Leveraging LLM to Generate Alt Text for Improving Image Accessibility

Introduction

The Importance of Alt Text

- People with visual impairments (PVI) users **have difficulty viewing or understanding searched images.**

- PVI users understand images through screen readers **that read aloud the alt text** provided for the images..

Poor digital content accessibility

- However, many web pages **do not provide appropriate alt text** for the images they use.

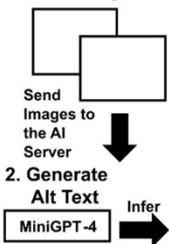- **Improving digital content accessibility** for PVI users is necessary.

Aim

- Develop a smartphone app that **generates detailed alt text** and **recommends images** based on user's voice input for the images in users' smartphone
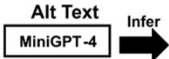- **Evaluate** the performance and usability of the developed app with **PVI users**.
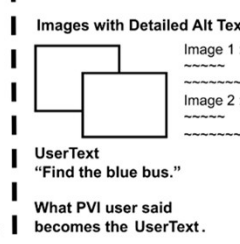
System

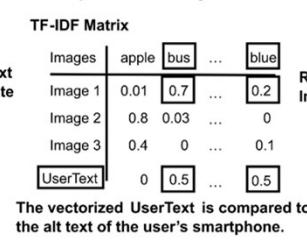### I. Generate Alt Text

1. Load Images
Send Images to the AI Server
2. Generate Alt Text
MiniGPT-4
Infer
Image 1 : ~~~~~
Image 2 : ~~~~~

Combine Images with the Alt Text
Detailed Alt Text

3. Save Alt Text
Images with Detailed Alt Text
Image 1 : ~~~~~ ~~~~~~
Image 2 : ~~~~~
SaGol searches for the images based on the alt text.

II. Search for Images

1. Tokenize the Alt Text
Images with Detailed Alt Text
Image 1 : ~~~~~
Image 2 : ~~~~~
UserText "Find the blue bus."
What PVI user said becomes the UserText.

Tokenize the Alt Text & Calculate TF-IDF

2. Compare Similarity of Vectors
TF-IDF Matrix

Images	apple	bus	...	blue
Image 1	0.01	0.7	...	0.2
Image 2	0.8	0.03	...	0
Image 3	0.4	0	...	0.1
UserText	0	0.5	...	0.5

Recommend Images

The vectorized UserText is compared to the alt text of the user's smartphone.

3. Select Images
Recommendation
Image 1 : blue~~~ ~bus~~~
Image 2 : ~~bus~~ ~blue~~~
SaGol selects the images with the most similar alt text to the UserText.
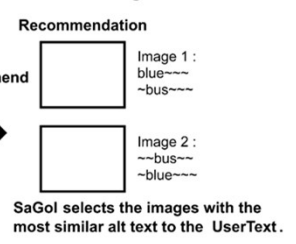
> (...)
> There are people walking on the street in front of the bus, and buildings in the background. The bus has large white letters on its side.
> (...)

The app works as follows:

1) **Search for the images** using voice recognition.
2) **Check** the voice recognition **results**.
3) **Listen to the alt text** of the images and **select** one.
4) **Replay** the alt text of the selected image, **share or delete** the image.
5) **Share** the image on other **social media platforms**.

CONTACT Hyunggu Jung(hyunggu@snu.ac.kr)