

How to Publish HITs on Amazon Mechanical Turk

Jianxiong Xiao

Abstract: This is a practical introduction to teach how to use Amazon Mechanical Turk, using the diversity experiments in my NIPS 2013 paper "Learning Deep Features for Scene Recognition using Places Database" as an example.

In general, there are 3 ways to use Turk.

1. template: the easiest way and probably most reliable way
2. command line tool: not easy, and not reliable
3. API: not easy but reliable. Useful for sophisticated control and

If you want to use 3.API, I have written a Matlab API interface that you can use
<http://vision.princeton.edu/pvt/MatlabTurkTool/>
see the demo.m file

For the places experiment, we use 1.template. The following text describes how to use template only.

Here is the the source code for our template
<http://vision.princeton.edu/projects/2014/places/12AFC.html>
right click on the webpage to see the source code.

The code starts is like this:

```
<html>  
<body>  
...  
</body>  
</html>
```

That means that we put all html, javascript, and css into the body, because we are going to copy the body content text to Amazon.

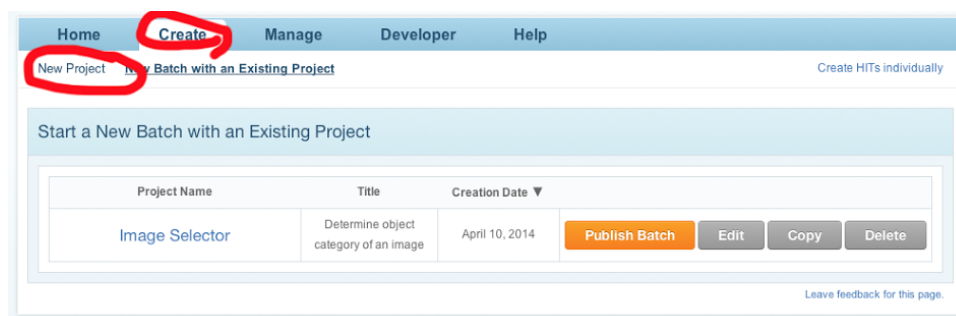
Now, login to the website

<http://requester.mturk.com>

or

<http://requestersandbox.mturk.com>

Create a new template



The screenshot shows the Amazon Mechanical Turk requester interface. The top navigation bar has links for Home, Create, Manage, Developer, and Help. The 'Create' link is circled in red. Below the navigation bar, there are links for 'New Project' and 'Batch with an Existing Project', with 'New Project' also circled in red. The main content area is titled 'Start a New Batch with an Existing Project'. It contains a table with columns for Project Name, Title, and Creation Date. The first row shows 'Image Selector' as the Project Name, 'Determine object category of an image' as the Title, and 'April 10, 2014' as the Creation Date. To the right of the table are buttons for 'Publish Batch', 'Edit', 'Copy', and 'Delete'. The 'Publish Batch' button is highlighted in orange. At the bottom right, there is a link to 'Leave feedback for this page.'

You can choose any default template provided by Amazon, because we are going to remove the content and replace it by our code completely.

Fill in the form:

Edit Project

Specify the properties that are common for all of the HIT's created using this project.

1 Enter Properties

2 Design Layout

3 Preview and Finish

Project Name: This name is not displayed to Workers.

Describe your HIT to Workers

Title

Describe the task to Workers. Be as specific as possible, e.g. "answer a survey about movies", instead of "short survey", so Workers know what to expect.

Description

Give more detail about this task. This gives Workers a bit more information before they decide to view your HIT.

Keywords

Provide keywords that will help Workers search for your HITs.

☐ This project may contain potentially explicit or offensive content, for example, nudity. [\(See details\)](#)

Setting up your HIT

Reward per assignment

Tip: Consider how long it will take a Worker to complete each task. A 30 second task that pays \$0.05 is a \$6.00 hourly wage.

Number of assignments per HIT

How many unique Workers do you want to work on each HIT?

Time allotted per assignment

Maximum time a Worker has to work on a single task. Be generous so that Workers are not rushed.

HIT expires in

Maximum time your HIT will be available to Workers on Mechanical Turk.

Auto-approve and pay Workers in

This is the amount of time you have to reject a Worker's assignment after they submit the assignment.

Advanced

Save Design Layout

Usually, we set an automatic approval time since we pay everyone. It is arguable how long to set this time.

Click on [Advanced]

Here you get to set the criteria for workers. Amazon sets it to use Masters by default, which means we will pay more money to Amazon and the jobs will be done by workers picked by Amazon as masters. Usually we don't do it. So we will remove that criteria, and Amazon will have a warning.

Advanced <

Advanced

Worker requirements <

Worker requirements:

Specify ALL the qualifications Workers must meet to work on your HITs:

remove

remove

(+) Add another criterion

(up to 5)

Only Workers who qualify to do my HITs can preview my HITs.

☐ Yes ☒ No

Now we can start to design the layout

Edit Project

Use the HTML editor below to design the layout of your HIT. This layout is common for all of the HITs created with this project. You can define variables for data that will vary from HIT to HIT ([Learn more](#)).

Project Name: Image Selector This name is not displayed to Workers.

Frame Height: 650 Height: pixels of the frame your HIT will be displayed in to Workers. Adjust the height appropriately to minimize scrolling for Workers.

Format Font U I B A Ix [Icons] Source

Is this \${question}? [ab] [ab]

Task

For each of the **N** images, answer yes or no to the above question. Only answer **Yes to the image with single, correct object**. Always answer **No to image with multiple object, object with preson, whole room, or part of object**. Here are some examples:

Usage

- Keep the right arrow [→] key (or [d] key) pressed down to move continuously from an image to the next. Release the key when you see an image for which the answer should be YES.
- Use [space] key to toggle the answer to YES. The default answer is NO.
- Use the left arrow [←] key (or [a] key) to go back to the previously seen images, if you skipped an image accidentally.

Submit

- After you answer all the images, the [submit] button will be enabled.
- If your accuracy is too low, it won't allow you to submit. You will have to improve your answers first.

By making judgments about these images, you are participating in a study being performed by Princeton Vision Group. If you have questions about this research, please contact XXX at XXX. Your participation in this research is voluntary. You may decline further participation, at any time, without adverse consequences. Your anonymity is assured; the researchers who have requested your participation will not receive any personal information about you.

body

Save Preview

First, there are two modes for editing. We only use the source code mode. Click the [Source] Button to switch to source mode, and then, delete whatever text is there. Replace the text there by our webpage without `<html><body></body></html>`. Only the content. Don't copy `<html><body></body></html>`. Now you can click and switch back to the WYSIWYG view. But don't change anything here, because the WYSIWYG editor may do something stupid that destroy our code.

Now, you need to set the height of the webpage. The HIT will be display as an iframe. So we need to set the height of the iframe. Usually we set a value that is big enough so that the turker doesn't need to be scrolling to finish the task. Otherwise, there will be two scrolling bar, one for the whole webpage, one for the iframe inside the webpage, which are very annoying and confusing. So, it is very important to set the height of the iframe.

Now you can [save] the template. Note that you can save only if it is not in the source mode (which is stupid, but Amazon decides to do this).

After we have a template, we just need to upload a CSV file for the list of images and we are done. Choose the template you want, and select [Publish Batch] to choose a .csv file.

Home Create Manage Developer Help

New Project [New Batch with an Existing Project](#) [Create HITs individually](#)

Start a New Batch with an Existing Project

Project Name	Title	Creation Date	
Image Selector	Determine object category of an image	April 10, 2014	Publish Batch Edit Copy Delete

[Leave feedback for this page.](#)

A .csv file is just a table.

In the first line, it will list the table header, which match to the \${XXXXX} in the webpage. Amazon will automatically replace the \${XXXXX} in our template by the actual content in the .csv file.

Here is an example file

http://vision.princeton.edu/projects/2014/places/input_first100.csv

For example, our template has

```

```

```

```

```

```

Then, the first line of the .csv file is

imgname0, imgname1, imgname2

The following lines are the image list. Each line is one HIT.

We can have as many columns as we need from the source code.

Usually, we generate the .csv file by Matlab, e.g.

<http://vision.princeton.edu/projects/2014/places/genInput.m>

You can open the .csv file by Excel to view the table. But don't save it. Excel may introduce some strange format and change the file so much that Amazon cannot recognize.

After we publish the HITs and the workers finish the jobs, we can download the result as a .csv file

[Manage Batches](#) > [Review Results](#)

[Review Results](#)

Select the check boxes on the left to approve or reject results. You only pay for approved results. To evaluate results offline, select Download CSV.

For additional batch information, [view batch details](#).

Image Selector 42

[Customize View](#)

[Filter Results](#)

[Upload CSV](#)

[Approve All](#)

[Download CSV](#)

0 of 27 assignments (FILTER APPLIED: only show assignments that are in 'Submitted' status)

No results match your filter

0 of 27 assignments (FILTER APPLIED: only show assignments that are in 'Submitted' status)

Then, you can use my Matlab function to read the .csv file into Matlab and do the analysis accordingly.

<http://vision.princeton.edu/code.html#readCSVfromMTurk>

For example, here is the code that I used to extract the result

http://vision.princeton.edu/projects/2014/places/extract_result.m

For the programming of HIT, it is basically just a html form.

But we want to create some hidden input to put the data.

So when the taker hits the submit button, the result will be passed on to Amazon to generate the .csv output file.

One special thing for Amazon is that we can tell if the HIT is accepted by

```
if (gup('assignmentId') == "ASSIGNMENT_ID_NOT_AVAILABLE")
```

If it is not, we don't allow the worker to work or submit.

And that is it!