




Crowdsourcing Annotations

School of Information Studies
Syracuse University

Crowdsourcing the Annotations

A cheap and fast approach for obtaining training examples

Amazon Mechanical Turk (AMT)



amazonmturk.com						
HITS	Dashboard	Qualifications	Search All HITS			
All HITS	Your HITS Queue					
HIT Groups (1-20 of 1131)						
			Show Details	Hide Details	Items Per Page: 20	
Requester	Title	HITS	Reward	Created	Actions	
KronoPin	Find the Website Address for a Consumer Product	37,858	\$0.03	3h ago	Preview	Qualify
ScoutIt	Classify Receipt	34,039	\$0.03	2m ago	Preview	Qualify
Corey	Find the Contact Information for the College Instructors	22,642	\$0.05	6h ago	Preview	Qualify
Spyridon Systems	Two tasks, one ask!	14,744	\$0.02	1d ago	Preview	Accept & Work
Alex Moss	Check Quantity Matching Data	10,074	\$0.10	5d ago	Preview	Qualify
Marketing Research	*商品名のカテゴリ入力*	8,630	\$0.04	2d ago	Preview	Qualify
Raj Singh	Find the Website Address for Companies	7,231	\$0.01	1d ago	Preview	Qualify
Alia Kaussar	Find company information for a credit card transaction description	7,189	\$0.15	8d ago	Preview	Accept & Work
Yusuke Watanabe	Similar Question Pair Labeling	6,090	\$0.01	2d ago	Preview	Accept & Work

To Be an AMT Worker

Sign up (takes up to 48 hours to be approved)

Search for tasks and accomplish

To Be an AMT Requester

Trial task: prepare ~\$5 and use your HW1 data

- Ask five AMT workers to label each comment
- Calculate inter-coder agreement
 - Among AMT workers (average Kappa)
 - Between each worker and your ground truth
 - Do the workers perform equally well? Do they share similar marginal distribution?

Step 1: Create Project

Home Create Manage Developer Help

New Project New Batch with an Existing Project Create HITs individually

sentiment toward target data breach — Judge the sentiment expressed by the following item toward: the Target company

Sentiment Project

Create Project

Provide Instructions

Upload Data File

Preview

Checkout

Project

What would you like to call your project?

sentiment toward target data bre:

Sentiment Question

What do you want Workers to judge attitude towards?
(see examples)

the Target company

How the Workers will see your question:

Judge the sentiment expressed by the following item towards: the Target company
(Placeholder for your items to be judged. You will be asked to upload them at a later step)

Strongly negative Negative Neutral Positive Strongly positive

Next »

Step 2: Provide Instructions

The screenshot shows a web interface for creating a project. On the left is a vertical sidebar with five buttons: 'Create Project', 'Provide Instructions' (highlighted with an orange arrow), 'Upload Data File', 'Preview', and 'Checkout'. The main area is titled 'Instructions for Workers' and contains the following text: 'For each point on the scale, define the criteria Workers should follow when selecting it. We have provided sample instructions that you can customize below.' Below this are five rows, each with a label and a text box for instructions:

Label	Instructions
Strongly positive:	Select this if the item embodies emotion that was extremely happy or excited toward the topic. For example, "Their customer service is the best that I've seen!!!!"
Positive:	Select this if the item embodies emotion that was generally happy or satisfied, but the emotion wasn't extreme. For example, "Sure I'll shop there again."
Neutral:	Select this if the item does not embody much of positive or negative emotion toward the topic. For example, "Yeah, I guess it's ok." or "Is their customer service open 24x7?"
Negative:	Select this if the item embodies emotion that is perceived to be angry or upsetting toward the topic, but not to the extreme. For example, "I don't know if I'll shop there again because I don't trust them."
Strongly negative:	Select this if the item embodies negative emotion toward the topic that can be perceived as extreme. For example, "These guys are terrific... NOTTT!!!!!!" or "I will NEVER shop there again!!!"

Below the instructions is a section titled 'Number of Responses' with the text: 'How many Workers do you want to rate sentiment for each item? (Help me choose)'. To the right of this text is a dropdown menu showing the number '5'. At the bottom right of the main area is an orange button labeled 'Next »'.

Step 3: Upload Data File

The screenshot shows the Amazon Mechanical Turk Requester interface. The top navigation bar includes 'Home', 'Create', 'Manage', 'Developer', and 'Help'. Below this, there are links for 'New Project', 'New Batch with an Existing Project', and 'Create HITs individually'. The main heading is 'sentiment toward target data breach — Judge the sentiment expressed by the following item toward: the Target company'. A sidebar on the left lists steps: 'Sentiment Project', 'Create Project', 'Provide Instructions', 'Upload Data File' (highlighted), 'Preview', and 'Checkout'. The main content area is titled 'Upload Data File - Choose Fields'. It contains instructions: 'Each column from your .csv file is shown as a field below. Tell us which fields you want to show Workers and whether the fields contain text, a link to an image, or a link to a website.' and 'Optionally, you can add a label to appear next to your data. For instance, if the field is a "tweet," you might want to add a label of "Tweet."'.

Field	Show Workers?	Type of Data	Label
text	<input checked="" type="checkbox"/>	Text	public comments

If this doesn't look right, you can [upload a different file](#).

[Next »](#)

Step 4: Preview

Sentiment Project

Create Project

Provide Instructions

Upload Data File

Preview

Checkout

How does it work?

Preview of Work Items

This is what Workers will see.

Instructions

Pick the best sentiment based on the following criterion.

Strongly positive	Select this if the item embodies emotion that was extremely happy or excited toward the topic. For example, "Their customer service is the best that I've seen!!!!"
Positive	Select this if the item embodies emotion that was generally happy or satisfied, but the emotion wasn't extreme. For example, "Sure I'll shop there again."
Neutral	Select this if the item does not embody much of positive or

Judge the sentiment expressed by the following item toward: the Target company

public comments: I love it that they remove negative but true comments. Way to go Target. You've let everyone down.

Strongly negative

Negative

Neutral

Positive

Strongly positive

You must ACCEPT the HIT before you can submit the results.

Preview 1 of the first 5 items

« Go back and edit

Continue »

SCHOOL OF INFORMATION STUDIES | SYRACUSE UNIVERSITY

8

School of Information Studies
Syracuse University

Step 5: Checkout

amazonmechanical turk REQUESTER

Home Create Manage Developer Help

New Project New Batch with an Existing Project Create HITs individually

sentiment toward target data breach — Judge the sentiment expressed by the following item toward: the Target company

How does it work?

Sentiment Project

- Create Project
- Provide Instructions
- Upload Data File
- Preview
- Checkout**

Checkout

Number of Items		30
Number of Workers per Item	X	5
Number of Worker Submissions	=	150
Reward per Submission (details)	X	\$0.020
Total Worker Rewards	=	\$3.000
Total Mechanical Turk fees (details)	+	\$1.350
Total cost	=	\$4.350

Your account does not have sufficient funds to publish these items. Click [here](#) to fund your account

Your available balance	\$0.00	Your balance after work is completed	\$-4.35
------------------------	---------------	--------------------------------------	----------------

Note: To accelerate Worker engagement, Worker submissions will be automatically approved in one hour.
If you prefer to customize the auto approval time, please [create a custom project](#).

Issues About Using AMT

Understand the AMT worker population

- U.S., India, and many other countries
- Language proficiency
- “Master” workers vs. others

Issues About Using AMT

How much do you pay?

- Not too high (attract spammers)
- Not too low (turn away workers and get bad reputations)

Check out customary payment level in similar tasks

The ethical issues around minimum wage

Issues About Using AMT

Strategies for spam control

- Hire “master” workers only.
- Insert “test cases” in your task; if a worker performed poorly on the test cases, you can refuse to pay.
- Refuse to pay is a serious matter. It can ruin your reputation in the AMT worker community.
- Motivate the workers with interesting design of your tasks.

How Good Are the Turkers?

Cheap and fast? But is it good? Evaluating non-expert for natural language tasks

Snow, R., O'Connor, B., Jurafsky, D., & Ng, A. Y. (2008, October). Cheap and fast—but is it good? Evaluating non-expert annotations for natural language tasks. *In Proceedings of EMNLP 2008* (pp. 254–263).