SE 3XA3: Problem Statement
google-images-downloader

Team 201, CAS Dream Team
Sam Crawford, crawfs1, 400129435
Joshua Guinness, guinnesj, 400134735
Nicholas Mari, marin, 400132494

Table 1: Revision History

Date	Developer(s)	Change
1/21/2020	Josh	Added template file to repo
1/21/2020	Sam	Added project information
1/22/2020	Josh	Added "What problem are you trying to solve?" section
1/22/2020	Sam	Added "Why is this an important problem?" section
1/22/2020	Nick	Added "Context of the problem" section
1/23/2020	Nick	Pushed Revision 0 to GitLab
1/24/2020	Josh	Reworded some phrases for increased clarity
1/24/2020	Sam	Minor improvements
4/4/2020	4/4/2020	Rev1 Changes: Formatting/Structure of document

What problem are you trying to solve?

Researchers, companies, artists, and hobbyists occasionally have situations where they need a large number of images related to a certain word, keyword, or topic, for example, three-legged black dogs. Although these can sometimes be obtained from a pre-existing data set, there are limitations to this approach in both the size and variety of what is out there. Our solution will make it easier to obtain these images by allowing users to easily download them right to a directory on their local machine.

~~One use case for this situation involves researchers or companies training machine learning algorithms or neural networks for image recognition purposes. Obtaining a large variety of images quickly is important to producing an accurate model in a decent time frame. Two other use cases include artists making a collage, or hobbyists looking to create a game, or a portfolio boosting project using machine learning.~~

~~There are currently two ways to solve the problem. The first is to manually assemble the set of images yourself. The second is to find an open dataset online that fits your needs. The problem with the first approach is that it can be extremely time-consuming and expensive, depending on the size of the data set. Although the second approach often yields great results, finding a reasonable result is less likely to be fruitful for a specific image criteria.~~

The original open-source program to solve this problem, google-images-download, has ~~some~~ many limitations and issues. ~~Some images cannot be downloaded and are skipped, and while there are a lot of available options for filtering which images are downloaded (like colour and size), the potential exists for some more options (like specifying bit depth to work with the user's specific image processor). Providing the user the opportunity to preview and confirm each image, if desired, will allow for easier vetting of acceptable images.~~

It no longer functions properly since Google changed its HTML structure and how images are stored on the front end of the website. This means its downloads zero images. The code is also unorganized, badly documented, and not tested. Our team will re-implement this project ensuring that it actually works correctly, has a more sensible architecture and program flow, is well tested, and well documented. In addition to solving these problems listed, our team will enhance the features of the original project by adding the ability to download the images to a server, as well as blacklist certain sites.

Why is this an important problem?

This problem is important to solve because the current solutions that exist use a lot of resources, or are not very accurate. More resources mean more money and less accuracy means less quality. This makes its an issue that needs better solutions. Solving this problem by providing a quicker way to obtain a large number of images will result in products being completed quicker and more accurately. This is especially true for machine learning models where the variety of images obtained will help to ensure the model is more accurate.

What is the context of the problem you're solving?

The primary stakeholders of this problem include developers aiming to gather a large amount of sample data for training a machine learning model focused on image processing, as well as artists wishing to gather a large amount of reference material at once.

A use case for the developer stakeholder involves training machine learning algorithms or neural networks for image recognition purposes. Obtaining a large variety of images quickly is important to producing an accurate model in a decent time frame. Some potential use cases for the artist includes collage making, game visual creation, or artistic inspiration.

Some additional stakeholders include potential customers like Univerities, creative agencies, and software companies. All three of these companies contain employees that could fall into one of the two roles discussed above, and would potentially use the tool for the use cases mentioned above.

There are currently two ways to solve the problem. The first is to manually assemble the set of images yourself. The second is to find an open dataset online that fits your needs. The problem with the first approach is that it can be extremely time-consuming and expensive, depending on the size of the data set. Although the second approach often yields great results, finding a reasonable result is less likely to be fruitful for a specific image criteria.

To meet the requirements of all the stakeholders, a desktop environment is the best fit for the program, as it will allow the software to be used by a variety of users without the need for extra third party softwares and technologies.