Conversational Search: A Sketch-Based Approach

Author: Daniel Braghis
Supervisor: Dr. Haiming Liu

Problem

Currently, conversational search is a hot topic with models like ChatGPT grabbing the attention of the Natural Language Processing community. However, natural language has its limitations in expression and accessibility. Image search, for example, is still challenging through natural language alone and special needs individuals struggle to express their intent through text, limiting their ability to use the latest developments in the field. While progress has been made on each of these fronts individually through image/symbol-based conversational models and sketch-based image retrieval, there is no system that offers the option to avoid text completely in conversational search or to switch between text and images seamlessly. For this reason, conversational search, while popular in the machine learning academic community, is yet to attract the average user who still favours the classic text-based image search of the last 25 years.

Goals

My project aims to address these two issues by augmenting conversational search with user-drawn sketches to facilitate the search for users using text as input and to offer an alternative way of interacting with language models through drawn symbols alone for users challenged by text-based input. The project aims to demonstrate that systems can be designed such that both types of users can use the same feature to interact with a machine-learning model in new ways.

To achieve this, I will leverage the latest generative AI and multimodal search models available to build a chat-style interface for the user to communicate with the model using text (the baseline ChatGPT-style interaction), text and sketch, and sketch-only communication. As of now, I am planning to do a qualitative comparison between interaction methods but if time permits a user-centred study would be a better way to assess the potential of sketch-based augmentation in image search.

Scope

As the project will involve interacting with third party data and individuals from minority groups, an ERGO application will be submitted for approval prior to starting the project.

It is important to point out that this project does not aim to develop from scratch or improve on the existing models but rather show that the existing ones have reached the maturity where they can be used to improve the way users search, and simultaneously can offer the same expressive power to users that can't use text-based search. The project will concentrate on pre and post-processing of the data and the human-computer interaction aspect of the problem.

It is also outside the project's scope to develop an image-based language for the system. The machine learning model will have to understand an already existing language and be able to map it to common text-based languages.