Amazon Mechanical Turk

Best Practices Guide

Amazon Mechanical Turk: Best Practices Guide

Copyright © 2008 Amazon Web Services LLC or its affiliates.

This book was last updated 2008-12-23.

Guidelines for Planning, Designing and Testing Your HITs

Topics

As an Amazon Mechanical Turk (MTurk) Requester, your goal is to design effective HITs. This chapter gives you guidelines that help you plan, design, and test your HITs.

HIT Planning

Topics

This section suggests topics to consider before creating your HIT.

Break your work into HITs

In the planning stage, your first job is to define the goal of your HITs. Once you define the goal, you need to specify exactly what the Workers are going to do to help you achieve your goal. Each of these Worker tasks typically becomes one HIT. For example, let's say that you want to create a travel guide. You might break this goal into three tasks and therefore three HITs: one about finding the

best restaurants, another about identifying the best entertainment, and finally, another about the best historic places to see in the area.

Make your HITs single-purposed

A single HIT can ask for more than one answer. For example, you might present four photos for a Worker to tag in one HIT. All of the answers required for a single HIT should have the same purpose behind them. In our example, the Worker would be asked to look for the same thing in all four photos. If you find that you have multiple tasks going on in one HIT, break the tasks into multiple HITs. For example, if your HIT asks Workers to specify if there are restaurants and historic sites in photos, break the HIT into two HITs: one about tagging photos showing restaurants, another about tagging photos showing historic sites. Making your HITs single-purposed will make the answers more focused and actionable.

Consider using qualification tests

Qualification tests serve several purposes. First, if your HIT requires special knowledge, such as being fluent in a foreign language, a qualification test can filter out Workers that do not meet a minimum standard of proficiency. Second, qualification tests benefit Workers. The tests help Workers understand what they will do when working on your HITs. To facilitate that purpose, you should make sure your test accurately reflects your HIT tasks. You might, for example, ask a Worker five questions that are actual HITs. If they answer all five correctly, they pass the test. By accurately representing the tasks Workers will be asked to do, Workers can self-select in or out of working on the HIT.

If your HIT contains adult material, you are required to use the Explicit Content qualification to screen out Workers who are either underage or offended by adult content.

MTurk automatically maintains statistics on Workers, such as the percentage of assignments rejected, submitted, and abandoned. These statistics can be used as qualifiers. For a complete list of system Qualifications, see "Qualification Requirement" in the *Amazon Mechanical Turk Developer Guide*. If the qualification test is not automatically generated by MTurk, you must use the API or command line tools to create a qualification test. For more information, see the documentation.

HIT Content

Topics

This section gives tips about creating effective HIT content.

Use a descriptive title

The more informative your title, the better the Workers will know what to expect. For example, using "Movie Survey" as a title is more descriptive than using just "Survey." The title is the first part of your HIT a Worker sees. Take time to make the title concise, engaging, and descriptive.

Make your instructions simple and concise

Short sentences that each contain a single idea increases the readability of your instructions and helps you get the answers you want from your Workers. Work to streamline your instructions without losing clarity. For example, the following sentence should be broken into multiple, simpler, shorter sentences. "Your job is to go to any resource you like and find restaurant reviewers in the specified area, include their names, addresses, and any other contact information you can find, and then specify whether they work for a particular publisher and, if so, for how long."

Screen space is a premium commodity, as you will see reading on. The more concise the instructions, the less screen space they occupy. Concise, clear instructions help make your Workers more effective.

Include sample answers

Include a sample answer for each task. Examples help make your expectations clear. Here's an example: "Tag each photo with the name of the city, state and country where the photo was taken, for example, Seattle, Washington, USA."

Include in your instructions things you do not want in the answers

You might specify things that you do not want in answers if you think it clarifies what it is you do want. For example, "Do not put the name of the structure in the photo in the response, for example, Eiffel Tower, Paris, France." Or, you might identify edge cases that complicate your tasks that you should clarify. For example, "Do not include words constructed with non-characters, such as sh*t."

Use negative examples sparingly. They are valuable in clarifying expectations but if you list too many of them, they become confusing.

Set time expectations in your HIT

Include in your instructions how long you expect an assignment should take to complete. This time directly impacts a Worker's pay rate. So, be honest. By setting a Worker's expectations, you will get a higher percentage of work completed on time.

Eliminate redundant questions

Redundancy wastes the Workers' time and leads to useless answers. Make sure you ask a question only once. Do not, for example, include yes and no radio buttons that ask, "Is this picture offensive?" and also include a drop down list that asks, "What is offensive in the picture?" In this case, eliminate the radio buttons and include in the drop down list the option, "None of the above."

Do you want an opinion or a definitive answer

In general, there are two kinds of questions you can ask of Workers; you can ask for their opinion or you can ask for a definitive answer. An example of a HIT that leads to an opinion is, "Is there a sports car in the photo?" An example of a HIT that leads to a definitive answer is, "Is there a car in the photo?" If, in this case, you meant to get a definitive answer but posed the question unintentionally as

an opinion, you will wind up with unusable results. So, make sure you write your instructions so they only can be interpreted as either seeking an opinion or a definitive answer. A good rule of thumb is to avoid the use of adjectives, if possible, when you want a definitive answer.

Make clear guidelines for getting paid

When asking for a Worker's opinion, as long as the answer is relevant to the question, the Worker should get paid. When you are asking for a definitive answer, such as, is there a car in the photo, make clear that you will not give rewards for inaccurate answers.

Provide a form where Workers can comment on the HIT

Workers can provide valuable feedback about your HIT. They might cite, for example, edge cases that you have not considered. They might explain why some of your instructions are unclear. These comments can help you provide mid-course corrections to your HITs and help you plan better for future HITs. So, make sure to leave a place for Workers to enter comments.

Do not ask Workers to sign up for email offers or to purchase anything

Amazon Mechanical Turk is designed to be an innovative tool to get work done. It is not designed to be an advertising medium. Do not use MTurk to advertise in any way.

Do not ask Workers for personal information

Asking Workers for opinions is fine but it is not acceptable to ask them for personally-identifying information, such as their name, address, phone number, or social security number.

Do not violate the Mechanical Turk Participation Agreement

Read the Mechanical Turk Participation Agreement to make sure that your HITs do not violate it.

Make clear if your HITs contain adult material

Adult material is offensive to some Workers and is illegal to present to Workers under the age of 18. If your HITs contain adult material, in the HIT title you must include, "May include explicit, adult content." Also, you must use the Explicit Content qualification for Workers and make the qualification a requirement. In this way, if a Worker is not qualified to work on adult content, they will not be able to view the HITs.

HIT Layout

Topics

This section provides tips for effective layout of HIT content.

Number your instructions

A numbered list makes it easy for Workers to see each instruction clearly and to refer back to an instruction quickly. In contrast, a paragraph of instructions requires that the Worker dig out of the paragraph each instruction.

Numbering your instructions quickly shows your Workers how many steps they must complete for each HIT and shows them the exact order in which to complete the steps.

Avoid scrolling as much as possible

Scrolling reduces Worker efficiency and effectively reduces their pay rate. Find a way to make your HIT fit on one page only. One way of doing this is making long instructions collapsible and collapsed by default. Making Workers close collapsible instructions in every HIT reduces their efficiency. Making workers more efficient effectively increases their pay rate and, thus, their motivation to work on your HITs.

Avoid using drop down menus for answers

Opening drop down menus to enter answers requires extra clicks and thus reduces a Worker's pay rate. To increase Worker efficiency, use radio buttons instead of drop down menus to present a choice of answers.

Minimize the need to visit other websites and view multiple windows

Making Workers click on one window to view HIT instructions or data and then another window to enter the HIT data reduces Worker efficiency and their pay rate. Avoid the need to open multiple windows by breaking the work into multiple HITs that each have a simpler set of instructions and a simpler layout that fits onto one page.

Don't open new web pages in the HIT window

If, for some reason, a Worker must open a new window, open it in a new browser instance so that the Worker can continue working on the HIT in the current window while referring to the data in the new window.

Place answer fields next to images

By placing answer fields adjacent to the images they're associated with, you reduce the possibility of Workers entering answers in incorrect comment fields.

Place the response areas at the bottom of the HIT near the Submit button

Putting the response areas and Submit button in close proximity reduces the amount of mousing required and therefore makes a Worker's job easier.

Provide separate entries for each distinct answer

Each Worker answer for a HIT is stored in a separate column when the .csv file is displayed in a spreadsheet. By breaking Worker's answers across multiple columns you can more easily access results. For example, if you have a Worker put in one entry field a restaurant reviewer's first name, last name, e-mail address, and employer, breaking out that information into reports is far more difficult than if you provide separate answer fields for each piece of information.

HIT Properties

Topics

This section provides tips about setting HIT properties.

Give Workers enough time to do the work

One HIT property is `Time Allotted`. This quantity is not the time you think it should take to do the work; it is the time a Worker is allowed to hold onto a HIT before it is taken away so another Worker can work on it. So, for example, if you set `Time Alloted` to one minute, a Worker would have to complete all of his or her assignments in one minute!

Do you want more than one Worker to work on a HIT

You can specify the number of assignments per HIT, that is, the number of Workers who work on the same HIT. You use one assignment per HIT when you want one Worker's answer to a task, such as, is there a car in the photo. In this example, there is only one right answer and the simplicity of the question does not suggest a need for multiple responses.

You use multiple assignments for several reasons. First, you might like to conduct a survey. In this case, one question might generate many different responses and you're interested in the diversity. Second, you might like to find if there is a commonality of opinion. In this case, you would be interested in finding out what a majority of the Workers consider to be the correct answer. An example of a survey question is, "Where is your favorite place to vacation?" An example of a majority vote question is, "Who do you think will win the presidential election? Obama or McCain?"

Set your rewards commensurate with the work required

MTurk groups HITs together based on HIT types. All HITs within one group of HITs that pay the same should require approximately the same amount of work. For example, if one set of HITs pays one cent to answer one question, it is ill advised to have a HIT in that group pay one cent to answer five questions.

Setting the price for your work requires looking at HITs similar to yours that are posted on MTurk. MTurk provides many standard HIT templates. You should look at the HITs that belong to the same

HIT template as yours. Posting too low a reward might disincline Workers from working on your HITs. You can change the reward for posted HITs but this makes the process take longer.

HIT Testing

Topics

- Make sure your instructions are clear and concise (p. 7)
- Check your links (p. 7)
- Use the MTurk sandbox to test your HITs (p. 7)
- Do a trial run (p. 7)

Testing your HITs before publishing them is the best way to see if they will return the kinds of answers you're looking for. Eliminating mistakes by testing your HITs is an effective way to save you money. This section describes how to test your HITs.

Make sure your instructions are clear and concise

Before publishing your HITs, give your instructions, without explanation, to a fellow worker and make them follow them. See if the worker proceeds in the way you intended. Ambiguous instructions often lead to erroneous answers, which compromise your results. For example, asking a Worker to say whether there is a sports car in a photo can lead to Workers making judgement calls, whereas asking if there is a car in the photo is clearer. Make sure your instructions are clear, specific, and unambiguous.

Check your links

If you have links to pictures or videos in your HITs, make sure the links work. The best way to check the links is to actually complete some of your HITs using the MTurk sandbox.

Use the MTurk sandbox to test your HITs

Testing your HIT yourself end-to-end is an important step in qualifying your HITs. Use the MTurk sandbox to test the workflow of your HIT, complete some assignments, and view the results. Consider whether the results enable you to achieve the overall goal of your project.

Do a trial run

Completing HITs yourself is good but having real Workers complete HITs can reveal subtle errors. Before loading up a .csv file with 10,000 HITs, do a trial run by publishing 100 HITs. Seeing the results of the trial run can help you revise your HITs and thereby be a real cost-saver.