Unified Devanagari Rendering Engine for Nepali Language

Team Members

Amar Dura [THA077BCT007]
Ayush Bhandari [THA077BCT014]
Harish Joshi [THA077BCT018]
Sugam Pokharel [THA077BCT044]

Supervised By
Er. Praches Acharya

Co-supervised By
Er. Santa B. Basnet

Department of Electronics and Computer Engineering
Institute of Engineering
Thapathali Campus
Kathmandu, Nepal

PRESENTATION OUTLINE

- Motivation
- Objectives
- Scope of Project
- Proposed Methodology
- Expected Result
- Project Applications
- Gantt Chart
- Estimated Budget
- References

MOTIVATION

- Lack of open source tools for pdf rendering for Nepali language
- Inconsistency in glyphs in different fonts for Devanagari script

OBJECTIVE

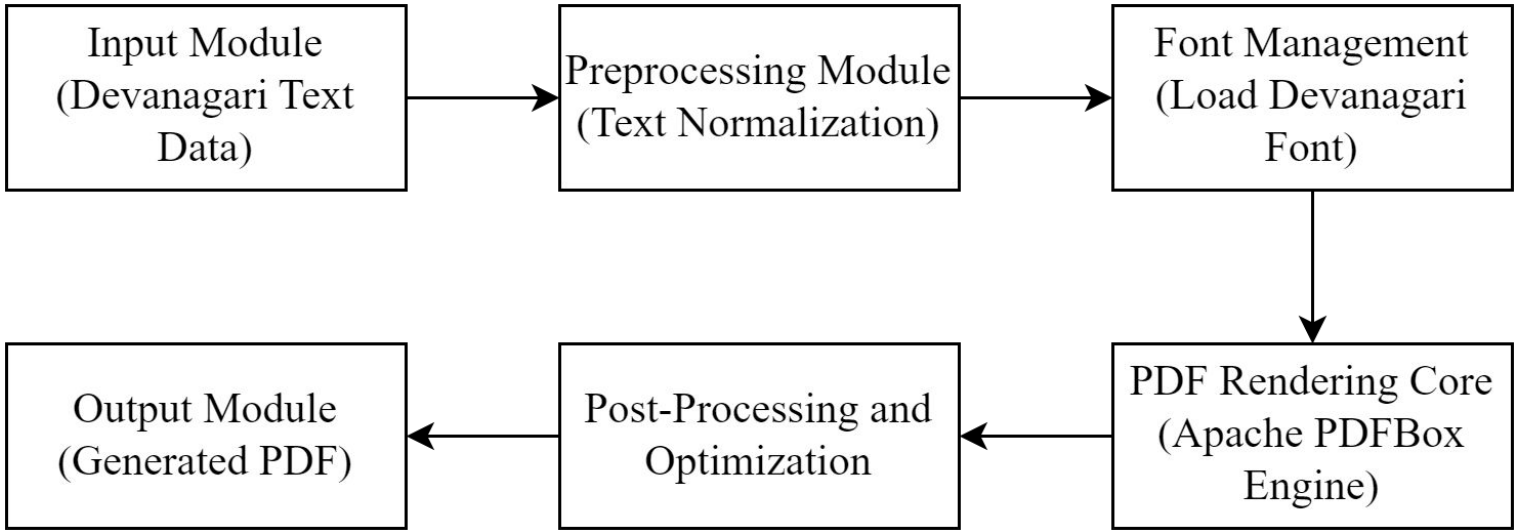
- To standardize glyph ordering mechanism using the Unicode Standards and Nepali Brihat Shabdakosh
- To develop an open source unified rendering engine for Nepali language in JVM using Apache PDFBox

SCOPE OF PROJECT

- Ensures the correct ordering of glyphs for Nepali text
- Focus on rendering on PDF file
- Only implemented for multi-byte Unicode fonts
- Apache PDFBox works on JVM only

METHODOLOGY - [1]

System Block Diagram



METHODOLOGY - [2]

Input Module

- Accepts Devanagari text
- Data can be provided through text files or databases

Preprocessing Module

- Handles glyph ordering to ensure that characters are correctly sequenced

METHODOLOGY - [3]

Font Management

- Ensures that the rendering of the glyphs is in same design as the chosen font.

PDF Rendering Core

- Handles the rendering of the Devanagari text onto the PDF.

METHODOLOGY - [4]

Post-Processing and Optimization

- Optimize the rendered PDF for size, performance and visual quality.

Output Module

- Responsible for saving and outputting the final generated PDF document

METHODOLOGY - [5]

Implementation details: First Part

- Algorithm for Devanagari Glyph Ordering
- Unification of Composite Character Representation
- Environment Setup
- Test Data Preparation

METHODOLOGY - [6]

Implementation details: Second Part

- Integration with PDF Rendering Engine
- Handling Composite Characters
- Testing and Validation
- Correctness Measure

METHODOLOGY - [7]

Algorithm for Devanagari Glyph Ordering

- Will be derived from the rules set by the Unicode Standards

Unification of Composite Character Representation

- Prepare comprehensive composite character chart based on Nepali Brihat Sabdakosh

METHODOLOGY - [8]

Environment Setup

- Learn Java Programming
- Understanding the Apache PDFBox codebase

Test Data Preparation

- Words of Nepali Brihat Shabdakosh(60-70K words) will be used as the test data.
- Categorization of mis-rendered glyphs will be done.

METHODOLOGY - [9]

Integration with PDF Rendering Engine

- Modifying the text rendering pipeline in PDFBox
- Include new glyph ordering logic

Handling Composite Characters

- Implement standard composite character chart developed in first part

METHODOLOGY - [10]

Testing and Validation

- Manual testing for visual correctness of rendered text
- Automated tests for validating glyph ordering algorithm
- Java testing framework will be used.

Correctness measure

$$\text{Accuracy} = \frac{\text{Correctly ordered glyphs}}{\text{Total words}}$$

$$\text{Average accuracy} = \frac{\sum_{i=1}^{k} \text{Accuracy (Category)}_i}{k}$$

where, k = Number of categories

EXPECTED RESULT - [1]

Current Output

Noto Sans Devanagari

वद्यिार्थीहरू अध्ययन गर्दैछन्। शक्षिकि अत्यन्तै ज्ञानवान छन्।

Mangal

वद्यिार्थीहरू अध्ययन गर्दैछन्। शक्षिकि अत्यन्तै ज्ञानवान छन्।

Output of Proposed System

Noto Sans Devanagari

विद्यार्थीहरू अध्ययन गर्दैछन्। शिक्षिका अत्यन्तै ज्ञानवान छन्।

Mangal

विद्यार्थीहरू अध्ययन गर्दैछन्। शिक्षिका अत्यन्तै ज्ञानवान छन्।

EXPECTED RESULT - [2]

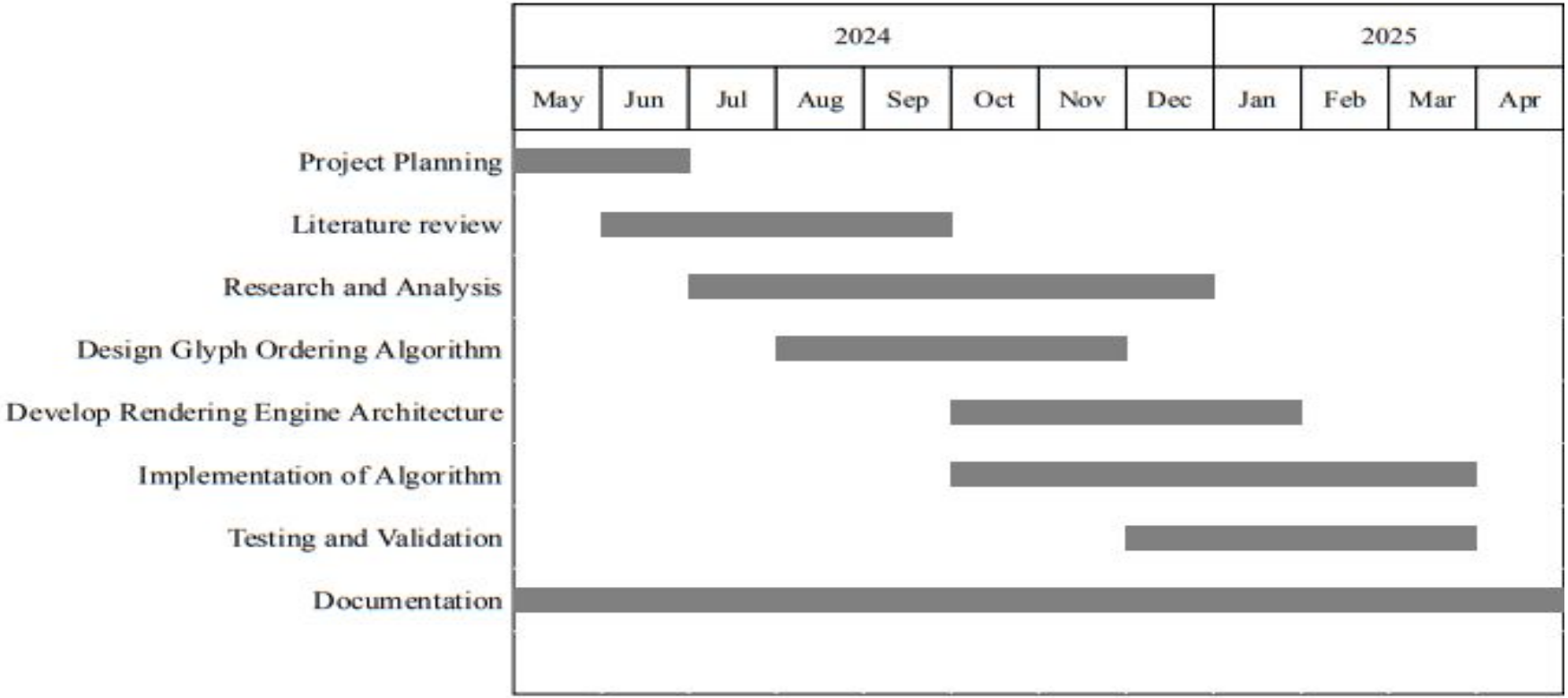
Uniform of composite character representation

Noto Sans Devanagari	अङ्क	ब्राह्मण	प्रह्लाद	र्याली	मङ्गल
Mangal	अङ्क	ब्राह्मण	प्रह्लाद	र्याली	मङ्गल
Kalimati	अङ्क	ब्राह्मण	प्रह्लाद	र्याली	मङ्गल
Kantipur	अङ्क	ब्राह्मण	प्रह्लाद	र्याली	मङ्गल
नेपाली बृहत् शब्दकोश	अङ्क	ब्राह्मण	प्रह्लाद	ऱ्याली	मङ्गल

PROJECT APPLICATIONS

- News compilation
- Financial reporting
- Educational materials
- Governmental outreach

GANTT CHART


2024
2025
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
Project Planning
Literature review
Research and Analysis
Design Glyph Ordering Algorithm
Develop Rendering Engine Architecture
Implementation of Algorithm
Testing and Validation
Documentation

ESTIMATED BUDGET

Particulars	**Amount (Rs.)**
Memory upgrade	8000
Printing expense	5000
Miscellaneous Costs	5000
Total	18000

REFERENCES - [1]

- S. Chandra, “Indian typefaces in digital platforms: Issues and challenges,” in Advances in Design and Digital Communication, N. Martins and D. Brandão, Eds. Cham: Springer International Publishing, 2021, pp. 125–136.
- S. B. Basnet and S. Trishna, “Unification of fonts encoding system of devanagari writing in nepali,” Nepalese Linguistics, vol. 32, no. 2, pp. 130–136, 2017.
- K. S. Lydon, “Designing devanagari type: the effect of technological restrictions on current practice,” Ph.D. dissertation, 2015.

REFERENCES - [2]

- M. Boualem, M. Leisher, and B. Ogden, “Encoding script-specific writing rules based on the unicode character set.”
- S. P. Mudur, N. Nayak, S. Shanbhag, and R. Joshi, “An architecture for the shaping of indic texts,” Computers & Graphics, vol. 23, no. 1, pp. 7–24, 1999.
- U. Consortium et al., “The unicode standard, version 14.0. 0–core specification,” 2021.