Background

Pixel Scrapper's service:

Pixel Scrapper is a digital scrapbooking website. Digital scrapbooking is a way for people to document their memories digitally, using their pictures, journals, and also graphics (digital papers, embellishments, layouts and templates) offered by the site. Pixel Scrapper markets itself as a space for creatives to express themselves and support one another. Using a feature called the *Commons*, users can share their designs and get feedback from other users. There are also various forums for users to engage in conversations on topics such as preferred digital scrapping software. Pixel Scrapper is a subscription service. Instead of paying a fee for each desired kit (a kit is a collection of graphics), a user just has to pay one monthly or yearly fee. Pixel Scrapper considers itself to be the "Netflix of digital scrapbooking".

People involved:

The project partners are Jordan Magnuson, Marisa Larin, and Thani Magnuson. We primarily will be working with Jordan.

Vision

All graphics that are uploaded to the Pixel Scrapper site need to be tagged (a tag should be a brief description of a graphic). When this project is completed, tags will automatically be generated for uploaded graphics. The auto-tagging will initially be used by the Pixel Scrapper team of designers, which are people uploading commercial use graphics to the site. This will help speed up the process for when they do bulk uploading (uploading 100's of graphics at a time), because they won't have to manually tag each one of their graphics. Pixel Scrapper will also release autotagging to users at a later stage, far after we've completed the project. This would be helpful for users because some of them have a hard time accurately tagging their graphics. It's crucial that the tagging is accurate, so that people can go on the site and use tags to find different kinds of graphics that they want.

Two project hypotheses

Growth hypothesis:

People adopt Pixel Scrapper for a couple reasons. Firstly, Pixel Scrapper is user friendly. It is very easy for users to download graphics that they like, upload their own

graphics to the site, and communicate with other users through different forums on the site. Without even reading the instructions provided on the site, it is very clear how to do these three things (this is thanks to the well-organized layout of the site). Secondly, Pixel Scrapper is an affordable service; users have to pay a relatively small fee for unlimited access to graphics.

The completion of our project will make the user experience even easier because users will no longer have to worry about coming up with accurate tags on their own (tags will be generated automatically). So our project will expand the Pixel Scrapper user base by boosting its already existing reputation of being user friendly.

Value hypothesis:

The high quality of graphics that Pixel Scrapper offers is what makes the site beneficial to users. The users are, generally speaking, creatives that don't want standard designs for their digital scrapbooking. Pixel Scrapper offers intricate designs that can't be found with a simple google search.

Pixel Scrapper users are, generally speaking, creatives; they aren't looking for standard designs for their digital scrapbooking. The graphics on Pixel Scrapper are high quality and unique. Also, there is a variety of different kinds of graphics (abstract, concrete, etc) that the site offers (variety is necessary because different kinds of graphics are needed depending on a user's preference and the context of what their scrapbook is documenting). Users cannot find graphics that are comparable to the ones on Pixel Scrapper just anywhere online. This ultimately is the value of Pixel Scrapper.

High-level project requirements

Functional requirements:

Our system should recommend tags that accurately describe the images that users will upload. Our system must be able to recognize both concrete visual elements and more abstract visual elements. The system should improve (the tag recommendations should get more accurate) over time, by using new images that are uploaded to the system as reference.

Non-functional requirements:

The system should accomplish automated tagging with the use of a machine learning platform to train a model to recommend tags. A database of already tagged graphics will be provided to the model.

It is to be determined how well the system should function, based on more research we do and dialogue we have with project partner Jordan Magnuson.

Prioritized Project Constraints:

Time:

The timeline for our project can be broken up by term. Currently, we are in the process of generally learning about machine learning and also researching available Machine Learning platforms. During the winter term, we will move into the actual implementation of the machine learning system and train our system. If there is sufficient time left, we will create an interface for the system and Pixel Scrapper during spring term. We are using burndown charts and tracking the number of hours we spend on each sprint to estimate our weekly time allowances. It is difficult to know the extent to which time limits us because we have no previous experience with machine learning. We will be able to make a better judgment once we are a little deeper into the project.

Resources:

At the moment we do not know exactly what financial resources we have access to. We will however be constrained by financial resources to some degree because machine learning platforms are not free services. Our project partner has made it clear that we will have access to the resources we need to accomplish our goals (he will pay for resources as needed), but we do not know exactly how much this is. After we thoroughly research and select our machine learning platform, we will be able to estimate the degree to which resources limit our project.

Scope

In a perfect world our project would include the machine learning system and the interface that allows pixel scrapper to communicate with our system. We may, however, be constrained to the machine learning system alone if we run out of time. The main goal is to implement the machine learning system, so this will be our priority throughout most of the project and what we can realistically accomplish will become more clear as research progresses.

Process Flows:

We need to use a machine learning platform such as AWS or Google Cloud AutoML to train a model to recommend tags for graphics that will be uploaded to Pixel Scrapper. The recommendations will be based on a database of already tagged graphics on the site. Many of the graphics are abstract, which is why an existing model cannot be used. We have to train our model, using many different kinds of tagged graphics that users have previously uploaded to the site.

The process diagram below (figure 1) shows that the user will upload the image using the website. The website will transfer the image to the AWS image database The ML model will pull from this database to use as training data. After the user tests the image against the ML model, it will send the tags back to the website. Then the user will be able to verify those tags to make sure that they match the content. Finally after the data has been validated, it will be sent back to the training dataset.
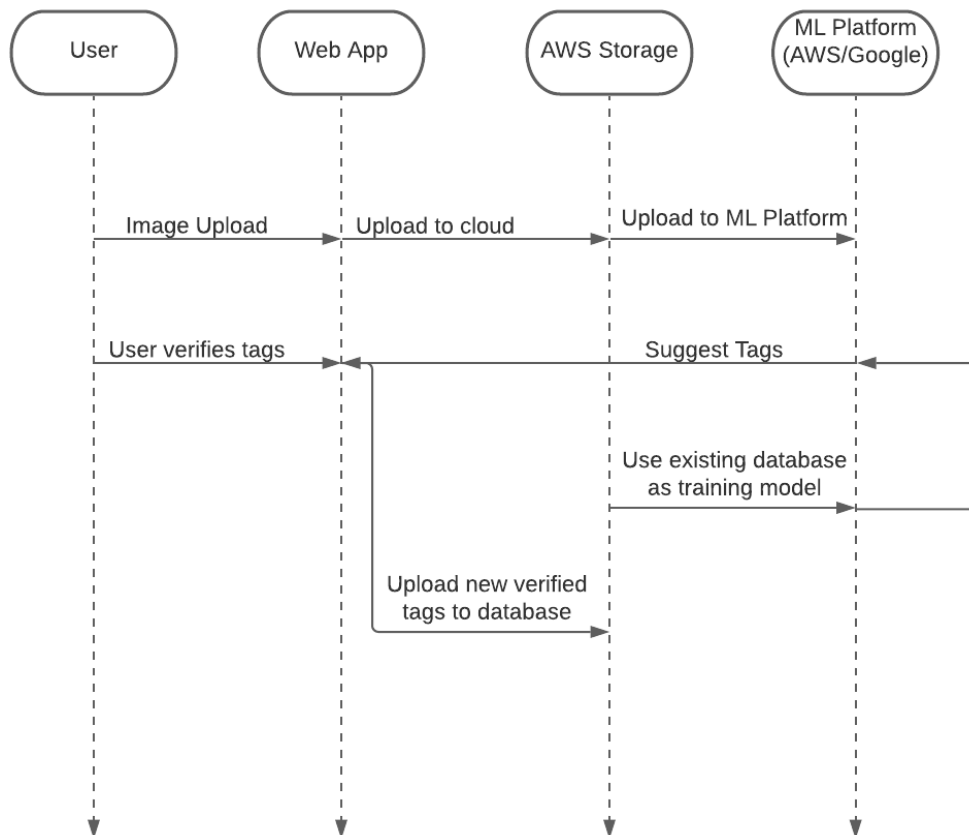

Figure 1: UML Diagram showing the flow of data and user interaction.
User Stories:

1. As a paying customer, I want an automated tagging system so that I can put relevant tags on pictures I upload.
2. As a Pixel Scrapper project partner, I want new images uploaded to the site to be used in the machine learning pipeline so that the service would get better over time.
3. As a paying customer, I want the machine learning system to be integrated with the existing site, so that it would be easy to use.
4. As a Pixel Scrapper project partner, I want the machine learning model to be efficient so that the cost of running the service does not exceed the budget.
5. As a developer I need to easily integrate the machine learning platform with the existing website so that the site does not change too significantly.
6. As a paying customer I want it to be easier to bulk upload pictures so that it is not so time consuming.
7. As a paying customer, I want the machine learning system to be accurate so that I would not have to go back and correct the tags later on.

Iteration Plan and Estimate

The first month will consist mostly of research of the different machine learning platforms and other techniques used for image classification. The team will gather resources and documentation that describes the various potential machine learning platforms we may use. With these resources we will have the necessary understanding required to begin experimenting with the platforms. This experimentation is the next step in our iteration plan. None of our group members have had any experience with machine learning, so we will focus on experimenting with different platforms and machine learning platforms in order to gain practical knowledge related to machine learning. We cannot accurately estimate how long we will spend in the experimentation phase, however could potentially last into winter term. By the end of the experimental phase we will have a final choice of machine learning platform as well as a more in depth plan for our next steps. The next sprint should begin in early to mid January. This sprint will involve the implementation of our machine learning system and the training of our machine learning model.

Once the machine learning model has been tested and verified on new digital art, the team will start working on implementing the service with the existing website. Since PHP is new for the whole team, we will have to do a bit of research to build the Drupal 7 module for the Pixel Scraper website. It is important that the tag suggestions are easy for the user to find and use, which will help them with large bulk uploads. During the entire process of this project, the team will use GitHub and other unit testing software to ensure that the code will be free of any bugs *before* being deployed to the user base.

- Scope out project (1 week)
  - Talk with project partner to gather information about the project and its requirements
  - Set expectations and commitments
  - Start planning weekly meetings and goals
- Research (1-3 weeks)
  - Explore machine learning services online
  - Gather data about pricing
  - Learn how effective image classification models are set up and deployed
- Testing (1-4 weeks)
  - Experiment with ML platforms
  - Walk through starter code to learn platform
  - Test model on real world objects
  - Continue to learn how to design an efficient way to train model
- First trial run (1-3 weeks)
  - Figure out ways to export Pixel Scrapper data
  - Build a script to generate CSV files with new data
  - Upload and test model
- Automate image upload (1-4 weeks)
  - Explore continuous learning with ML platform
  - Write scripts to automate image upload and tagging
  - Fine tune first ML model
- Deploy ML model (1-2 weeks)
  - Research ways to deploy ML model using rest API's or Python libraries
  - Use Deployed services to start PHP module