System Requirements Specification

1. Introduction

1.1 Purpose
This document outlines the requirements for the Textify system. The main feature of Textify will be recognizing text from images using Optical Character Recognition (OCR). Additionally, it will allow users to input text in a source language and receive a translation in the desired target language, as well as support features such as exporting translations.

1.2 Scope
The Textify application shall provide real-time text translation, text recognition from images via OCR, and the ability to export translations in various formats. The application shall offer a user-friendly interface accessible via web browsers on both desktop and mobile platforms.

1.3 Definitions, Acronyms, and Abbreviations

- **API**: Application Programming Interface
- **OCR**: Optical Character Recognition
- **NMT**: Neural Machine Translation
- **TTS**: Text-to-Speech
- **STT**: Speech-to-Text
- **Export**: Saving or sharing translation output in formats such as PDF, Word, or TXT

1.4 Overview
This document outlines the functional and non-functional requirements for Textify, including features for text translation, OCR-based text recognition, TTS, STT, and exporting. The responsibilities of various project roles are also defined.

---

2. General System

2.1 Product Overview

- Textify shall be a machine translation system capable of text translation and exporting outputs in multiple formats.
- The system shall support OCR to extract and translate text from images.
- Textify shall allow speech input transcription via STT and convert text to audio files using TTS.
- The application shall function as a web-based platform accessible on any standard web browser.

2.2 Roles and Responsibilities

- **Customer**: Shall define high-level requirements and review progress.
- **Manager**: Shall oversee timelines and deliver reports.
- **Designer**: Shall create the user interface with a focus on usability and aesthetics.
- **Programmers**: Shall implement functionality for text translation, OCR, TTS, and STT.

- **Testers**: Shall verify that all features meet the requirements and function as intended.

---

3. Functional Requirements

3.1 Text Translation

- **Requirement**: The system shall accept input text in a source language and translate it into a target language.
- **Acceptance Criteria**: Users shall enter text and select source/target languages. The system shall provide a translation within 2 seconds for text under 500 words.

3.2 Text-to-Speech (TTS)

- **Requirement**: The system shall convert text into speech and allow users to download audio files in standard formats.
- **Acceptance Criteria**:
    1. The system shall support playback control (pause, stop, resume).
    2. Users shall adjust speed (slow, normal, fast) and volume via a slider.

3.3 Speech-to-Text (STT)

- **Requirement**: The system shall transcribe spoken input into text with 90% accuracy in moderate noisy environments.
- **Acceptance Criteria**: Users shall upload audio files or provide live input. Transcriptions shall include proper punctuation and capitalization.

3.4 Optical Character Recognition (OCR)

- **Requirement**: The system shall extract text from uploaded images with 95% accuracy for standard fonts.
- **Acceptance Criteria**:
    1. Users shall upload images in formats such as PNG and JPEG.
    2. The system shall preprocess images (e.g., de-skewing, noise reduction) before text extraction.

3.5 User Interface

- **Requirement**: The application shall have a responsive design for desktops and mobile devices.
- **Acceptance Criteria**:
    1. The interface shall allow users to easily access translation, OCR, TTS, and STT features.
    2. Key features shall load within 1 second on standard devices.

3.6 Exports

- **Requirement**: The application shall provide an export function for the converted text.
- **Acceptance Criteria**:
    1. The interface shall allow users to easily export into various formats.

---

4. Non-Functional Requirements

4.1 Scalability

- **Requirement**: The system shall support multiple concurrent users with a response time under 3 seconds.

4.2 Security

- **Requirement**: All data transmissions shall be encrypted using SSL/TLS. User data shall not be stored in unencrypted formats.

4.3 Privacy

- **Requirement**: User data shall only be used for training models with explicit consent.

4.4 Maintainability

- **Requirement**: The system shall have a modular architecture with clear interfaces, minimal dependencies, and centralized management for code, dependencies, and configuration to ensure maintainability and ease of updates.
- **Acceptance Criteria**:
    1. Each module shall have clear interfaces and minimal dependencies.
    2. Code shall include inline comments and detailed documentation.
    3. Code shall be managed via a version-control-system.
    4. Dependencies and third party integrations shall be managed via central Dependency management.
    5. Deployment and Configuration of the application shall be managed via a central file.

4.5 Reliability

- **Requirement**: The system shall provide error detection, logging with timestamps and error codes, and clear error messages, supporting debugging and automated testing where feasible.
- **Acceptance Criteria**:
    1. Logs shall include timestamps and error codes.
    2. Users shall receive clear, actionable error messages.
    3. The application shall have automated tests if possible.

---

5. Use Cases

5.1 Use Case: Text Translation

- **Actors**: User
- **Description**: The user shall input text and select source/target languages for translation.
- **Preconditions**: The user is on the translation page.
- **Postconditions**: The translated text is displayed to the user.

5.2 Use Case: OCR (Image-to-Text)

- **Actors**: User
- **Description**: The user shall upload an image for text extraction using OCR.
- **Preconditions**: The user is on the OCR page and has an image to upload.
- **Postconditions**: The extracted text is displayed.

5.3 Use Case: Export Translation

- **Actors**: User
- **Description**: The user shall export translations into formats such as PDF or Word.
- **Preconditions**: A translation is completed.
- **Postconditions**: The translated text is downloaded in the selected format

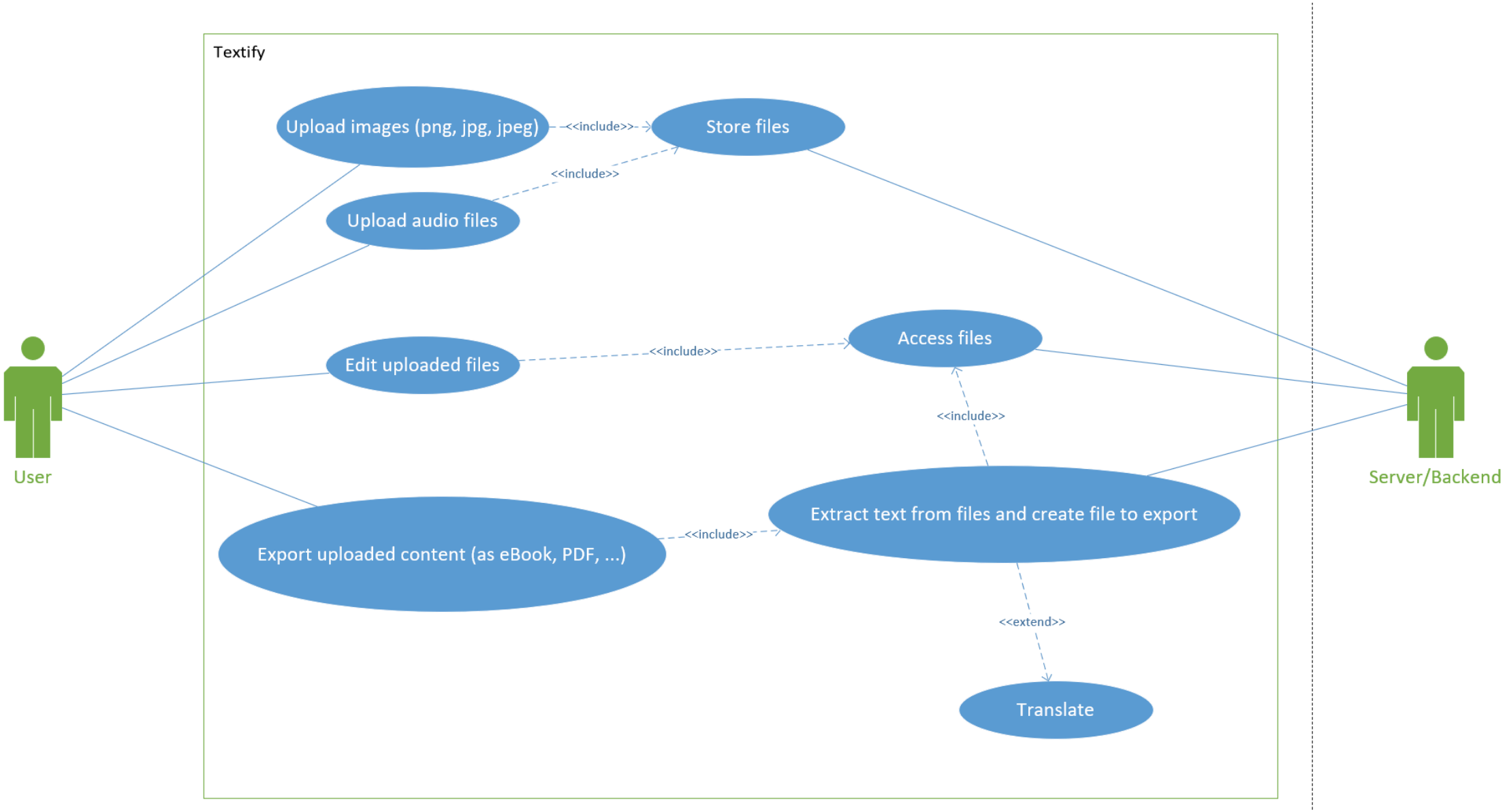
5.4 Use Case: STT

- **Actors**: User
- **Description**: The audio file is transcribed so that a text file containing the content of the audio file is output.
- **Precondition**: The user uploads an audio file.
- **Postcondition**: The user exports his text file or can translate it.

5.5 Use Case: TTS

- **Actors**: User
- **Description**: The system synthesises the speech in an audio file. This can be played on the website and downloaded as a file.
- **Precondition**: The user copies or writes a text into an input field.
- **Postcondition**: The user has an mp4 file.

5.6 Use Case Diagram



6. System Design and Architecture

6.1 Architecture Overview

- **Frontend**: A web-based user interface for interaction.
- **Backend**: A server hosting the AI translation model and managing API communication.
- **Database**: Storage of user data, logs, and translation history.

6.2 Sketch System Design: