

Microtask Crowdsourcing: Fundamentals (Part I)

Maribel Acosta, Amrapali Zaveri



 Maastricht University

Institute of Data Science

Microtask Crowdsourcing

Work is broken down into smaller (“micro”) pieces that can be solved independently.

Outline

1. **Dimensions of Microtask Crowdsourcing**
2. Microtasks: Terminology
3. Creating Microtasks
4. Microtask Platforms: Figure Eight and CrowdFlower
5. Summary

Dimensions of Microtask Crowdsourcing



What

Tasks based on human skills not easily solvable by machines:

- Visual recognition
- Language understanding
- Knowledge acquisition
- Human communication
- Many others



Who

- Open call or crowd accessible through a platform
- Call may target specific skills and expertise
- The crowd is composed of a large pool of (unknown) people



How

- Tasks are broken down into smaller units undertaken in parallel by different people
- Partial answers are consolidated and aggregated into a complete solution
- Coordination is required to handle cases with more complex workflows

Example:

Citizen Science via Microtasks

What is crowdsourced

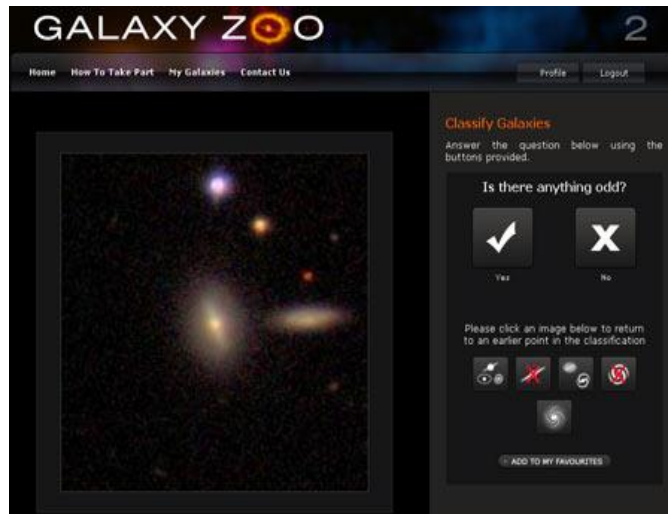
- Object recognition, labeling, categorization in media content

Who is the crowd

- Anyone

How is the task crowdsourced

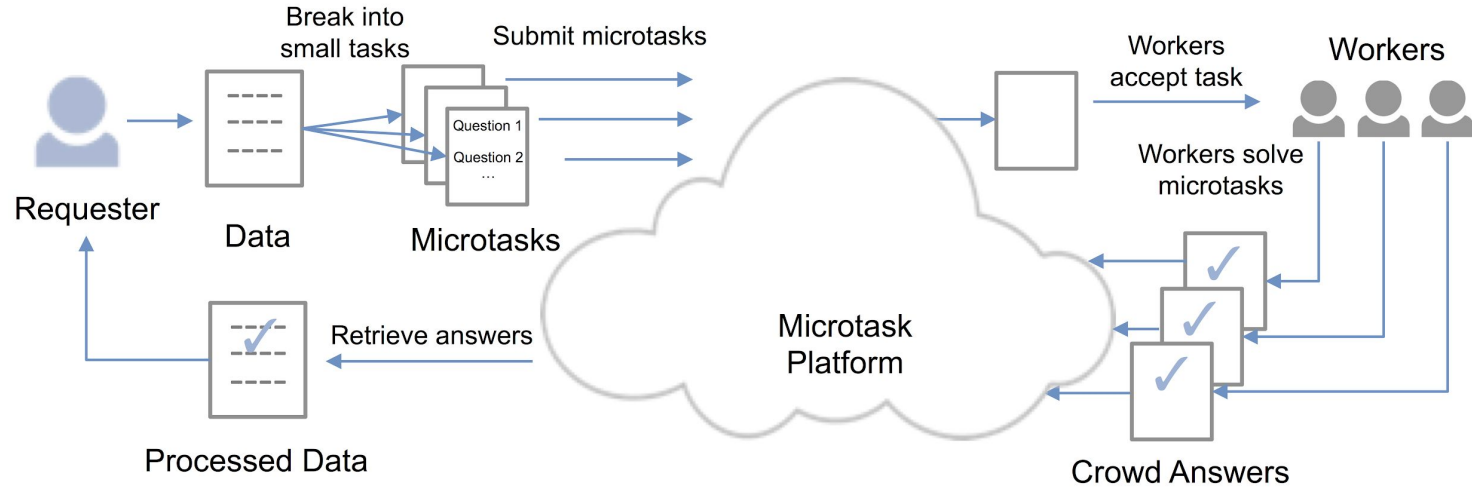
- **Highly parallelizable** tasks
- Every item is handled by **multiple annotators**
- Every annotator provides an answer
- Consolidated answers solve problems



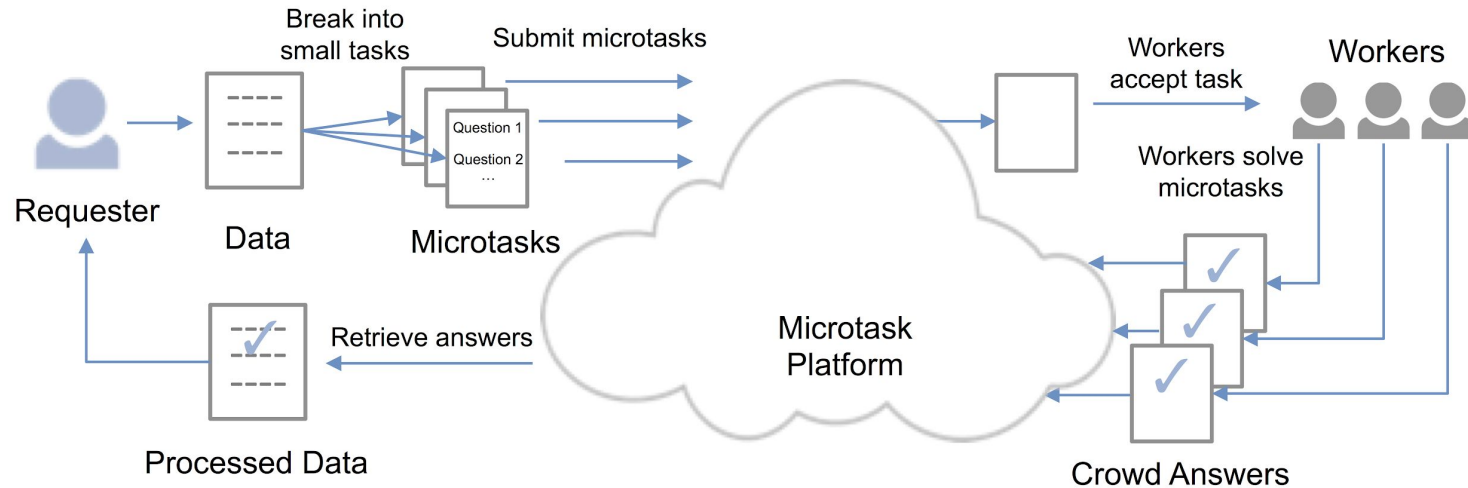
Outline

1. Dimensions of Microtask Crowdsourcing
2. **Microtasks: Terminology**
3. Creating Microtasks
4. Microtask Platforms: Figure Eight and CrowdFlower
5. Summary

Microtask Crowdsourcing Process

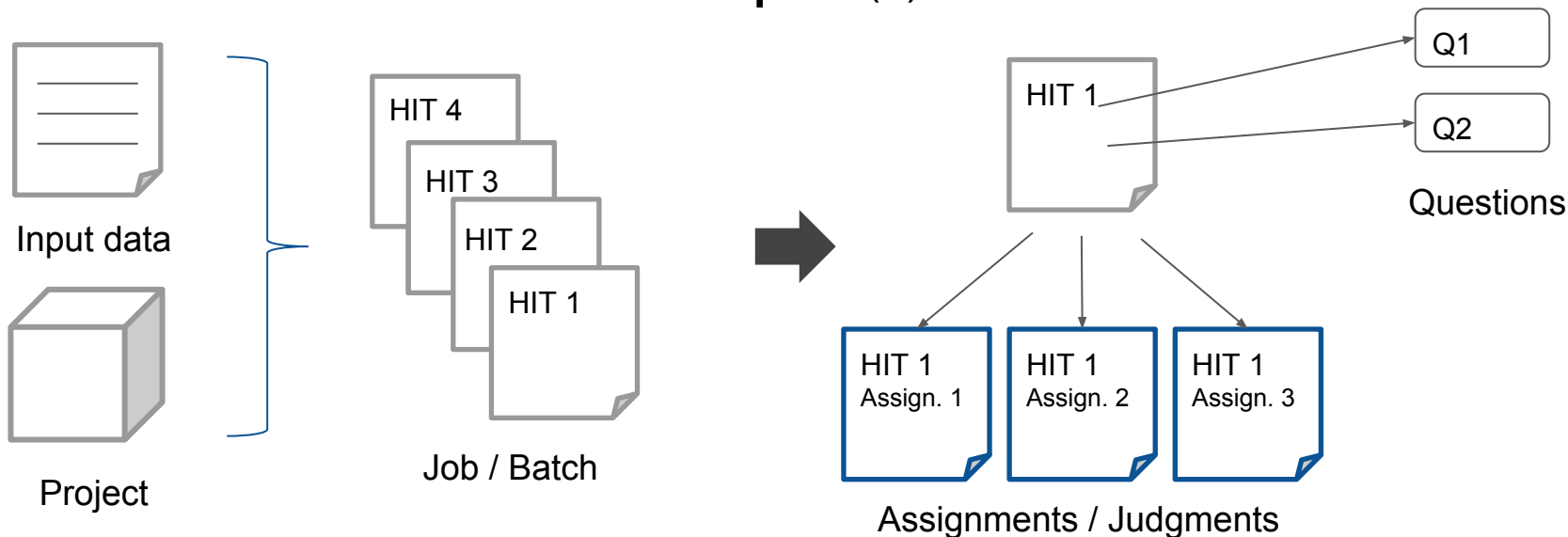


Microtask Crowdsourcing Process



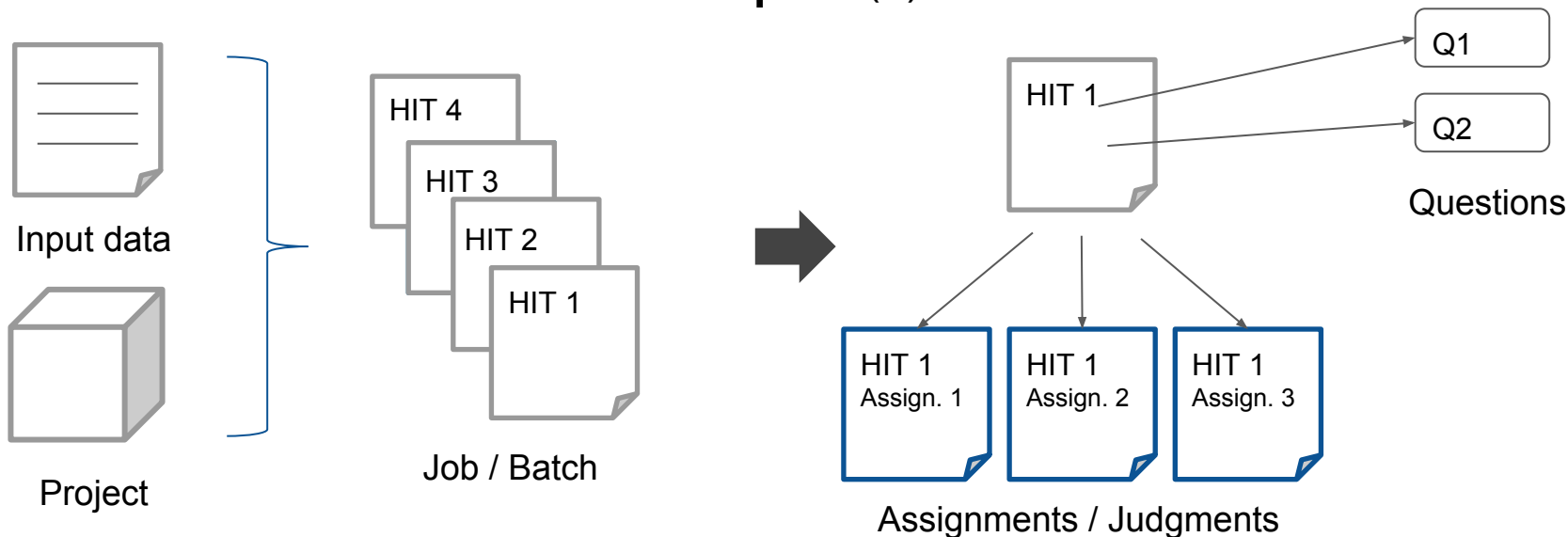
- **Requester:** Creates and submits microtasks to the platform.
- **Microtask or Human Intelligence Task (HIT):** Work unit.
- **Worker:** Person who solves tasks.

Microtasks: Basic Concepts (1)



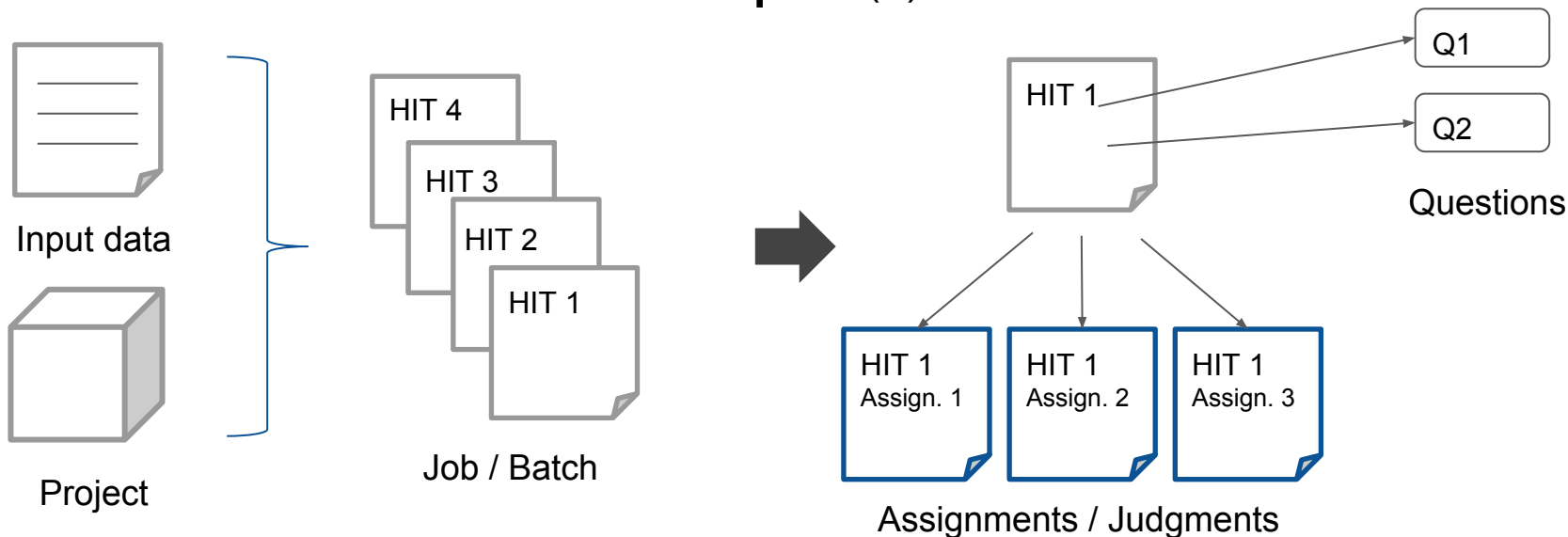
- **Input data:** Typically, a CSV or TSV file where
 - Each column corresponds to a variable
 - Each row is a question
 - Each file corresponds to a job / batch

Microtasks: Basic Concepts (2)



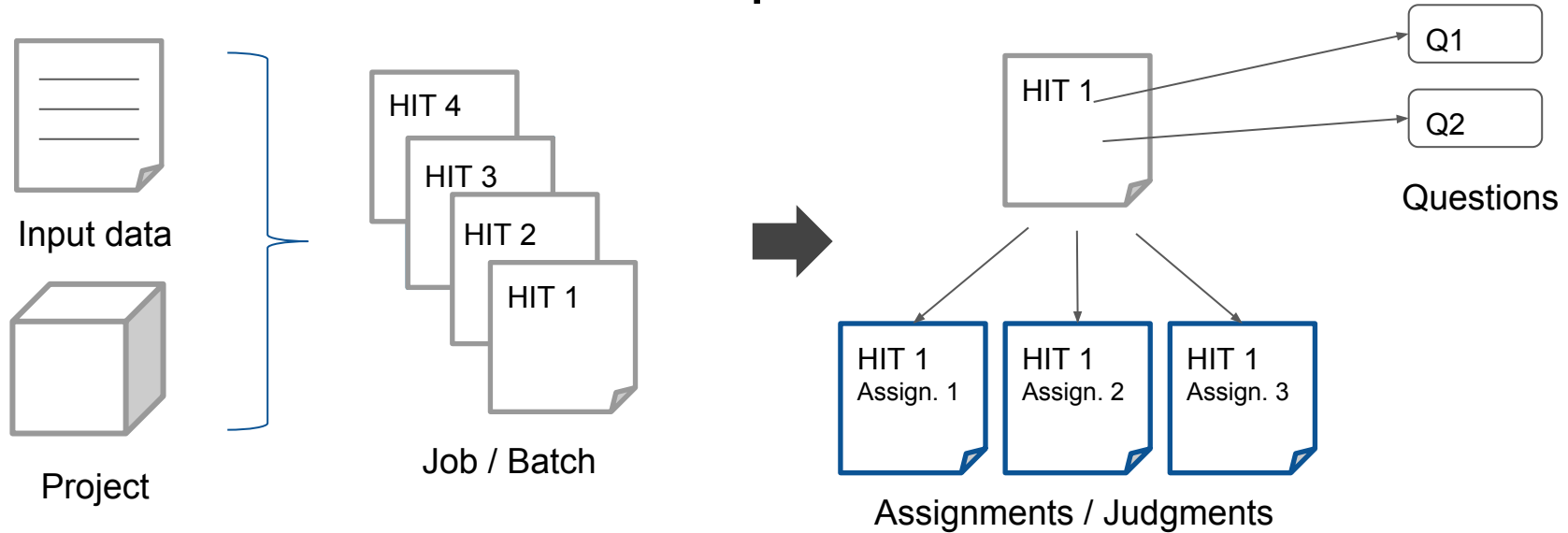
- **Project:** Includes the template for the microtask interface and further metadata
 - The template for the interface can reference columns (variables) in the input data
 - All the microtasks generated within a project have the same properties (payment, quality control, duration, etc.)

Microtasks: Basic Concepts (3)



- **Job / Batch:** Group of microtasks or HITs created in a project based on the input data
- **Questions:** Individual questions inside of a microtask.
- **Assignments / Judgments:** Answer from a worker to a microtask. The number of assignments determine the redundancy per microtask.

Microtasks: Basic Concepts (4)



Total cost = No. of HITs in Job x No. of assignments x (Reward per HIT + Platform Fee)

Example: Input Data, HITs, and Questions

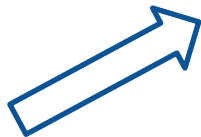
- Problem: Tagging (describing) images

image_url
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9008.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_8982.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9016.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9017.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9018.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9020.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9023.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_8984.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9026.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9027.JPG

Input data:
900 images

Create tags for image
{{image_url}}

Project



HIT:

Create tags for image
IMG_9008.JPG



HIT:

Create tags for image
IMG_8982.JPG



...

No. of HITs = 900

Example: Input Data, HITs, and Questions

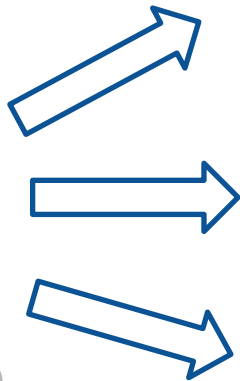
- Problem: Tagging (describing) images

image_url
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9008.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_8982.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9016.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9017.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9018.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9020.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9023.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_8984.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9026.JPG
https://s3.amazonaws.com/pichost/segmentor/us101/IMG_9027.JPG

Input data:
900 images

Create tags for image
{{image_url}}

Project



HIT:

Create tags for image
IMG_9008.JPG,
IMG_8982.JPG,
IMG_9016.JPG

HIT:

Create tags for image
IMG_9017.JPG,
IMG_9018.JPG,
IMG_9020.JPG

...

Several questions in a HIT!

No. of HITs = 300

Instead of 900 HITs

Properties of a Project

- **General information:** includes title and (short) description of the HIT.
Depending on the platform, requester can specify keywords which are used by workers for searching HITs.
- **Microtask duration time:** time allotted to solve a HIT.
- **Microtask life time:** how long will the HIT be available on the platform.
- **Redundancy:** number of different persons that assess the exact same HIT.
- **Reward:** payment for correctly solving an assignment.
- **Quality control:** settings to improve the quality of the results (more about this topic in the next lecture)

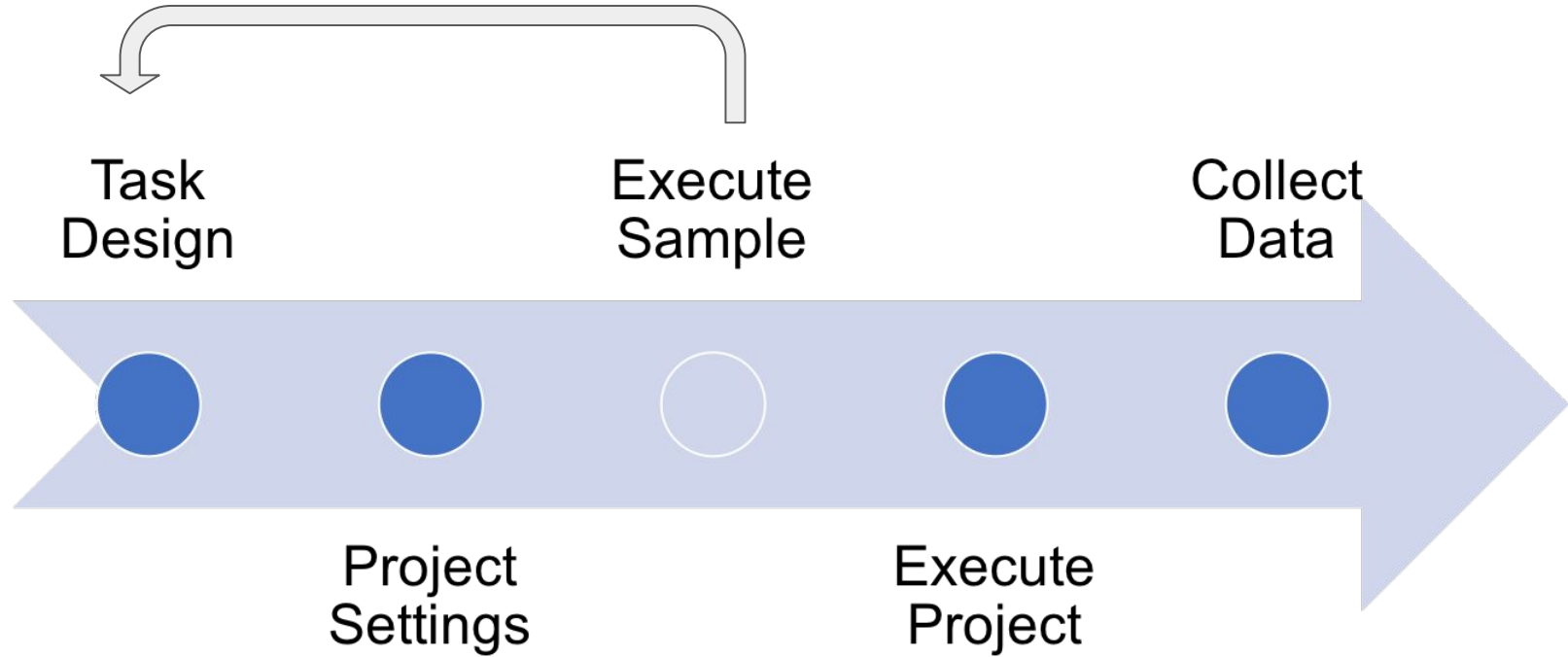
Summary of Terms

Figure Eight (CrowdFlower)	Amazon Mechanical Turk	Meaning
Requester / Customer	Requester	Person who submits a microtask.
Page	Human Intelligence Task (HIT)	Microtask to be solved.
Row	Question	A question in a microtask.
Job	Batch	Group of microtasks.
Contributor	Worker / Turker	Person who solves the microtask.
Judgements per Row	Assignments	Number of distinct persons that complete the same microtask (redundancy).
Judgements	Responses	Answers from the crowd for a microtask.
Payment	Reward	Payment granted to a worker for completing a microtask.
Quality Control	Qualification Types	Quality control mechanisms.

Outline

1. Dimensions of Microtask Crowdsourcing
2. Microtasks: Terminology
- 3. Creating Microtasks**
4. Microtask Platforms: Figure Eight and CrowdFlower
5. Summary

Creating Microtasks: Typical Workflow



Task Design

- The design on the task impacts on:

The quality of the crowd answers

The time performance of the crowd

- Ask the right questions
 - Provide clear instructions
 - Show examples
 - Workers may not be experts: Do not assume the same understanding of terms. Simplify the wording.
- Reduce ambiguity**

Example: Bad Task Design

Help us describe How-To Videos! Earn \$2.50 bonus for every 25 videos entered!

Watch a how-to video, and write a keyword-friendly synopsis describing the video.

1. Click on the link to watch the **Film & Theater** how-to video ==> [332492 Get a 35mm film look with a depth of field adapter](#)
2. Write a description of the video linked in 4 or more sentences.
3. Be detailed in your description. Describe how the procedure is done.
4. Description should be at least 100 words.
5. Description should be fewer than 2000 characters.
6. Use the character and word counters below to help you stay within the limits.
7. You must complete **25 video descriptions** in order to earn the \$2.50 bonus. Bonuses are distributed after HITs have been completed. The more HITs completed and approved, the more you will earn.
8. It is **not necessary** to repeat the headline in your entry. It will **NOT** count toward your word count.
9. Do **NOT** describe the following: the format, where the video comes from, or how long the video is. This information is **IRRELEVANT**.
10. Do **NOT** describe the video in the following manner: "She turns around to face the camera. Then she faces left." Follow the examples below.

Current Word Count: 0 Current Character Count: 0 / 2000

Criteria for **REJECTION**:


1. Entries with obvious and multiple spelling or grammatical errors will be rejected.
2. Entries with fewer than 100 words will be automatically rejected.
3. Text copied from the web or other places will be rejected. Multiple plagiarized answers will lead to being **BLOCKED**. You may use a quotation, but the majority of your content must be **ORIGINAL**.
4. Incomplete and blank answers will be rejected. Multiple blank answers will result in being **blocked**.
5. Tasks submitted without descriptions will be rejected.
6. Tasks submitted with inaccurate descriptions will be rejected as well.
7. Do **NOT** add any personal opinions. Entries with personal opinions or reviews will be automatically **REJECTED**.
8. If you notify us that a link is broken, we appreciate it but will not be able to accept the submission. The notification will result in rejection.
9. Entries that transcribe the video will be **REJECTED**.

- Asking too much
- Task is not engaging

Example: Good Task Design

Task

Please evaluate the relevance of the following document for the query **milton keynes**.



The screenshot shows a Bing search engine interface. At the top, there are navigation links for Web, Images, Videos, Shopping, News, Maps, More, MSN, and Hotmail. The search bar contains the text 'milton keynes'. Below the search bar, there are links for 'Milton Keynes Map', 'Milton Keynes Restaurants', and 'Milton Keynes Hotels'. The main search results section shows 'ALL RESULTS' for '1-20 of 7,020,000 results'. The first result is 'Milton Keynes - Wikipedia, the free encyclopedia', which includes a brief description of the town and its location. To the right of the search results, there is a section for 'Sponsored sites' featuring 'Milton Keynes Hotels' with a promotional message about saving up to 50% on hotels.

Please rate the above document according to its relevance to **milton keynes** as follows. Note that the task is about how relevant to the topic the document is.

☐ **Relevant.** A relevant document for the topic.

☐ **Not relevant.** The document is not good because it doesn't contain any relevant information.

All information is available:

- What to do
- Question to answer
- Relevant resources (search result) is included in the task

Task Design: General Recommendations

- Microtasks should be self-contained
- Keep it short and simple. Brief and concise
- Be very clear with the relevance of the task
- Engage with the worker. Avoid “boring” tasks
- Request feedback:
 - Including an optional open-ended question (input box) in the task
 - Some platforms (e.g., Figure Eight) have built-in surveys for collecting worker feedback

Other Design Principles

- Presentation: text alignment and legibility
- Reading level: complexity of words and sentences
- Attractiveness: worker's attention & enjoyment
- Special needs communities (e.g. simple color blindness)
- Cognitive load: mental rigor needed to perform task
- Multi-cultural / multi-lingual

Impact of Task Design on Crowd Accuracy: Completing Life Sciences Knowledge Graphs

Simple Task Interface:

What is the ICD of Carotid artery dissection?

Does Carotid artery dissection have a ICD?

Choose one answer (required)

- ☐ Yes
☐ No
☐ I don't know

Familiarity with the topic (required)

	1	2	3	4	5	6	7	
Not familiar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very familiar

“Enriched” Task Interface:

What is the ICD of Carotid artery dissection?

[Search in Google: Carotid artery dissection](#)

Short description: Carotid artery dissection is a separation of the layers of the artery wall supplying oxygen-bearing blood to the head and brain, and is the most common cause of stroke in young adults. (In vascular medicine, dissection is a blister-like delamination between the outer and inner walls of a vessel, generally originating with a partial leak in the inner lining.)

Wikipedia page: http://en.wikipedia.org/wiki/Carotid_artery_dissection

Picture:



Does Carotid artery dissection have a ICD?

Choose one answer

- ☐ Yes
☐ No
☐ I don't know

Familiarity with the topic

	1	2	3	4	5	6	7	
Not familiar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very familiar

Impact of Task Design on Crowd Accuracy: Completing Life Sciences Knowledge Graphs

Simple Task Interface:

What is the ICD of Carotid artery dissection?

Does Carotid artery dissection have a ICD?

Choose one answer (required)

- ☐ Yes
☐ No
☐ I don't know

Average F1 = 0.624

Familiarity with the topic (required)

	1	2	3	4	5	6	7	
Not familiar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very familiar

“Enriched” Task Interface:

What is the ICD of Carotid artery dissection?

[Search in Google: Carotid artery dissection](#)

Short description: Carotid artery dissection is a separation of the layers of the artery wall supplying oxygen-bearing blood to the head and brain, and is the most common cause of stroke in young adults. (In vascular medicine, dissection is a blister-like delamination between the outer and inner walls of a vessel, generally originating with a partial leak in the inner lining.)

Wikipedia page: http://en.wikipedia.org/wiki/Carotid_artery_dissection

Picture:

Average F1 = 0.924



Does Carotid artery dissection have a ICD?

Choose one answer

- ☐ Yes
☐ No
☐ I don't know

Familiarity with the topic

	1	2	3	4	5	6	7	
Not familiar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very familiar

Setting Project Properties

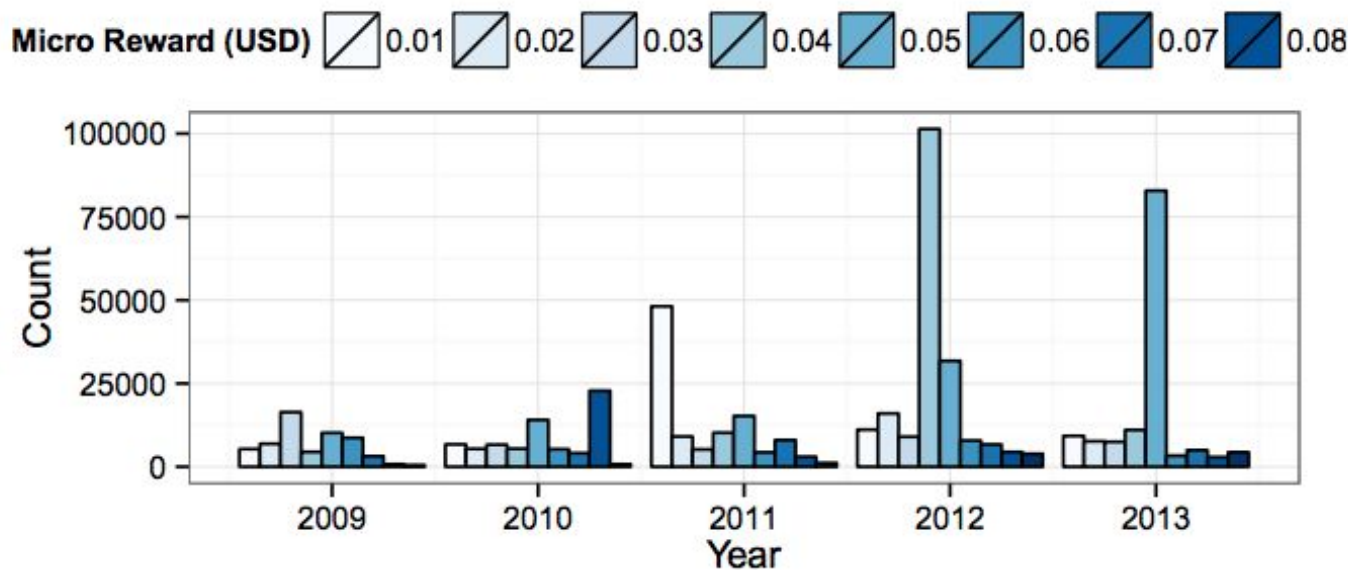
- Microtask duration time:
 - Try to solve a HIT (including reading the instructions and examples)
- Redundancy:
 - Typically between 3 and 5, depending on the difficulty of the task
 - Can be increased on-the-fly (if needed) for some specific tasks
- Rewards:
 - Besides the basic payment, microtask platforms allow for rewarding bonuses to workers who completed additional work

Setting Project Properties: Payments

- Price commensurate with task effort
 - Example: \$0.02 for yes/no answer + \$0.02 **bonus** for optional feedback
- Ethics and market-factors
 - Consider the characteristics of the audience and the marketplace
- Uptake & time-to-completion vs. Cost & Quality
 - Low rewards: no crowd interest or slow answers
 - High rewards: may attract spammers
- Accuracy and quantity
 - More pay does not necessarily produce better answers [Mason and Watts, 2009]

Payments:

Popularity of HIT Rewards in Amazon Mechanical Turk



Difallah, Catasta, Demartini, Ipeirotis, Cudré-Mauroux. *The Dynamics of Micro-Task Crowdsourcing: The Case of Amazon MTurk*. In WWW, 2015.

Execute Sample

- Select a subset of your data
- Run a test of your project to make sure that the settings are fine
- Monitor workers' satisfaction and re-adjust settings (if necessary)
 - Example of workers' satisfaction in one of our tasks executed in Figure Eight.



Execute Project

- **Monitor** the execution of the project
 - Take into consideration the workers' satisfaction and feedback
 - It could be that it is necessary to re-adjust the redundancy of some tasks in the project. This increases the overall cost of the project (check your budget)
- **Accept/reject** crowd answers:
 - In some platforms (e.g., MTurk) **the requester has full control** of accepting and paying for crowd answers/responses. In these cases, workers expect a fast reaction to their contributions
 - **In case of accepting a response:** the worker receives the pre-defined payment
 - **In case of rejecting a response:** the worker receives no payment. The requester may decide to re-submit this microtask to collect an answer from another worker

Collect Data

- Crowd answers are available as soon as they are ready and can be retrieved as: a **CSV file** or **data stream** (using the platform's API)
- The data collected from the microtask platform may include:
 - Timestamps when the task was created and when each crowd answer was submitted
 - Further metadata about the workers (worker ID, country, IP, confidence, total tasks solved)
- Some platforms allow for aggregating the data before retrieving the data:
 - Data can be aggregated by answer field
 - Typical aggregators: majority voting, average (in case of numbers)

Outline

1. Dimensions of Microtask Crowdsourcing
2. Microtasks: Terminology
3. Creating Microtasks
4. **Microtask Platforms: Figure Eight and CrowdFlower**
5. Summary

Components of Microtask Platforms

Task Design and
Workflow
Support

Crowd access
and crowd
selection

Task publication
and assignment

Reward

Quality Control

Interface and
programming
support

Amazon Mechanical Turk (MTurk)

<https://www.mturk.com/>

Information Gathering

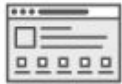
The diversification and the scale of the MTurk workforce allows you to gather a breadth of information that would be almost impossible to do otherwise such as:



Allowing people to ask questions from a computer or mobile device about any topic and have Workers return the results



Filling out market research or survey data on a variety of topics



Writing content for websites



Finding specific fields or data elements in large legal and government documents

Data Processing

Companies take advantage of the power of the MTurk workforce to understand and intelligently respond to different types of data including:



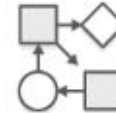
Audio editing and transcription



Human powered translation services



Rating the accuracy of results for a search engine



Categorizing information to match a given schema or taxonomy

Amazon Mechanical Turk (MTurk)

<https://www.mturk.com/>

Image/Video Processing

MTurk is well-suited for processing images. While difficult for computers, it is a task that is extremely easy for people to do. In the past, companies have used MTurk to:



Tag objects found in an image to improve your search or advertising targeting



Review a set of images to select the best picture to represent a product



Audit user-uploaded images or videos to moderate content



Classify objects found in satellite imagery



Removing duplicate content from business listings



Verifying restaurant details such as phone numbers or hours of operation



Identifying incomplete or duplicate product listings in a catalog



Converting unstructured data about locations into well-formed addresses

Amazon Mechanical Turk (MTurk)

Start a New Project

Survey Link

Survey

Tagging of an Image

Moderation of an Image

Choose image A or B

Writing

Collect data

Collect data from a Website

Transcription from A/V

Transcription from a Receipt

Transcription from an Image

Sentiment Wizard

Sentiment Project

Sentiment of a Tweet

Categorization Wizard


Categorization Project

Other

Example of Categorization Wizard

Create Project »

Choose the best category for this image



☐ kitchen

☐ living

☐ bath

☐ bed

☐ outside

View Instructions ↓

Select the room location in home for this picture. Seating areas outside are outside not living. Offices or dens are living not bedrooms. Bedrooms should contain a bed in the picture.

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

Create Project »

[Leave feedback for this page.](#)

Figure Eight (Former CrowdFlower)

<https://www.figure-eight.com/>

We transform messy, real world data into the training data powering AI



Over 100 million images
labeled



Over 10 billion human
judgments



Over 10 years enabling
AI projects

Figure Eight (Former CrowdFlower)

figure eight

Create a new job

Select a template or start from scratch

RECENT
Image Categorization

Use this template

What would you like to do?



Sentiment Analysis



Search Relevance



Data Categorization



Data Collection & Enrichment



Data Validation



Image Annotation



Transcription



Content Moderation

Comparison: MTurk vs. Figure Eight

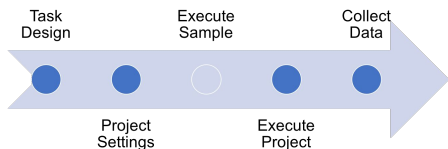
Criterion	MTurk	Figure Eight
Target tasks	Any	Focus on generation of training data for Machine Learning. But any type of task can be submitted
Crowd access	Own pool of workers	Reach to other microtask platforms (channels) to recruit workers
Sandbox support	Yes	No
Quality control	Quality control is responsibility of the requester	The platform filters out spammer or low-quality answers
Features	Basic task interfaces	Support advanced interface language including figure annotation
Fee	A percentage of the task payment	In general, higher than MTurk

Outline

1. Dimensions of Microtask Crowdsourcing
2. Microtasks: Terminology
3. Creating Microtasks
4. Microtask Platforms: Figure Eight and CrowdFlower
5. **Summary**

Summary

- Microtask dimensions: **What** (problem), **Who** (crowd), **How** (task)
- Microtask terminology: requester, worker, HIT, assignment/judgment, reward
- Typical workflow for executing microtasks:



Prepare to iterate!

- Microtask Platforms:
 - MTurk vs. Figure Eight
 - They differ in microtask workflow, worker pool, fees

References

- MTurk: Concepts
<http://docs.aws.amazon.com/AWSMechTurk/latest/RequesterUI/mechanical-turk-concepts.html>
- Figure Eight: Glossary of Terms
<https://success.crowdfunder.com/hc/en-us/articles/202703305-Getting-Started-Glossary-of-Terms>
- The Mechanical Turk Blog
<https://blog.mturk.com>
- Tips For Requesters On Mechanical Turk
<http://turkrequesters.blogspot.co.uk/>
- Requester Best Practices Guide
http://mturkpublic.s3.amazonaws.com/docs/MTURK_BP.pdf

Acknowledgements

Slides adapted from the tutorial “Microtask Crowdsourcing to Solve Semantic Web Problems” by Gianluca Demartini, Elena Simperl, and Maribel Acosta at ISWC 2013.

Source: <https://github.com/maribelacosta/crowdsourcing-tutorial>

Hands-on I: Designing a Task on Figure Eight (CrowdFlower)

<https://www.figure-eight.com/>