Creating and running experiments on Mechanical Turk and CrowdFlower

Crowdsourcing works for tasks that are:

Decomposable into simpler tasks that can be joined together

and/or

Parallelizable into small, quickly completed chunks

Crowdsourcing works for tasks that are:

Balanced so that each task contains the same amount of work

  - Don't have tons of work in one assignment but not another

  - Don't ask workers to annotate something occurs in the data <<10% of the time

Crowdsourcing works for tasks that are:

Well-suited to **quality control** (e.g. some data has known correct answers)

but

Robust to some amount of noise/errors (e.g. the end goal is to train a statistical model)

Preset Templates

Start a New Project

- **Categorization**
- Data Collection
- Moderation of an Image
- Sentiment
- Survey
- Survey Link
- Tagging of an Image
- Transcription from A/V
- Transcription from an Image
- Writing

amazon mechanical turk™
Artificial Artificial Intelligence

Business URL Collection A

Content Moderation

Search Relevance A

Basic Demographic Survey

Sentiment Analysis

CrowdFlower

Paraphrase judgement
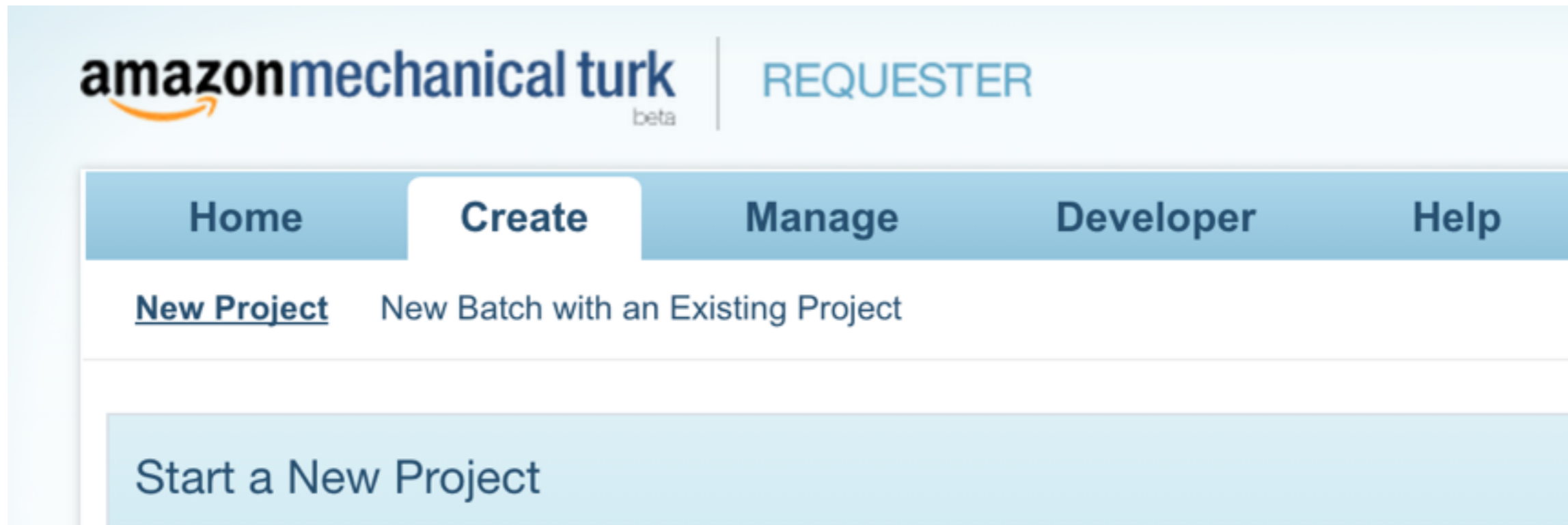
(Running example for today…)

17. occasionally , students spend the night in a youth hostel , where they share stories with peer travelers from around the world .

☐ now and then
☐ just sometimes
☐ only occasionally
☐ now and again
☐ every now and then
☐ some times
☐ once in a while
☐ for once
☐ here and there
☐ None of these paraphrases are good

Demos!

1. Mechanical Turk web interface

2. Mechanical Turk API, via Python
   (http://boto.readthedocs.org/en/latest/ref/mturk.html)

3. CrowdFlower web interface

4. CrowdFlower API, via Python
   (https://github.com/peoplepattern/crowdflower)

Development sites

https://requestersandbox.mturk.com

Development sites

Internal Contributors

On   **Share your job with team members**

If you want your team members to work on this job, we will provide you with a link to share with your teammates

Link to share:

https://tasks.crowdflower.com/channels/cf_internal/jobs/731405/work?

secret=7WhLuOulaFoykCgNx8O7pXwLkzCiPNOVKwmnM3KaJKNH

External Contributors

On   **Share your job with CrowdFlower Contributors**

When enabled, we'll allow our online contributors to work on your job

No   **Job may be offered to Unleveled contributors**

No   **Job contains explicit content**

https://make.crowdflower.com/jobs/JOB_ID/settings/channels

Show Contributor Channel Table

Writing Instructions

Writing Instructions

Please check **all** paraphrase which could replace the blue word in the context shown, using your best judgement. The paraphrase should **retain the meaning of the original phrase** and make sense in context. Most sentences will have at least one good paraphrase, but some will not. If none of the paraphrases listed fit the context shown, **select NONE**.

I have a question about these hits. When the substitution makes the sentence end up saying something like, "The small girl girl was jumping rope," or other repetition of the same word, what should that be categorized as?

I'm a little confused about when one word is a variant of another. On the linked page of expanded definitions it says "other relation" should be used when it is two forms of the same word. It uses the example of ancestor and ancestry. But, on the instructions on each hit page the first example uses snow and snowy and says to put equivalent. Could you please help me understand how to classify this type?

Writing Instructions

Please check **all** paraphrase which could replace the blue word in the context shown, using your best judgement. The paraphrase should **retain the meaning of the original phrase** and make sense in context. Most sentences will have at least one good paraphrase, but some will not. If none of the paraphrases listed fit the context shown, **select NONE**.

It is okay if the paraphrased sentence is contains small grammatical errors, as long as they do not drastically alter the meaning. E.g. the following is an **okay paraphrase**:

The **boy** is running.

The **boys** is running.

but the following is a **bad paraphrase**:

I am running.

You are running.

It is okay if the paraphrased sentence contains more/less detail than the original sentence, as long as it does not drastically alter the meaning. E.g. the following is an **okay paraphrase**:

The **person** is running.

The **boy** is running.

but the following is a **bad paraphrase**:

The cells undergo **development**.

The cells undergo **economic development**.

The paraphrase should make sense in the particular sentence shown. E.g. the following is a **bad paraphrase**, even though *center* might be a good paraphrase for *heart* in some other contexts.

The **heart** is an organ.

The **center** is an organ.

These HITs will be **quality controlled**. If you are unsure whether or not you are doing them well, we recommend you do a small number and wait for feedback before continuing.

Thank you in advance!

Writing Instructions

✔ If your task can't be expressed in one paragraph + bullets, then probably break it up

✔ Include an example of a right answer and a wrong answer

✔ Clarify what you expect if the HIT is not doable (e.g. because of missing data or other problems)

Approving / rejecting

amazon mechanical turk™
Artificial Artificial Intelligence

Approve	Reject

HIT ID ▲	Worker ID	Lifetime Approval Rate	Input.HIT Num	Input.Instance Id1	Input.Doc Id1	Input.Sent Id1	Input.Sentence1	Input.Two Before1
302U8RURJZ1S2C6CV3ZQM0GTGEONVJ	A2YTOX7SEDJ1TC	100% (23/23)	9	238	P155842	16	IMPRESSION: Developmental anomaly of the right ...	No definite abnormality is seen in the visualiz...
302U8RURJZ1S2C6CV3ZQM0GTGEONVJ	A312KFQ78WB3A4	99% (67/68)	9	238	P155842	16	IMPRESSION: Developmental anomaly of the right ...	No definite abnormality is seen in the visualiz...

Approving / rejecting

amazon mechanical turk™
Artificial Artificial Intelligence

- <u>YOU</u> have all the power

- After reviewing a Worker's submission, you can

    - approve it (worker gets paid)

    - reject it (worker does not get paid)

    - block the worker (worker is banned from working on any of your HITs, and may get blacklisted from MTurk)

Approving / rejecting

Your job requires more funds than what was previously allocated ⓘ

Previous estimate: $10.09

New estimate: $47.76
373.32% above expected

Crowd costs	$39.80
Transaction Fee (20%)	$7.96

$37.67 due now

Approving / rejecting

CrowdFlower

Your ... d (i)

- CrowdFlower decides who to approve/reject

- Downloaded final results come with confidence/trust scores

- Prepay for work, based on CrowdFlower's projected cost estimates

  - Good gold data/Quiz Mode helps prevent problems