Team Name:Shazam
Name of College(s)/University(s): Anurag University

Team Members Details:

1. Rajidi Raguveer Reddy
2. Dattatreya Dussa
3. Tata Vishnu Vardhan
4. Vedhin Govind

Proposed Solution :

The integration of voice recognition technology into web-based applications has fundamentally transformed user interfaces, making them more accessible and intuitive. Our initial phase involves fine-tuning the Whisper model on a curated set of commands, including "zoom in," "zoom out," "zoom in to Hyderabad," "change to satellite view," and "add marker to a specific location.”.Whisper is a multilingual model which fine-tuned on multiple languages . Once speech is converted to a text string, this text undergoes processing through an NLP framework, such as a TensorFlow-based model or a fine-tuned DistilBERT variant. The output is a list of text classified into distinct categories.

Named Entity Recognition (NER), a specialised sub-task within NLP, is employed to identify and categorise named entities within the text. These entities, categorised into location and event classes, are compared against predefined event parameters (e.g., "zoom in," "zoom out," "change view," "add marker") and location parameters (e.g., Hyderabad). For instance, the command "zoom to Hyderabad" should result in a zoom-in action centred on Hyderabad city, addressing potential ambiguities and context-dependent queries.

This action is executed using the OpenLayers GIS library, with rendering facilitated by WebGL for smooth performance. We optimise the rendering pipeline to enhance efficiency and performance. Additionally, robust error handling is critical in a voice-enabled geospatial map application to ensure a seamless and user-friendly experience. This involves setting confidence thresholds for speech recognition and implementing fallback mechanisms and clarification prompts to handle uncertainties effectively.

Tools and Technology Used:

- Programming Languages: JavaScript, Python
- Voice Recognition: Whisper for accurate speech-to-text conversion
- NLP Framework: TensorFlow for command processing and entity recognition
- GIS Library: OpenLayers for geospatial data management
- Map Rendering: WebGL for high-performance graphics rendering

1. How different is it from the any of the other existing ideas?

There are currently no voice-enabled user interfaces for geospatial map-based web applications. Qwhery, a voice assistant-based application, operates through Alexa and Google Assistant, integrating geospatial maps to process queries or commands but lacks a user interface like ours.Existing interfaces like Google Maps and Earth can't process human language effectively. For example, a voice command like "zoom in on Hyderabad" on Google Maps doesn't differentiate between action and location, showing places named "zoom in on" in Hyderabad.We address this by building an NLP system with TensorFlow that identifies and classifies actions and locations.

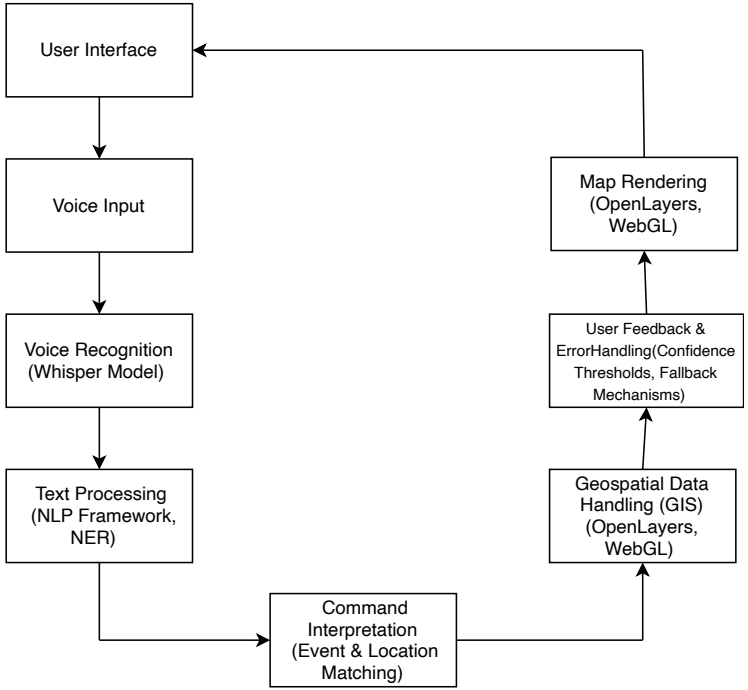
2. How will it be able solve the problem?

We combine voice interpretation with a user interface enabled through voice input. Using a fine-tuned Whisper model for commands like "zoom in [location]" and "show [satellite] view," our TensorFlow-based NLP parses and categorises actions and locations.We also use entity recognition model to classifies event entities and pass them to event functions Each action triggers JavaScript code interacting with a web interface, using OpenLayers for maps and WebGL for rendering. We refine our model based on user feedback.

3. USP of the proposed solution:

We enhance user experience with personalised voice command shortcuts. For example, “how is the day outside” uses user location to provide weather and traffic data via voice. Voice integration confirms actions like “show the terrain view.” Our application predicts weather using past data with ARIMA or SARIMA. Users can add multiple markers to a map, measure distances, find common points, and check conditions at specific markers.

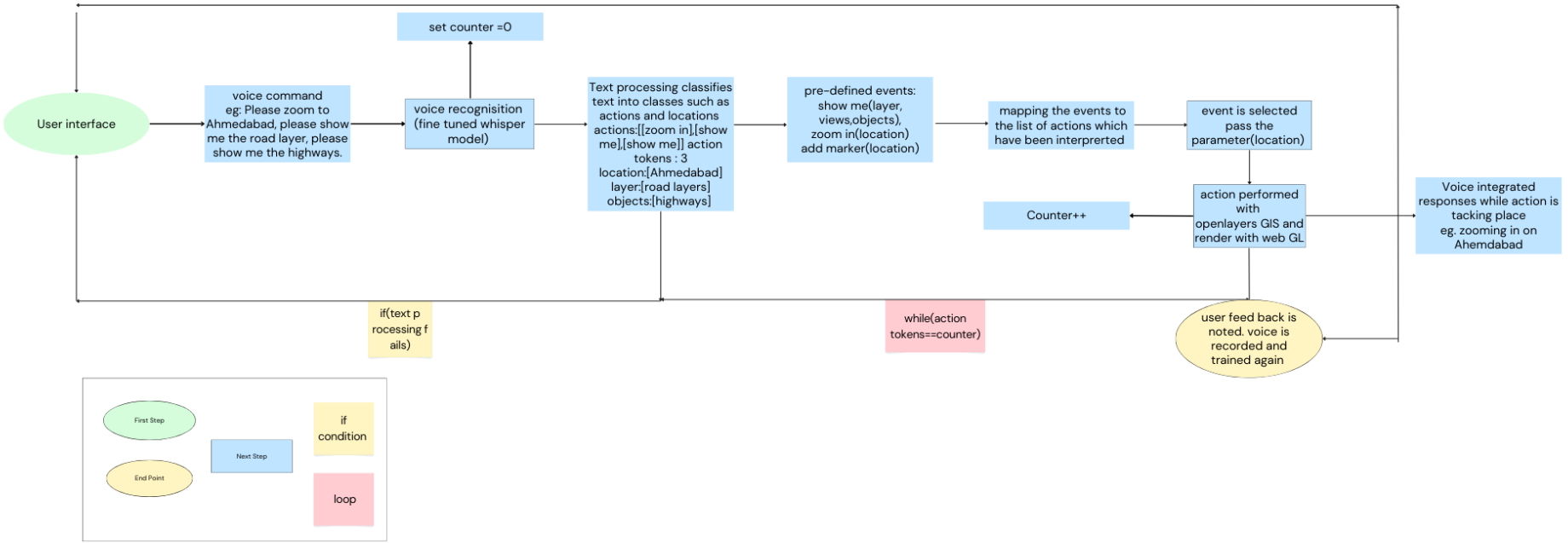
Proposed architecture/user diagram



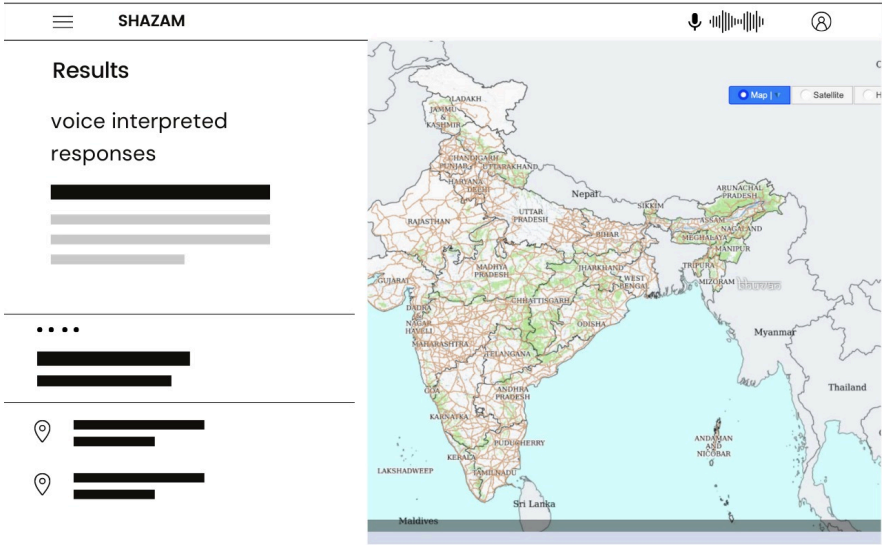
List of features offered by the solution

1. Voice command control
2. Multilingual voice recognition
3. Real-time geospatial maps interface
4. Voice command shortcuts
5. Multiple actions such as zoom in and out, hide or show layers, add markers etc..
6. Marker operations such as distance between two markers, common point among three markers, etc..
8. Runs mostly on machine hardware hardly leverages APIs for weather and Traffic data.
7. Voice integrated responses.

Process flow diagram or Use-case diagram



Wireframes/Mock diagrams of the proposed solution



Mock GUI of the website.

Solution Brief (Overall)

Our project introduces a revolutionary voice-enabled interface for geospatial web applications, combining advanced voice recognition with dynamic map interactions. We fine-tune the Whisper model for specific voice commands like "zoom in [location]" and "show [layer]" .Whisper is a multilingual model which fine-tuned on multiple languages . These commands are converted into text and processed through a TensorFlow-based NLP system, which uses Named Entity Recognition (NER) to categorise actions and locations accurately. For instance, a command such as "zoom to Hyderabad" results in a zoom-in action centred on Hyderabad.

The actions are executed using the OpenLayers GIS library, and map rendering is managed by WebGL for high performance. We optimise the rendering pipeline for efficiency and implement robust error handling to manage uncertainties in voice recognition. This includes confidence thresholds and fallback mechanisms to ensure a seamless user experience.

Unique Aspects and Problem Solving:

Existing solutions like Qwhery and Google Maps lack integrated user interfaces that process complex voice commands effectively. Our approach overcomes these limitations by combining voice interpretation with a user-friendly interface. Fine-tuning Whisper and leveraging TensorFlow’s NLP ensures accurate classification and execution of commands.

List of Features Offered:

1. Voice Command Control: Interact with the map using voice commands for intuitive navigation.
2. Dynamic Geospatial Maps Interface: Real-time updates and interactions with the map.
3. Voice Command Shortcuts: Personalised commands like “how is the day outside” for weather and traffic updates.
4. Multiple Actions: Includes zooming in and out, hiding or showing layers, and adding markers.
5. Marker Operations: Perform calculations such as distance between markers and finding common points among markers.
6. High Graphics Map Rendering: Enhanced visualisation with WebGL for smooth and detailed graphics.
7. Hardware-Based Execution: Runs primarily on machine hardware with minimal reliance on external APIs for weather and traffic data.
8. Voice Integrated Responses: Provides audible feedback for actions, ensuring clarity and confirming command execution.

Proposed Architecture/User Diagram:

The architecture includes:

1. Voice Input Module: Captures and processes voice commands.
2. NLP Processing: Converts text commands into actionable data using TensorFlow.
3. Action Execution: Implements commands through OpenLayers and WebGL.
4. User Interface: Features an interactive map with voice-enabled controls.
5. Feedback Loop: Collects user feedback to refine and improve the system continuously.

This solution redefines interaction with geospatial maps by integrating voice commands and advanced processing, Creating a more intuitive and user-friendly experience.


BHARATIYA ANTARIKSH HACKATHON 2024
Innovation partner H2S
इसरो isro
THANK YOU