Google Summer of Code (GSoC) Application

Basic Details

Full Name: Ajit Shah
Email: ajitshah6124@gmail.com
GSoC Nickname: AKAIZIC
First Language: English
Location: Mumbai, India (IST, UTC+5:30)

Previous Open-Source Contributions

- Contributed to an open-source retirement plan parser using AWS Textract and LLMs (https://codeberg.org/Remedy6784/emp_table_verify2.git)
- Built a topic classifier with spaCy for analyzing academic papers (https://github.com/AKAIZICOO7/Discussion-Microscope

Your Motivation

Why GSoC and GFOSS?

Google Summer of Code provides a structured avenue to contribute to impactful open-source projects while refining my technical expertise. I chose GFOSS due to its commitment to democratizing access to open technologies, particularly through projects like Archimedes that leverage AI for civic engagement. My goal aligns with GFOSS's mission: to build tools that foster transparency and informed discourse in digital communities.

Why This Project?

Online discussions are often fragmented, making it challenging to track arguments, sentiments, and user interactions systematically. By developing an open-source tool that combines NLP and interactive visualizations, I aim to empower researchers, educators, and moderators to dissect complex debates and identify biases.

Expectations from GFOSS

During the program, I seek mentorship to refine my technical approach and prioritize features that align with user needs. Post-GSoC, I hope to continue collaborating with GFOSS to expand the tool's capabilities, ensuring its long-term sustainability.

Technologies

- Core: Python, Hugging Face Transformers, NetworkX, Gephi.
- Visualization: Plotly, Dash, BERTopic.
- Data Processing: Convokit, Pandas.

Project Details

What Are You Making?

An open-source tool to analyze and visualize arguments, sentiments, and user interactions in online discussions. **Key features:**

- **Sentiment/Emotion Detection**: Using Hugging Face Transformers for nuanced analysis
- **Argumentation Graphs:** Dynamic network visualizations via Gephi/NetworkX
- **Interactive Dashboards:** Plotly/Dash for timelines, heatmaps, and hybrid visualizations

Impact on GFOSS

This tool will empower researchers, educators, and moderators to analyze discussions at scale, fostering transparency and inclusivity in digital spaces. By open-sourcing the codebase, GFOSS can strengthen its portfolio of tools for civic engagement and education.

Timeline

Community Bonding (May 8 – June 1)

- Finalize project scope with mentors.
- Set up development environment and study GFOSS's existing projects (e.g., Archimedes).

Phase 1: Data Pipeline & Sentiment Analysis (Weeks 1–4)

- **Weeks 1–2:** Implement data ingestion using Convokit, standardize formats (Reddit/Twitter).
- **Weeks 3–4:** Build sentiment analysis pipeline with Hugging Face models (e.g., fine-tuning for emotion detection).
- **Deliverable for Evaluation 1:** Functional data pipeline and basic sentiment analysis module.

Phase 2: Argument Graphs & Visualization (Weeks 5–8)

- **Weeks 5–6:** Create argumentation graphs using NetworkX; explore Gephi for dynamic layouts.
- **Weeks 7–8:** Develop interactive timelines and heatmaps with Plotly.
- **Buffer Week (Week 6):** Troubleshoot graph scalability issues.
- **Deliverable for Evaluation 2:** Prototype of argument graphs and visualization dashboard.

Phase 3: Integration & Testing (Weeks 9–12)

- **Weeks 9–10:** Combine visualizations into a Dash-based dashboard; integrate BERTopic for theme tracking.
- **Weeks 11–12:** User testing, documentation, and code optimization.
- **Buffer Week (Week 10):** Address feedback and refine UI.
- **Final Deliverable:** Fully functional tool with documentation and deployment guide.

Post-GSoC Plans

- Add multilingual support using NLP translation models.
- Collaborate with the Archimedes team to integrate LLM-based moderation features.

Why I'm a Good Fit?

Alignment with GFOSS Values:

Passionate about open-source education; contributed to tools used in academic research.

Committed to long-term collaboration, as seen in my post-GSoC plans.

Time Commitment

Weekly Hours: 35–40 hours (full-time).
Off-the-Grid Periods: None planned.

Final Statement

This project represents my commitment to leveraging open-source technologies for societal impact. With GFOSS's mentorship, I am confident in delivering a robust tool that enhances understanding of online discourse.