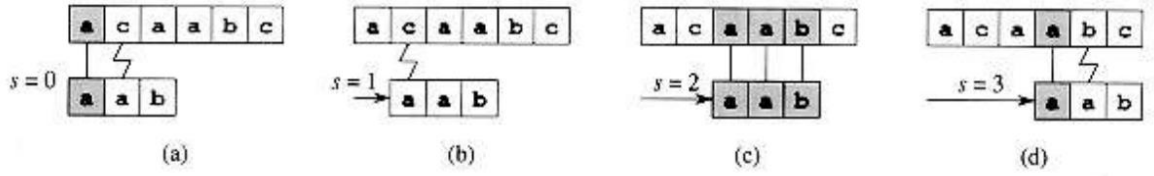


Fig 4.2: Processing on String Matching Algorithm

3.1.2 Edit Distance:

Edit distance is a category of such type algorithm which determine the dissimilarity between two or multiple strings. There are numerous varieties of edit distance. they are –

- Levenshtein distance
- Longest common subsequence (LCS)

3.1.3 Levenshtein distance:

This distance measures how much minimum operations or single character edits are needed to change one string to another. Single character edits mean insertion, substitution or deletion.

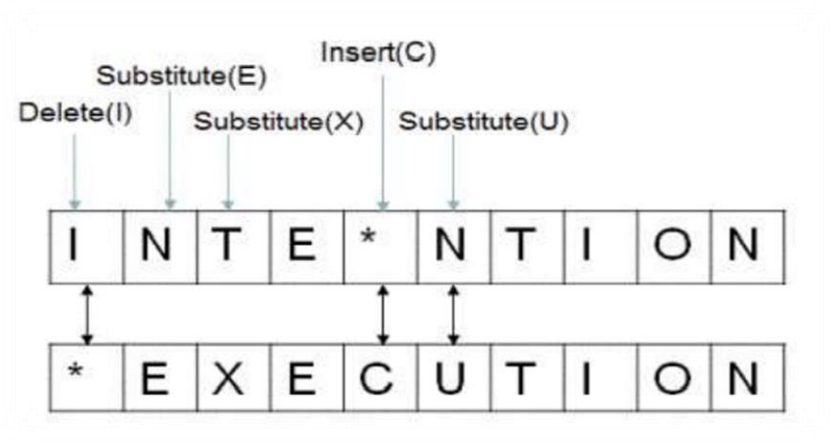
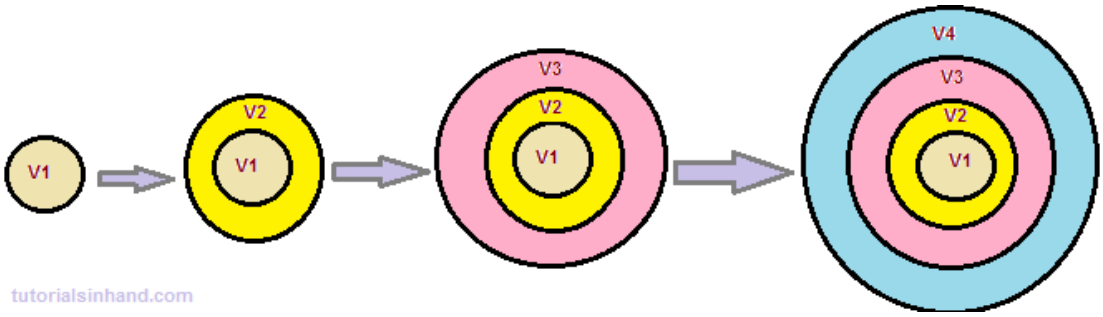


Fig4.3: Processing on Levenshtein Distance

4. Models, Tools and Resources:

Model:

Evolutionary model is a combination of iterative and incremental approach to software development. Evolutionary model is commonly used when the client wants to start using the core features instead of waiting for the full project. Evolutionary model is also used in object oriented software development because the system can be easily portioned into units in terms of object. The Evolutionary development model divides the development cycle into smaller, incremental waterfall models in which users are able to get access to the product at the end of each cycle. We have followed this model in our project because, in this model it reduces the error because the core modules get tested thoroughly. And a user gets a chance to experiment partially developed system.



IDE:

An IDE (Integrated Development Environment) contains a code editor, a compiler or interpreter, and a debugger, accessed through a single graphical user interchange (GUI). Our IDE is Eclipse.

Language:

- Java (Framework : Swing)

5. Project Members:

Developed by:

Sanjida Akter Samanta

BFH2025010F

Session:2019-20

Year:2nd Term:1st

Email: samanta2515@student.nstu.edu.bd

6. User Guide:

Step-1: (Version Guideline)

- User must have java 17.0.2 or above version updated on their device.

Step-2:

Go to link. JDK Download for Java download JDK 17.0.2:

- Accept License Agreement

- Run the exe for install

Figure 6.1: JDK Installation

Step-3:

Once you install Java in windows, click Close

Figure 6.2: JDK Installation

Step-4:

After installing JDK, check with any command prompt to make sure jdk is set in device environment

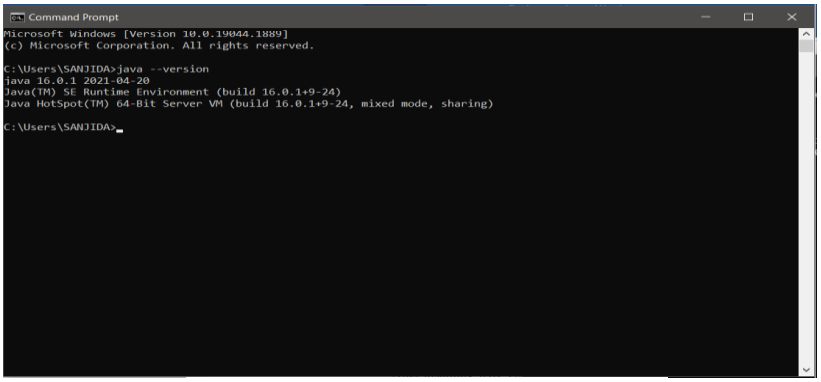


Figure 6.3: Java Version Check

Step-5:

Then download from here this project.

https://github.com/Hima0X2/Bangla_Spell_Checker

Step-6:

Then change Text-file encoding in UTF_8.

Properties->Resource-> Text-file encoding in UTF_8.

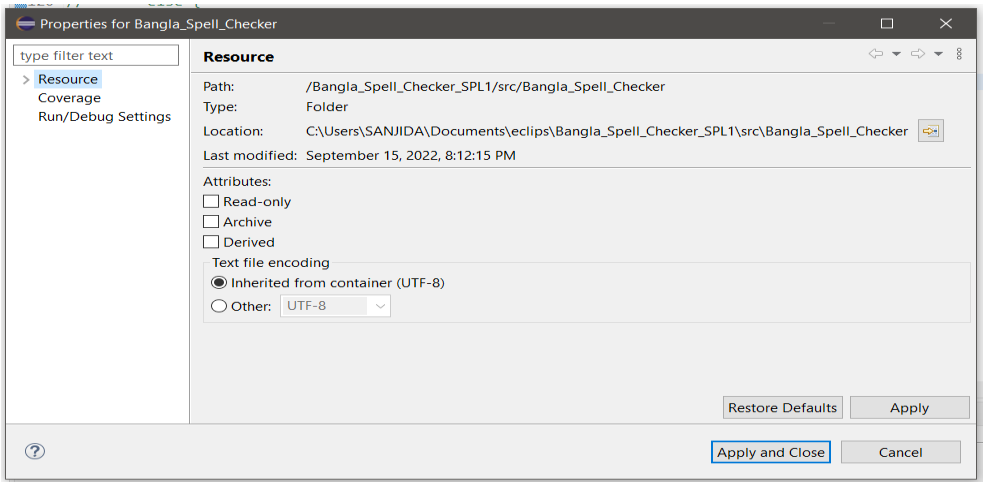


Figure 6.4: Encoding change

Step-7:

After running my code, user will see such kind of Window.

Fig6.5: First Window

Step-8:

When run the main method, user will see such kind of screen, Then after completing the loading Progress bar there will be open the below figure.

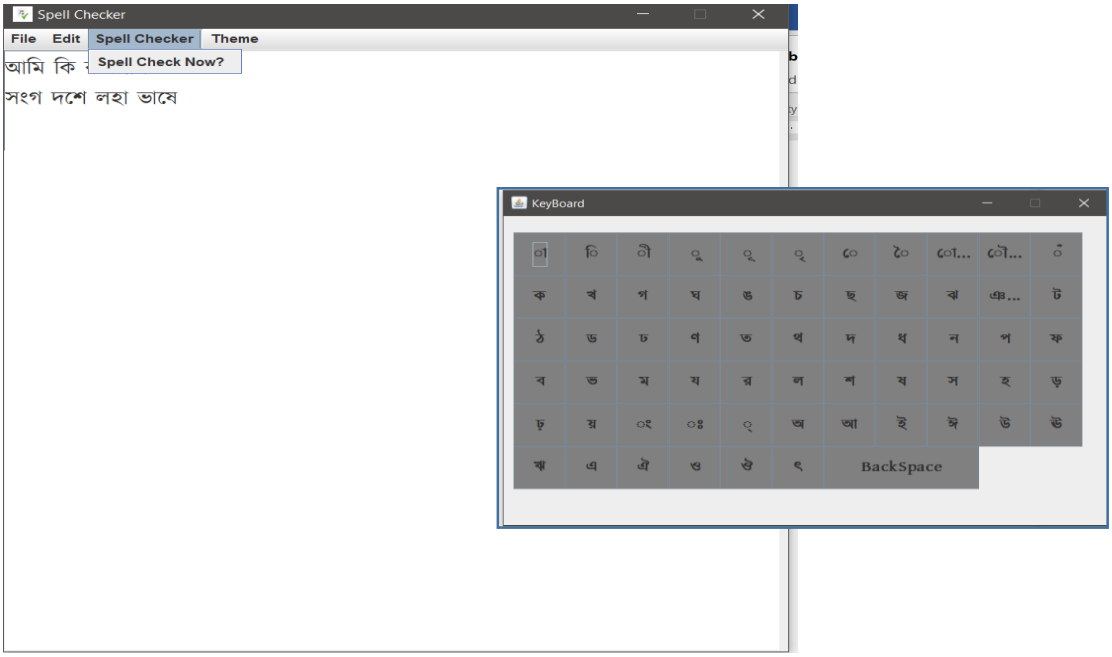


Fig6.6: Spell Checking Window

In this part,user write some words and click check spell now? If the word is correct then it will remain the same.If not then there will be open the below figure.

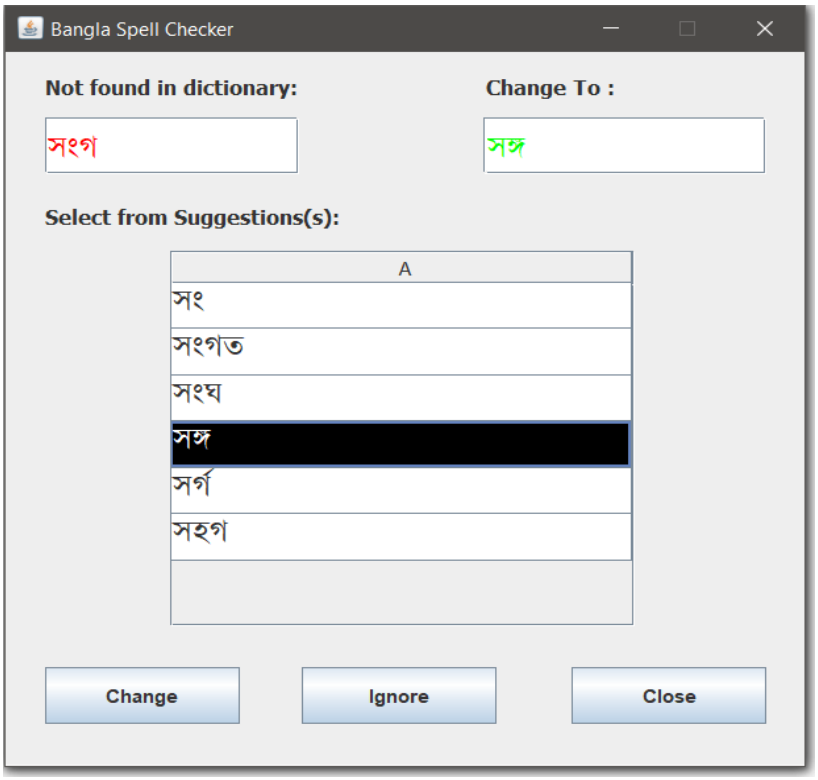


Fig6.7: Suggest word Window

Here which word is not found that is show in not found in Dictionary. Here a list of possible words is shown in a table. If any word is click then show the Change to textField if change Button is click then change the word in spell checking window,if press ignore then ignore this one and if close then close the window.

- It also open files

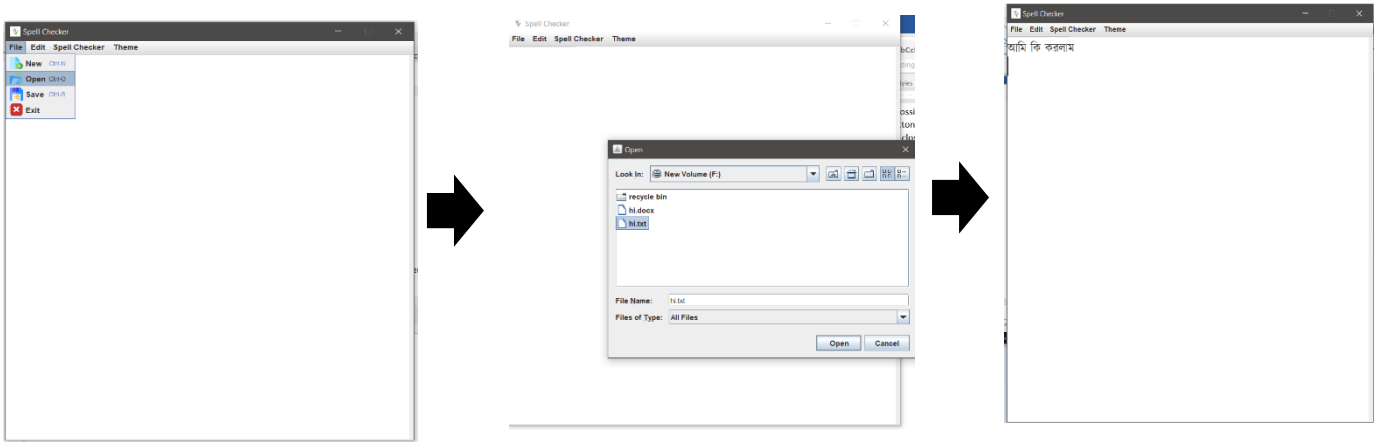


Fig6.8: Open a File from Spell Checking Window

- It also save files

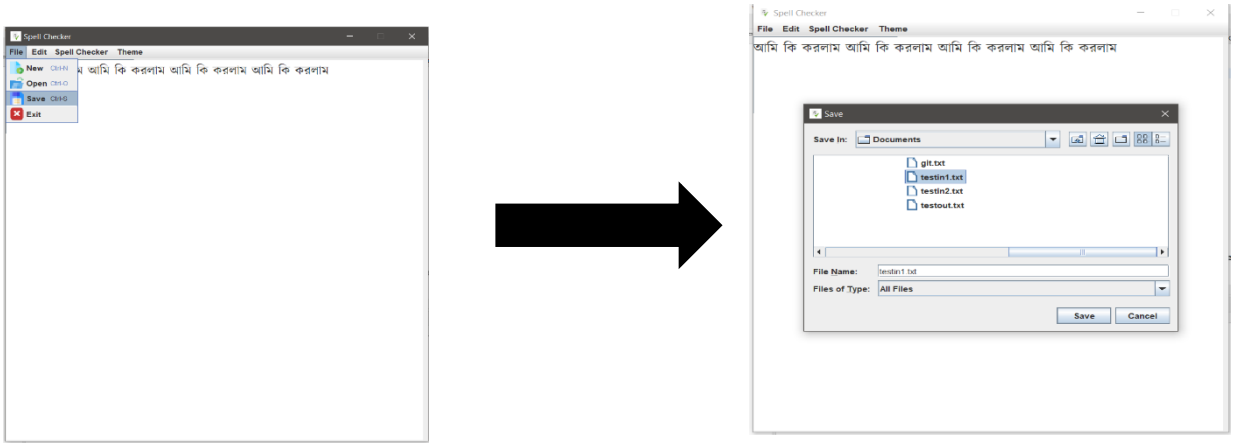


Fig6.9: Save a File to Spell Checking Window

- For Exit File

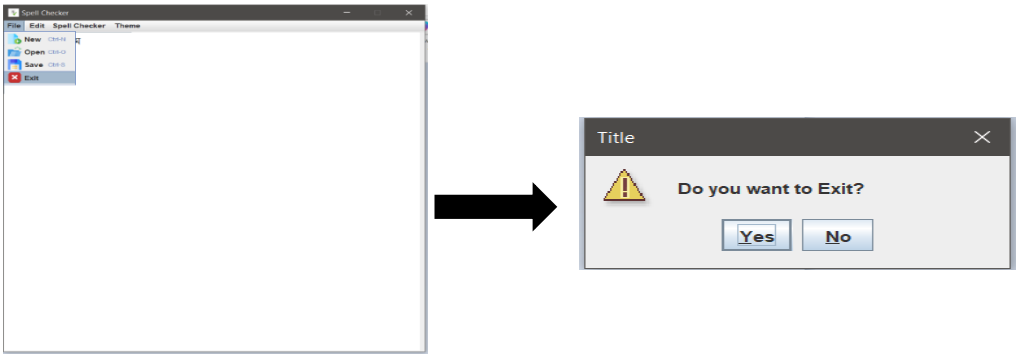
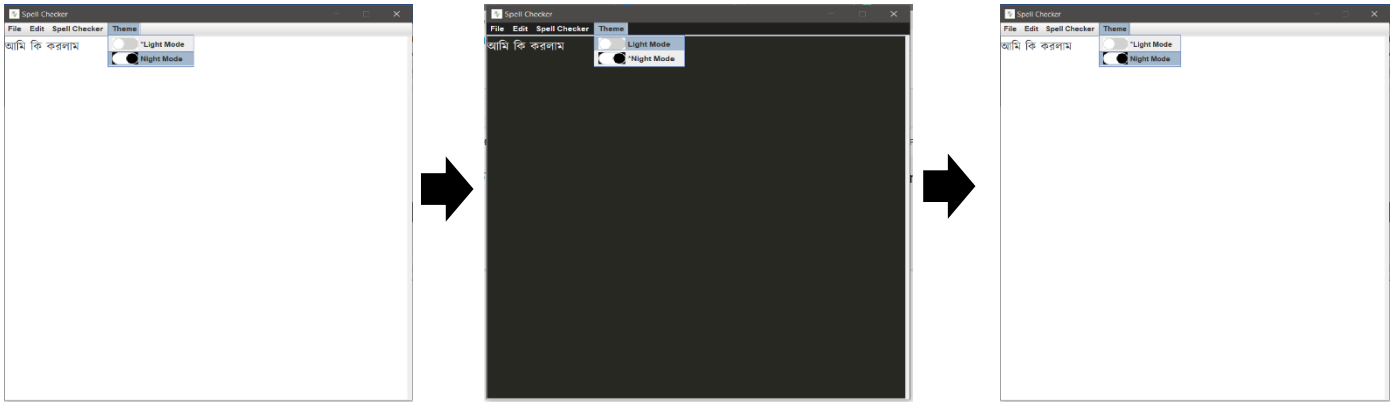


Fig6.10: Exit File

- For Change Theme



Light Theme Night Theme Light Theme

Fig6.11: Theme of Spell Checking Window

7. Challenges:

- ***Bangla word:*** As unicodes of Bangla words are not well funrnished That's why it's a hard to handle whose unicodes.
- ***Bangla "Juktoborno":*** As I worked with Bangla language, the uniqueness and difficulty also in the Bangla is juktoborno.
- ***File handling****:* As I have restriction of using database, so I have to work with text file. And for storing huge amount of words and working with file was a challenge!
- ***Algorithm Implementation****:* The algorithm I used it could create feature creep problem. The feature might not be handy and seems complex to the user.
- ***Tree and Recursion:*** It was hard to create possible words using Tree and Recursion for Bangla as Bangla Unicode Doesn't well funrnished.

Future Direction: I will do more with spell checker. I will work with word suggestion and replace correct word with wrong word.It will work fine if we write the wrong word of a complex word like some Juktoborno and large length of word.

- ***Create Keyboard:*** I will try to create a keyboard as like as Avro keyboard.

8. Source Code Documentation:

In my code there are 8 classes I use:

[1] Dic_Index.java

Constructor:

Description	Name	Parameter	Access Modifier
Index calculate for bangla unicode	Dic_index()	char	public

Method:

Description	Name	Parameter	Return type	Access Modifier
Takes Index Value	get_index()	No perameter	int	pubic

Here is a constructor here all bangla Unicode characters store by numerical value.

get_index method which is return int value is return the index value.

[2] Dictionary.java

Method:

Description	Name	Parameter	Return type	Access Modifier
Insert Value from Dictionary	insert()	Node root, String	void	pubic
Create some possible words according to misspelled word	Suggestions()	Node root, String	void	public
If modify words is found in tree or not	checkPresent()	Node root, String	boolean	public
If a word is correct or not	searchword()	String	boolean	public
works how many characters are different in two words that is counted	editDist()	String,int	int	public
Modify word which is incorrect and change it.	modify()	String	TreeSet<String>	public

[3] Main.java

When compile this code Intro class is Open.

[4] Intro.java

Constructor:

Description	Name	Parameter	Access Modifier
Create window and going running method	intro()	no parameter	public

Method:

Description	Name	Parameter	Return type	Access Modifier
Loading bar processing	running()	No perameter	void	pubic

[5] Filecontroller.java

Constructor:

Description	Name	Parameter	Access Modifier
Path detecting using Unicode	Filecontroller	String	public

Method:

Description	Name	Parameter	Return type	Access Modifier
Return the path	pathName()	No perameter	String	pubic

[6] Notepad.java

Constructor:

Description	Name	Parameter	Access Modifier
Create Spell Checking Window	Notepad()	no parameter	public

[7] Spell.java

Constructor:

Description	Name	Parameter	Access Modifier
Unicode constructing	Spell()	String	public

Method:

Description	Name	Parameter	Return type	Access Modifier
Store value from path file	construct()	String	LinkedHashMap<String, Integer>	pubic
If word correct or not	trueString()	String	boolean	public
Misspelled word possible word check(manually)	getSuggestionSpell()	String	TreeSet<String>	public

[8] Suggest_word.java

Constructor:

Description	Name	Parameter	Access Modifier
Create suggest word window and show a table from using possible words and change the word with misspelled word	Suggest_word	No perameter	public

This is a flow chart of my whole work:

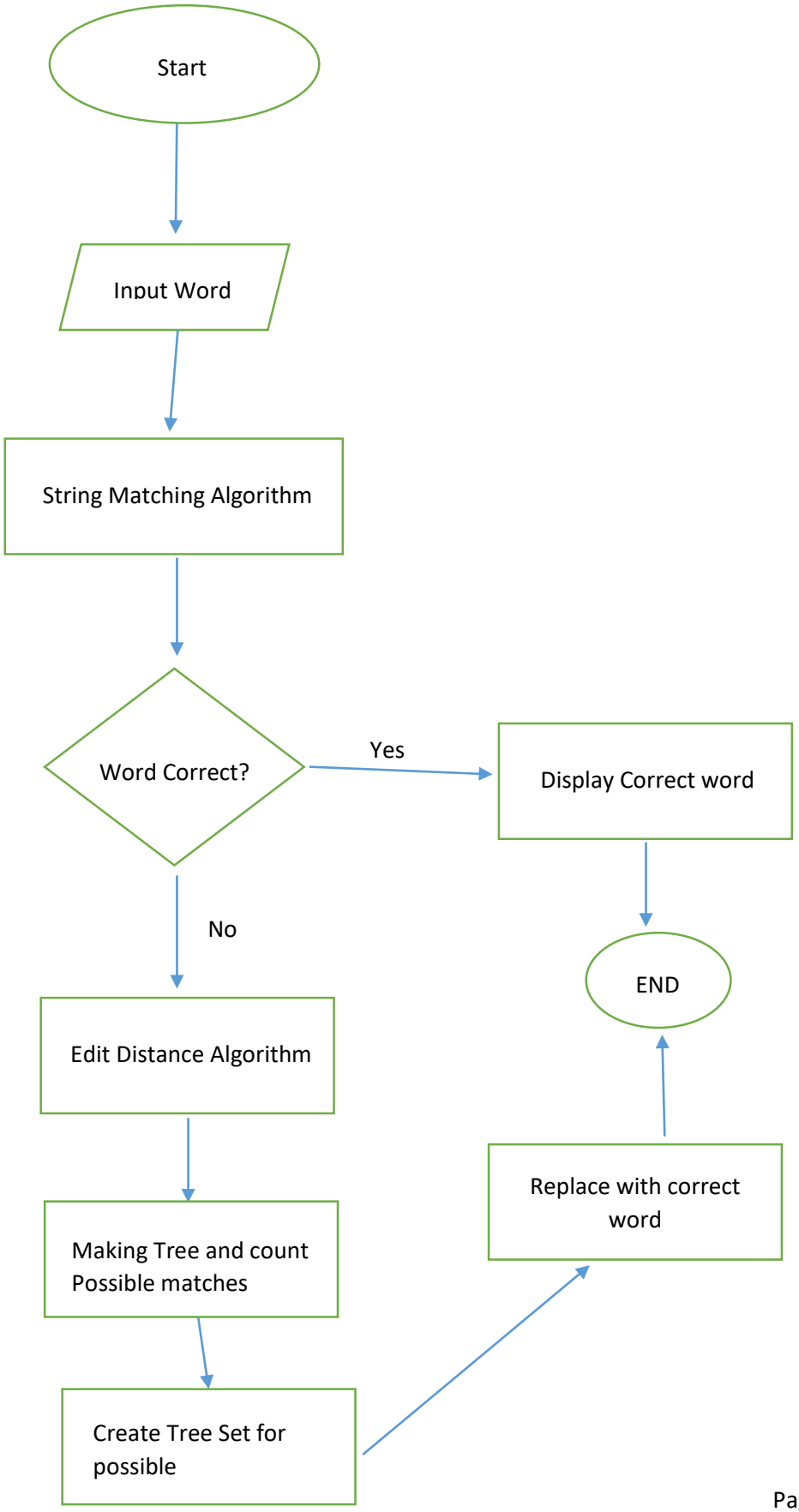


Figure 8.1: Encoding change

9. References:

[1] Introduction to Algorithms by Thomas H. Carmen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. [Third Edition]

[2] Java The Complete Reference by Herbert Scheldt. [Eleventh Edition]

[3] Data Structures by SEYMOUR LIPSCHUTZ [Third Edition]