Leveraging LLM to Generate Alt Text for Improving Image Accessibility

Yunseo Moon[1], Hyunggu Jung[*][2]

[1]University of Seoul, [2]Seoul National University

*: Corresponding author

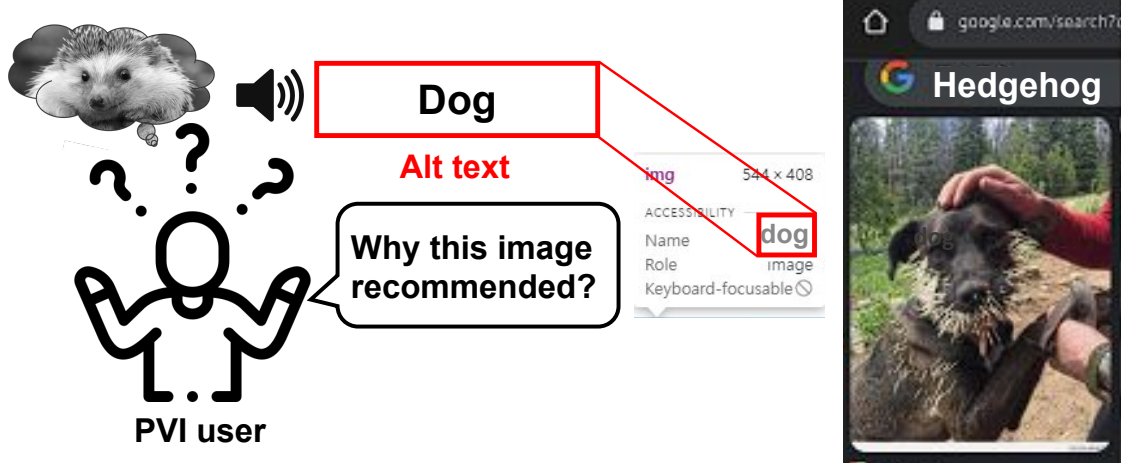# Introduction

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

- **Generate detailed alt text** for **improving image accessibility** with leveraging **Large Language Models**.
- Develop a smartphone app **to generate alt text** and help **PVI users search for** their **smartphone images** using alt text.

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 : ~~~~~
Detailed Alt Text
Combine Images with the 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 : ~~~~~ ~~~~~
Tokenize the Alt Text & Calculate TF-IDF
UserText "Find the blue bus."
What PVI user said becomes the UserText.

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

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

3. Select Images
Recommendation
Recommend Images
Image 1 : blue~~~ ~bus~~~
Image 2 : ~~bus~~~ ~blue~~~
SaGol selects the images with the most similar alt text to the UserText.

1. **Search for the most similar images based** **the search term the user entered via voice.**

2. **Search for the most similar images based** **the search term the user entered via keyboard.**

3. **Confirm** **the search term the user entered.**

4. **Provide alt text via voice** when hovered over the image.

5. If the user choose the image, a screen with an **enlarged image** and **alt text is shown.**

The image shows a bicycle standing next to a house. The bicycle is white with a basket on the back and a red bow tied to the handlebars. There is a small white dog sitting next to the bicycle. The dog is looking up at the camera. The house behind the bicycle has a brown roof and white walls. (...)

Acknowledgement

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