Information and Society-E2

Crowdsourcing and Human Computation

Rafik Hadfi

Department of Social Informatics

Kyoto University

Email: rafik.hadfi@i.kyoto-u.ac.jp

These slides may only be used with class enrollees. Distribution of these slides is prohibited. ILAS, Kyoto University

Human Computation

Introduction

- Some ***problems*** are still ***hard*** for computers, e.g.:
  - Visual perception
    (e.g. Object recognition)
  - Natural language analysis
    (e.g. Sentiment analysis)
  - Planning and reasoning
    (e.g. Which route to take? What should you see in Paris?)

Introduction

- Developing ***AI algorithms*** to solve these problems can be difficult or/and needs time
- However, ***humans*** are able to perform these tasks much better and require little effort
- In general, ***understanding (making sense of something)*** is more difficult for computers but relatively easier for humans

➔ Instead of spending time and effort developing new and sophisticated algorithms, it might be easier to ***mobilise a large number of people*** to help you perform these tasks

Human Computation

What is Human Computation?

Firstly, we have to ask ourselves: what is computation?

- *The process of receiving some data input, processing it, and producing some output*

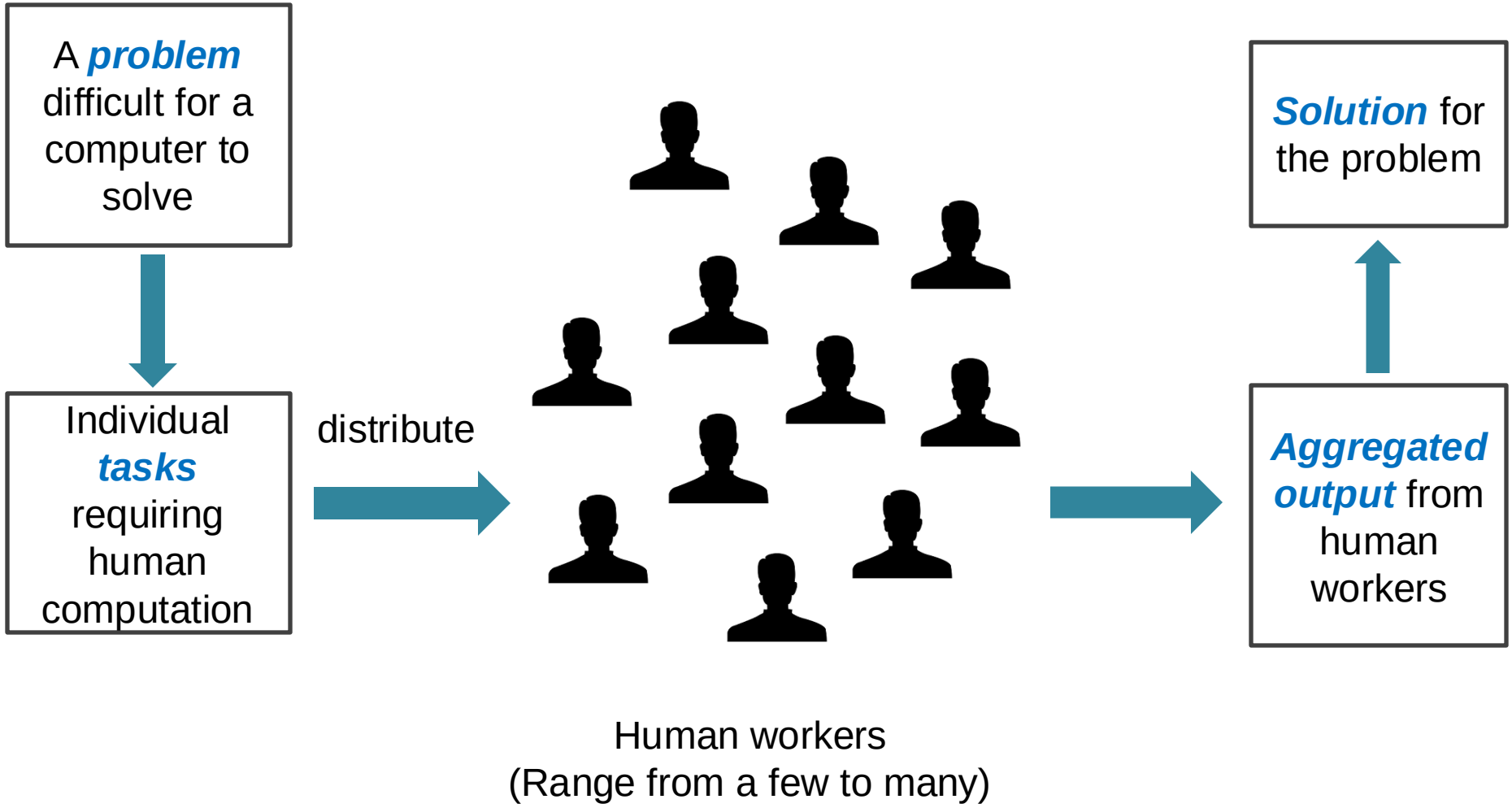
Human Computation

- Computation carried out by human beings

→ A new research area, continuously evolving

Human Computation

What does a human computation system look like?



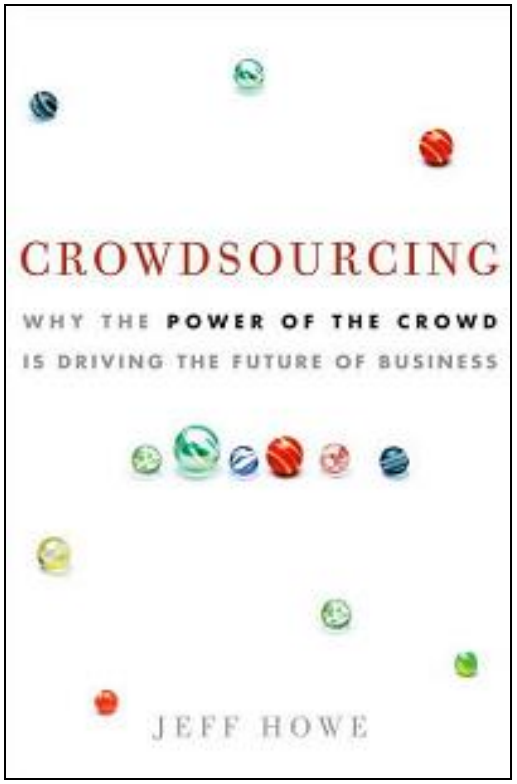
Crowd-sourcing

Crowd-sourcing

- What is ***crowd-sourcing***? (ref. out-sourcing?)
- To obtain solutions, ideas, services, etc., by soliciting input from a large number of people

"*Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call.*"

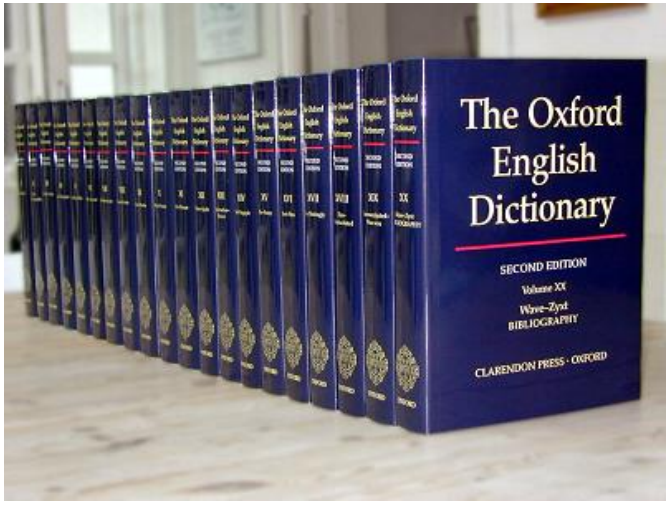
Jeff Howe (Contributing Editor of Wired Magazine)
(http://crowdsourcing.typepad.com/)

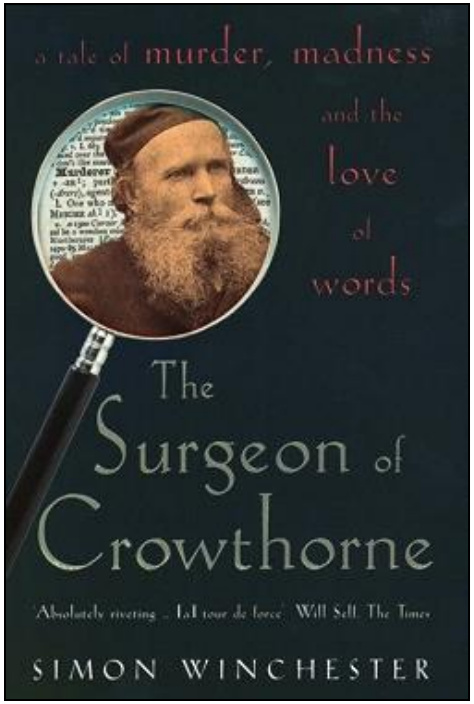


Crowd-sourcing Examples

The Oxford English Dictionary (OED)

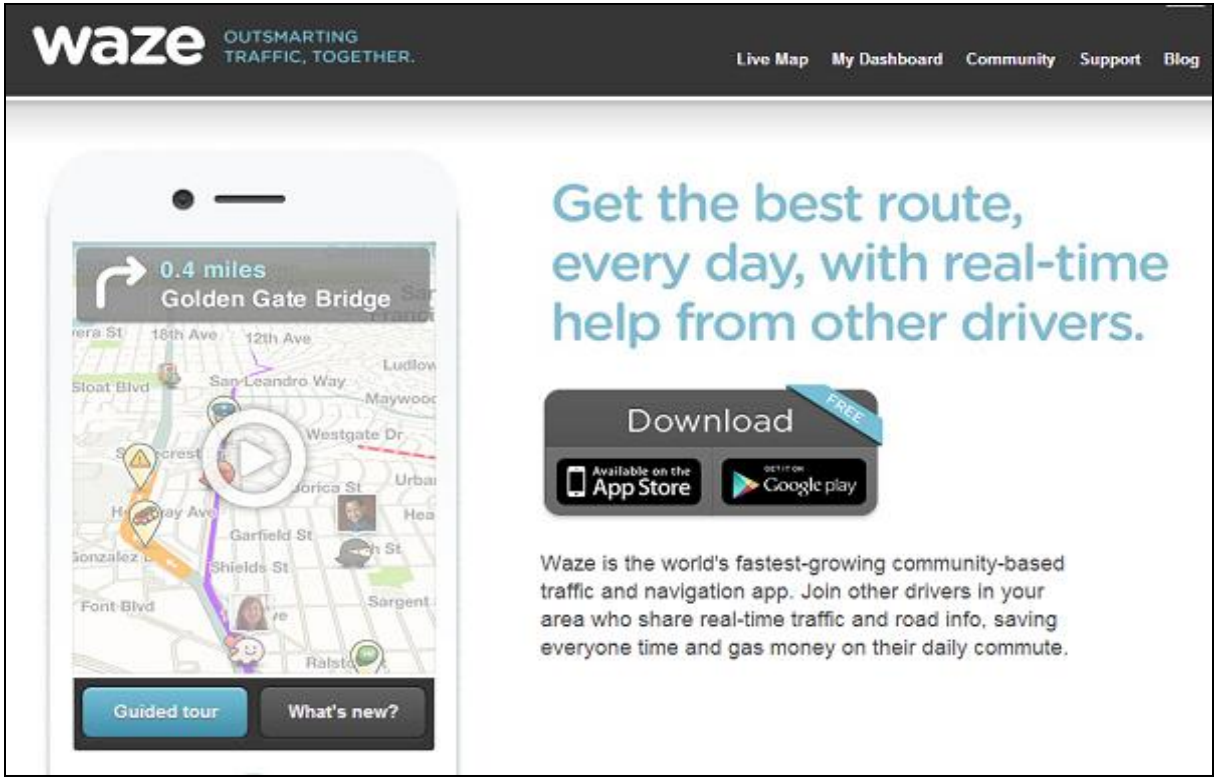
- The dictionary asked people to ***contribute words in the English language and their examples***
- Over 70 years, they received more than ***6 million submissions***





Crowd-sourcing Examples

Waze: Crowd-sourced navigation and traffic information



Citizen Science

- **Public participation in *scientific research***

- New technologies such as ***mobile devices***, ***mobile apps***, and ***mobile broadband Internet***, allow people to ***collect data*** easily

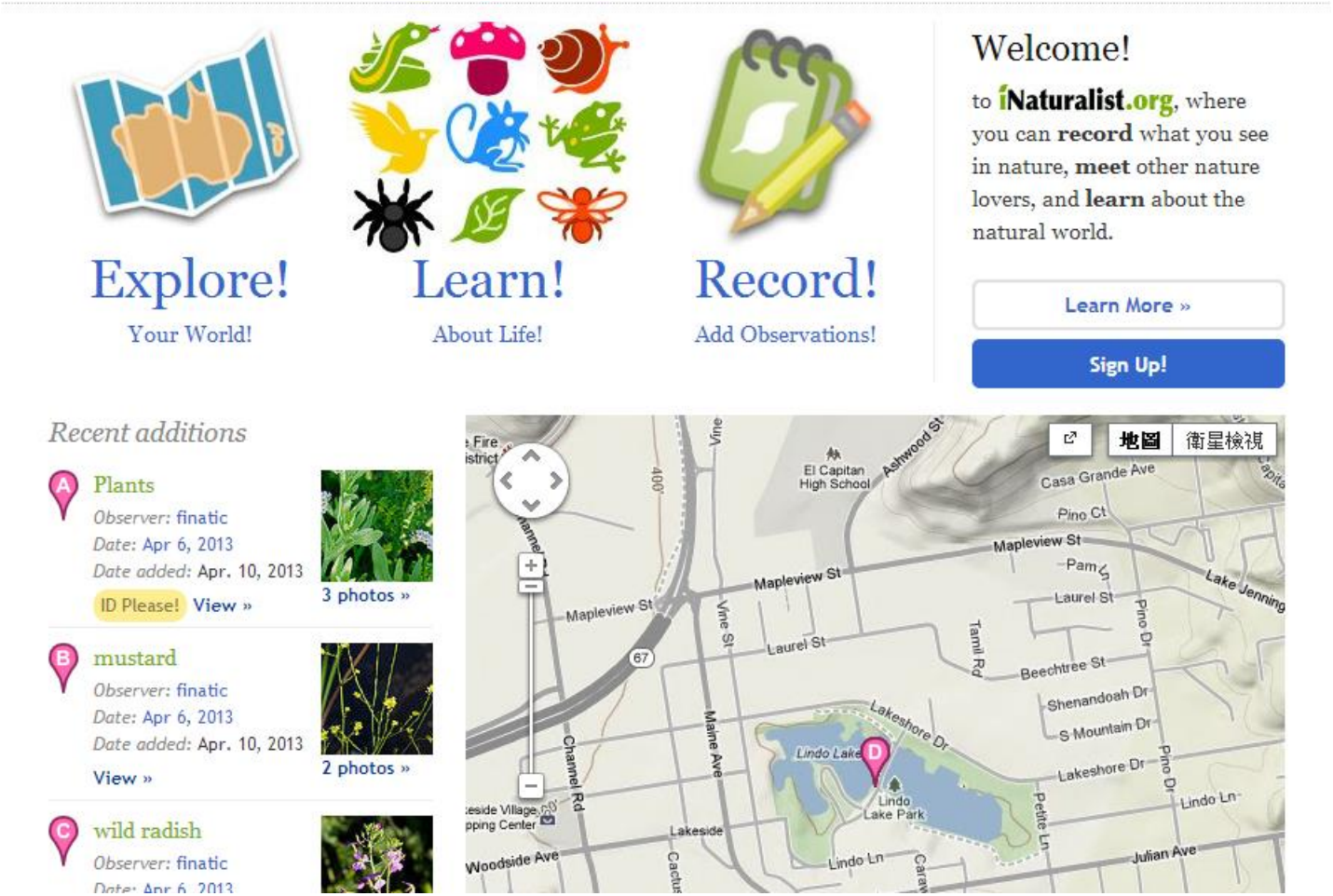
- ***Activities***:
  - Gather scientific data for experiments
  - Gather and share information
  - Carry out labour-intensive tasks to speed up scientific research

Citizen Science

- Project Noah "Networked Organisms And Habitats" (http://www.projectnoah.org/)

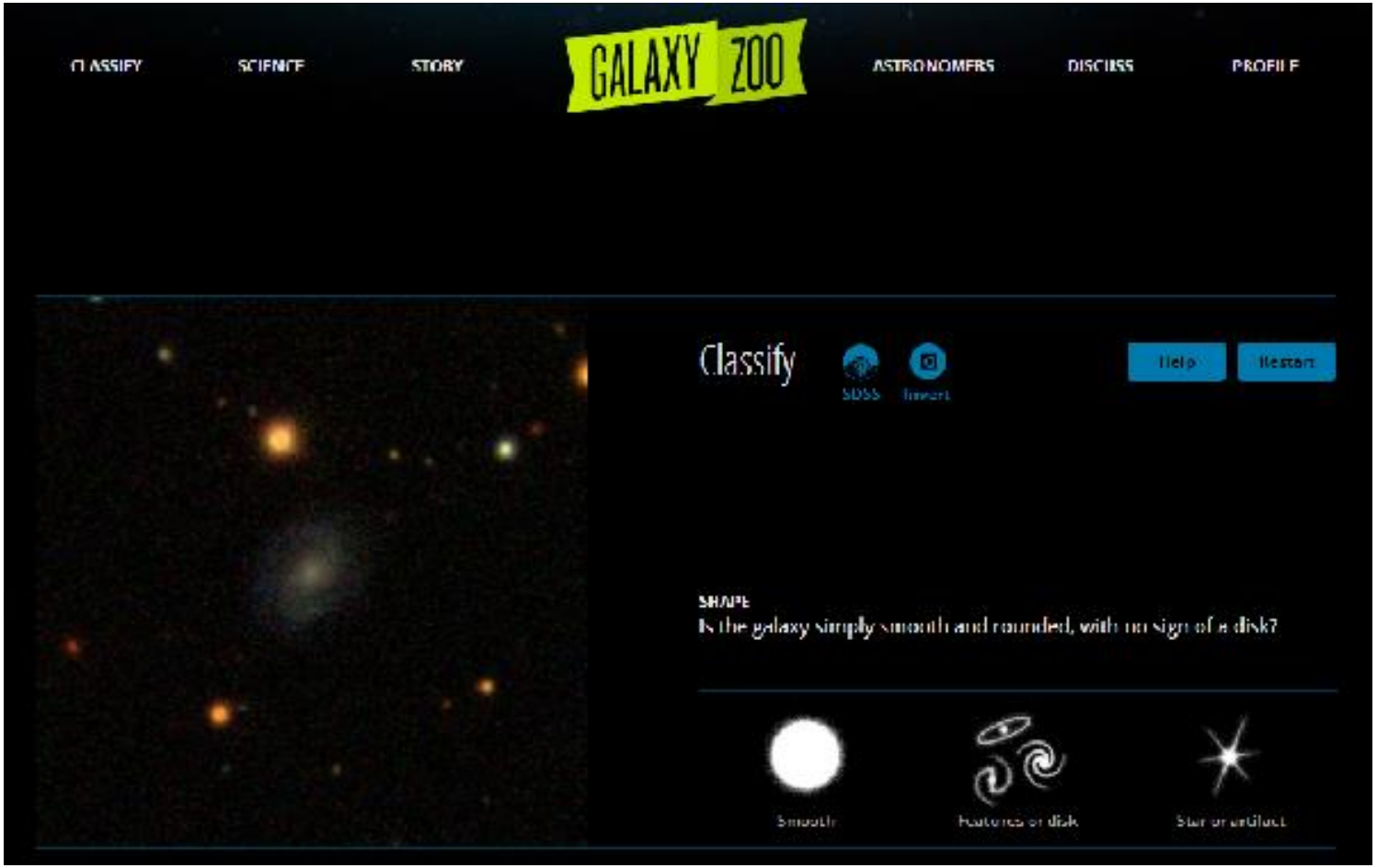
Citizen Science

- iNaturalist (http://www.inaturalist.org/)



Citizen Science

- Galaxy Zoo (http://www.galaxyzoo.org/)



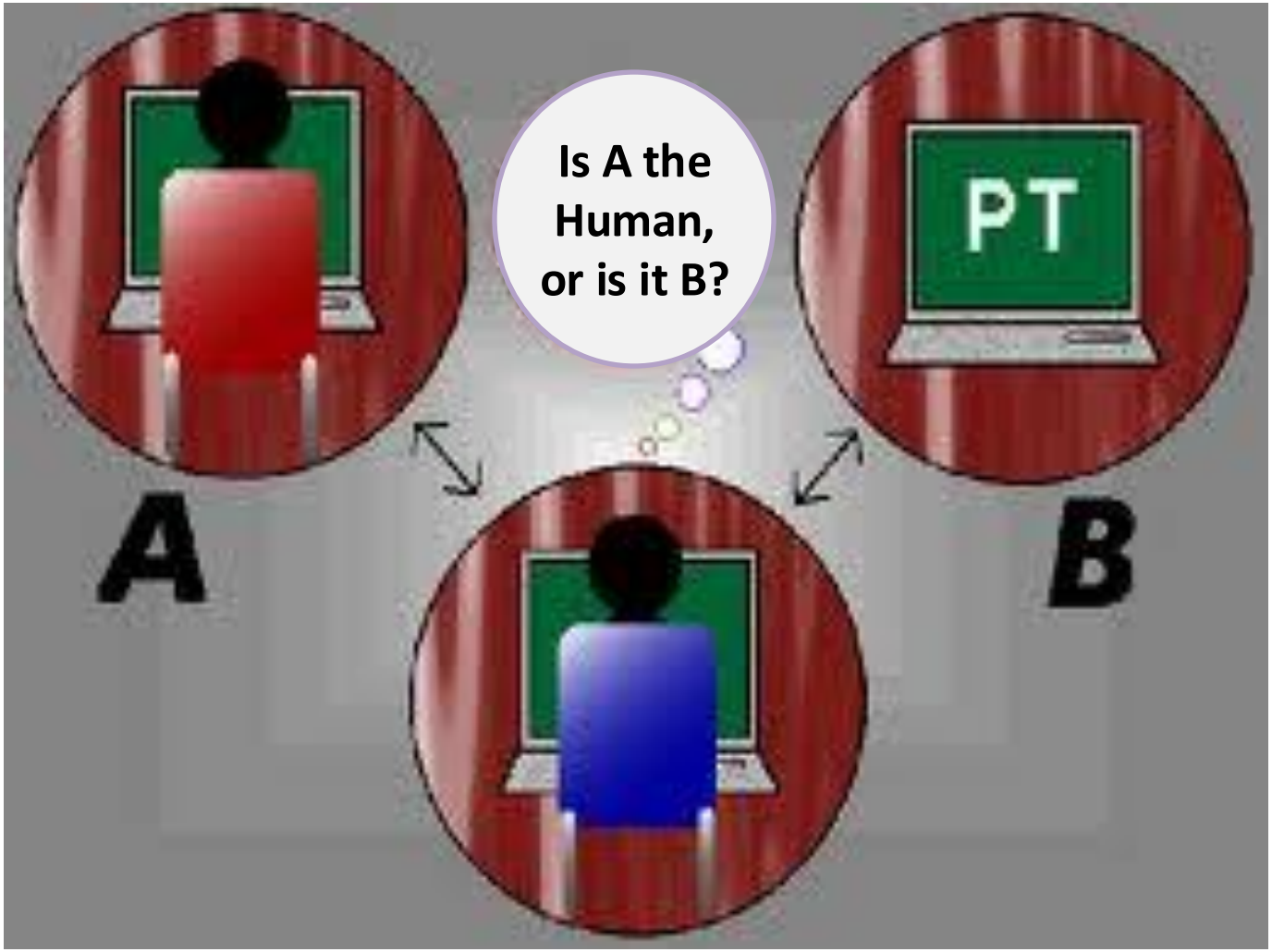
reCAPTCHA

Problem 1:

- Online forms (e.g. registrations, applications) can be filled in ***programmatically*** (using some automatic programs or scripts)
- ***Spamming*** becomes a severe problem for online forms on Websites
- How can we ***verify that it is a human*** that is filling in the form?

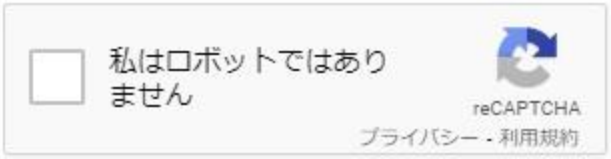
vs.

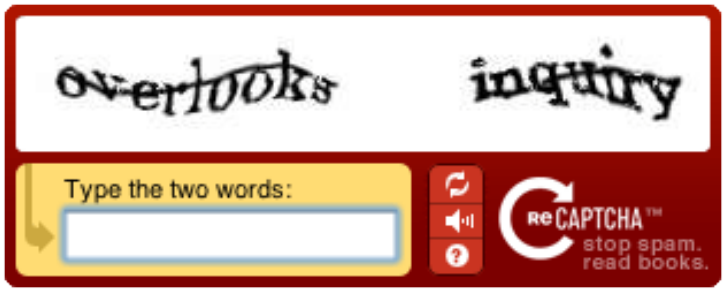
Turing Test and Reverse Turing Test

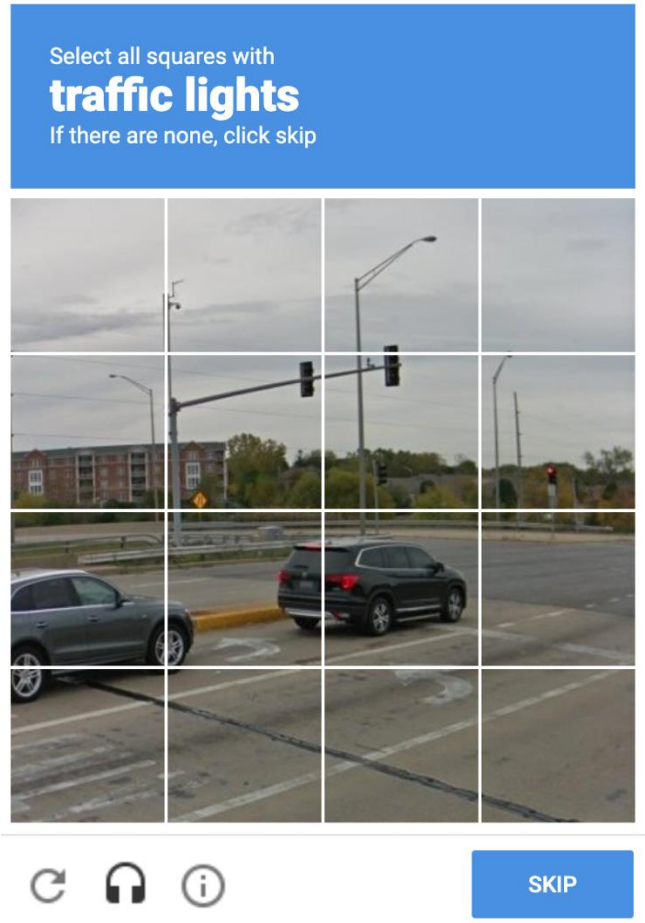


reCAPTCHA

What is this?







reCAPTCHA

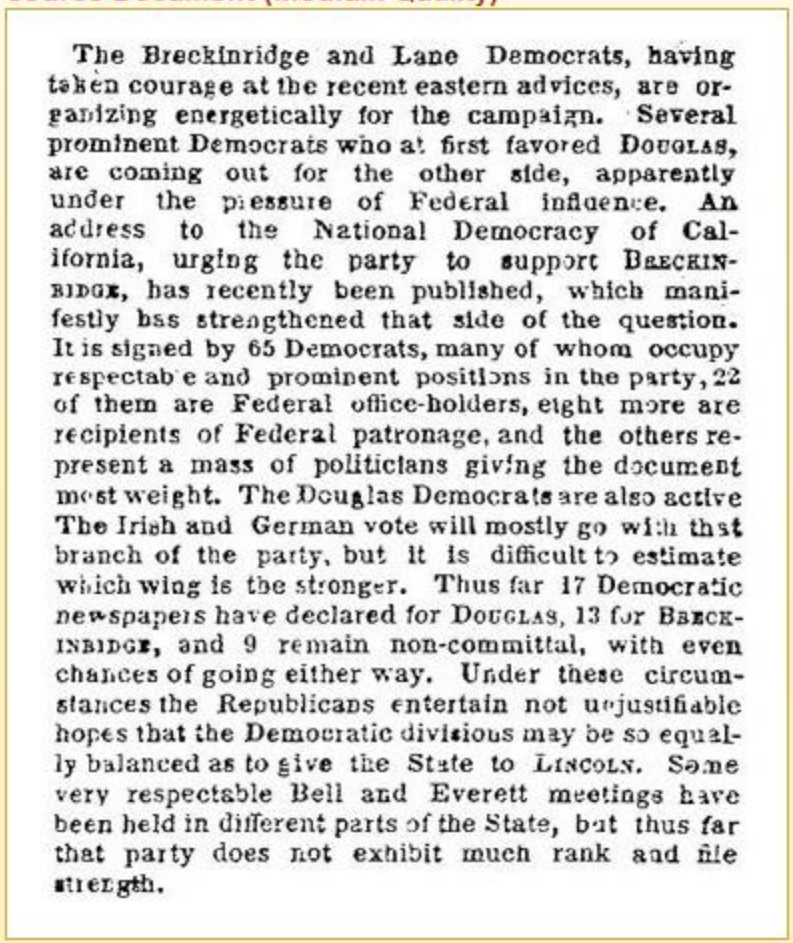
Problem 2:

- We are motivated to ***digitize old documents*** (e.g. newspapers, magazines, books) for various reasons:
  - To preserve the content
  - To search the content digitally

- However, old texts are printed using ***different fonts***, paper in a book may have ***deteriorated***, printing ***may not be clear***
- Difficult for ***OCR programs*** to achieve high accuracy (< 70%)

reCAPTCHA

Source Document (Medium Quality)



The Breckinridge and Lane Democrats, having taken courage at the recent eastern advices, are organizing energetically for the campaign. Several prominent Democrats who at first favored DOUGLAS, are coming out for the other side, apparently under the pressure of Federal influence. An address to the National Democracy of California, urging the party to support BRECKINRIDGE, has recently been published, which manifestly bss strengthened that side of the question. It is signed by 65 Democrats, many of whom occupy respectab e and prominent positions in the party, 22 of them are Federal office-holders, eight more are recipients of Federal patronage, and the others represent a mass of politicians giving the document most weight. The Douglas Democrats are also active The Irish and German vote will mostly go with that branch of the party, but it is difficult to estimate which wing is the stronger. Thus far 17 Democratic newspapers have declared for DOUGLAS, 13 for BRECKINRIDGE, and 9 remain non-committal, with even chances of going either way. Under these circumstances the Republicans entertain not unjustifiable hopes that the Democratic divisions may be so equally balanced as to give the State to LINCOLN. Same very respectable Bell and Everett meetings have been held in different parts of the State, but thus far that party does not exhibit much rank and file strength.

OCR Transcription

The Hreckinridge' and Lane Democrats, having taken courage at the recent eastern advises, are [xxxxxxxxx] energetically for the campaign: Several prominent Democrats who at first favored DonoLea, are coming out. for the other aide, apparently under the [xxxxxxx] of Federal [xxxxxxxx]. An address to the National Democracy of ,1ifornia, urging the party to support HaeeslipslDas, has recently been published, which manifestly bss strengthened that aide of the [xxxxxxxx]: It is signed by 65 Democrats, many of whom occupy respectab e and prominent positions in the party, 22 of them are Federal office-holders, [xxxxx] more are recipients of Federal patronage, and the others represent a mass of politicians giving the document [xxxx] [xxxxx] mTheDcu8las Democrats are also active The Irish and German vote will mostly go with ths# branch of the party, but it is [xxxxxxxx] to [xxxxxxx] [xxxxx] [xxxx] [xx] the stronger. Thus far 17 IT newspapers have declared for DonGres, 13 for Base$- laalDGS and 9 remain non-committal, with even chances of going either way. Under these circumstances the Republicans entertain not unjustifiable hopes that the Democratic divisions may be so equal,- ly balanced as to give the State [xx] LIaCOLV. Same very [xxxxxx] Bell and Everett meetings have been held in different parts of the State, bat thus far that party does not exhibit much rank sad ale air en.

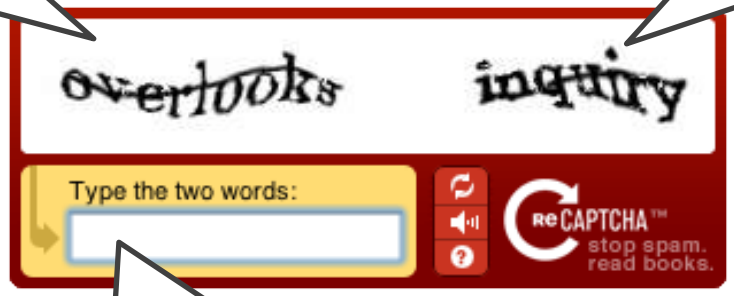
Source Document
(Scanned Image)

OCR Output
(Text)

reCAPTCHA

A word selected from the database, the correct answer of which is known (***control word***)

A word from the output of OCR on some old texts. ***Human verification*** needed.



It is difficult to design a computer program/algorithm to do this task, while a human being is able to do so rather easily.

reCAPTCHA

http://www.google.com/recaptcha

reCAPTCHA

CAPTCHA = "**C**ompletely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans Apart"

Past statistics

- Over 100 million CAPTCHAs are displayed everyday
- 750 million people (**10 % of humanity**) have helped digitize at least one word
- Current transcription rate / day: > 160 books

Try if you have a Website:
https://developers.google.com/recaptcha/

Games with a Purpose (GWAP)

- People like to play ***games***
- What if we incorporate human computation elements into a game?
- Players are made ***to solve some computational tasks*** that are difficult for computers, while they are playing the game

Games with a Purpose (GWAP)

GWAP Examples

Object Recognition

- In computer vision research, object recognition is an important task
- While developing new algorithms, training data with ***ground truth*** (***annotations***) are usually needed

Ferrari, Sports Car, Red Car

A380, Airbus, Aeroplane

The ESP Game

- Two users are shown the same image
- When they come up with the same word for describing the image, they can proceed to the next stage

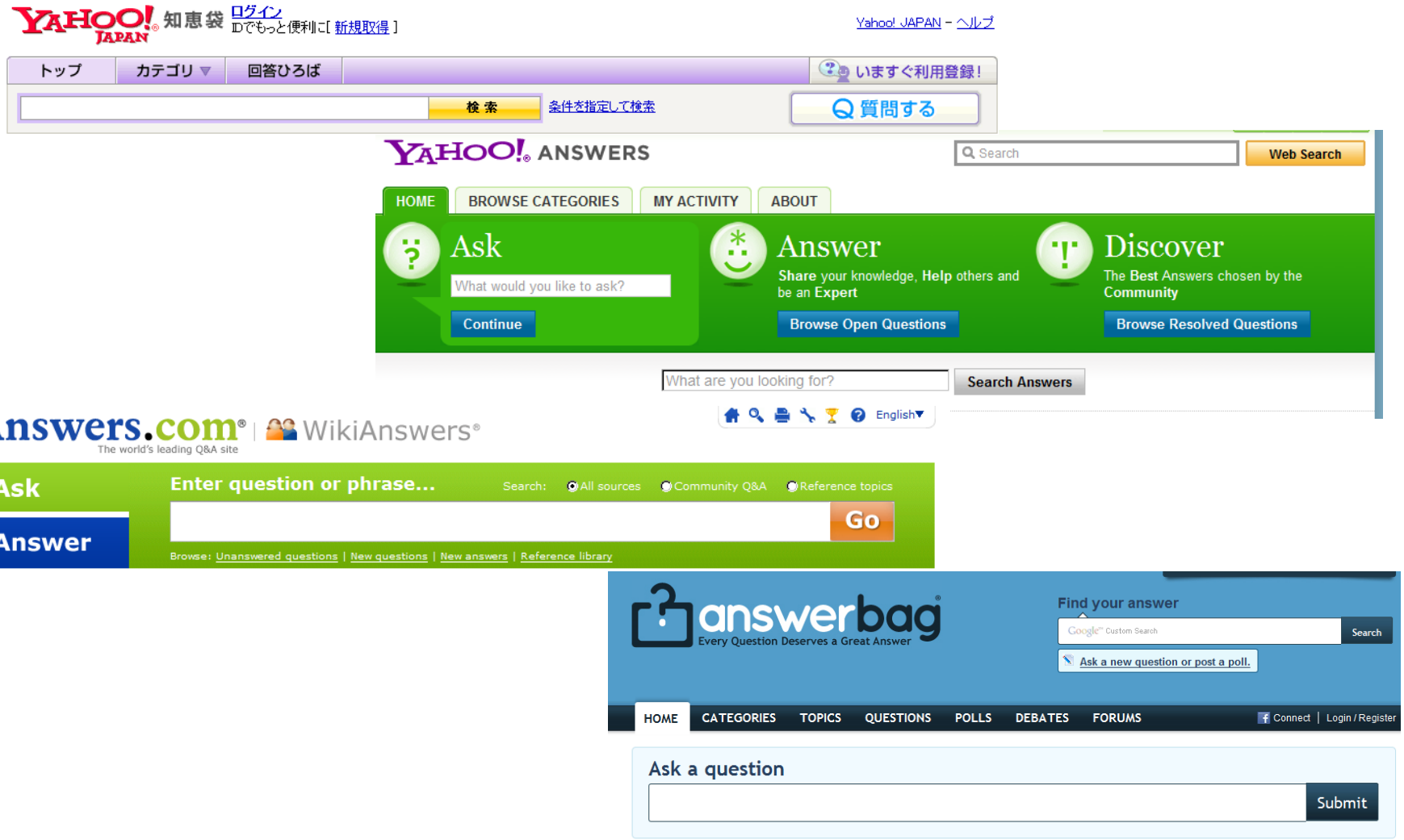
The ESP Game

Sheep
Animal
Grass
…

Tree
White
Sheep
…

‘sheep’ is a label for this image.

Other Social Computing Examples: Community QA (CQA) Sites, Wikipedia, etc.



Crowd-sourcing Market Places

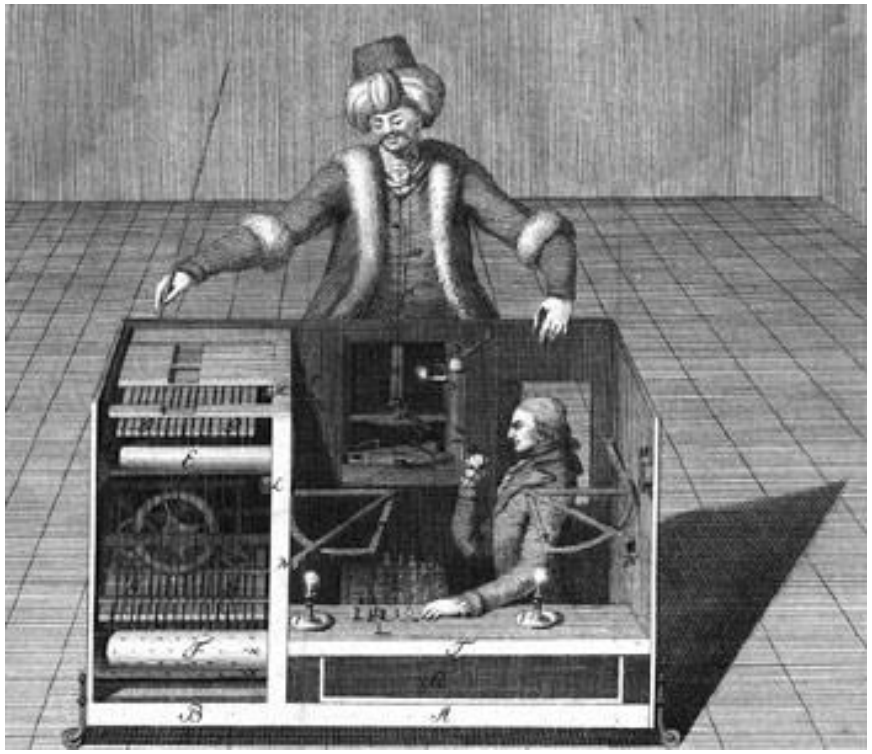
Amazon Mechanical Turk

Amazon Mechanical Turk
https://www.mturk.com/mturk/

Amazon Mechanical Turk

What is 'the Mechanical Turk'?

- A fake chess-playing machine constructed in the late 18th century.
- A human chess player hiding inside the machine was actually playing the game
- More at: http://en.wikipedia.org/wiki/Mechanical_Turk



Amazon Mechanical Turk

- An online platform established by Amazon, where:
  - You can create a human computation / crowd sourcing task
  - Other people can come to work on the task and get paid

- Tasks are called ***HITS*** '***Human Intelligence Tasks***'

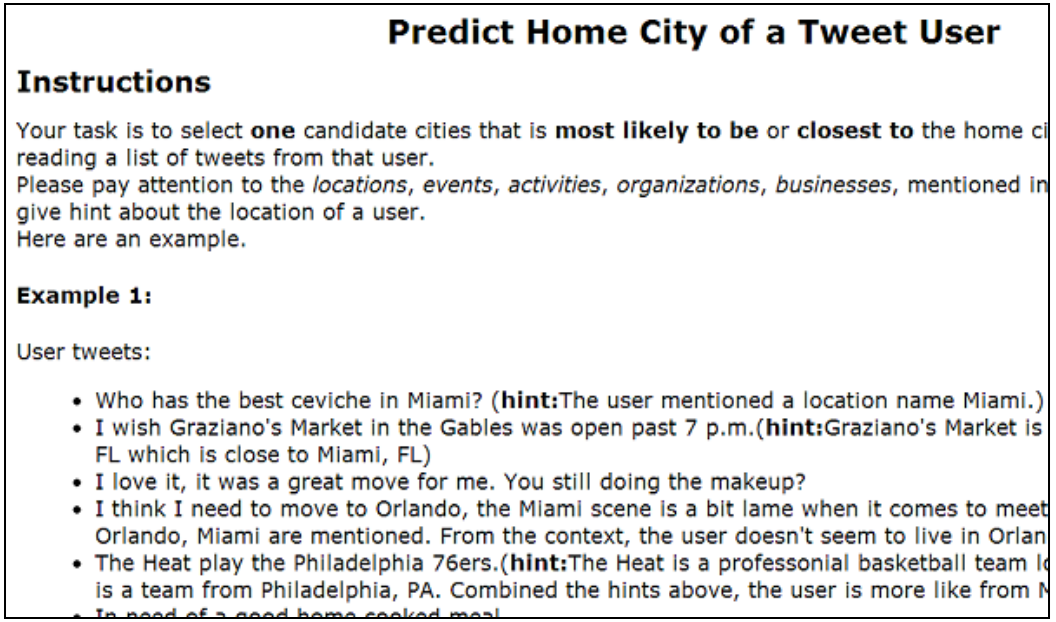
- 90% of the HITS have rewards < US$0.10

Amazon Mechanical Turk

What kind of tasks are created on AMT?

Created by researchers for their research projects

- Does the tweet involve positive or negative emotion?
- What is the person doing in the image?
- Extract person names mentioned in a text document



Predict Home City of a Tweet User

Instructions

Your task is to select **one** candidate cities that is **most likely to be** or **closest to** the home ci
reading a list of tweets from that user.
Please pay attention to the *locations*, *events*, *activities*, *organizations*, *businesses*, mentioned in
give hint about the location of a user.
Here are an example.

Example 1:

User tweets:

- Who has the best ceviche in Miami? (**hint:**The user mentioned a location name Miami.)
- I wish Graziano's Market in the Gables was open past 7 p.m.(**hint:**Graziano's Market is
  FL which is close to Miami, FL)
- I love it, it was a great move for me. You still doing the makeup?
- I think I need to move to Orlando, the Miami scene is a bit lame when it comes to meet
  Orlando, Miami are mentioned. From the context, the user doesn't seem to live in Orlan
- The Heat play the Philadelphia 76ers.(**hint:**The Heat is a professonial basketball team l
  is a team from Philadelphia, PA. Combined the hints above, the user is more like from N
- In need of a good home cooked meal

Amazon Mechanical Turk

What kind of tasks are created on AMT?

Routine tasks

- Transcription (e.g. audio to text)
- Simple translations
- Copying information (e.g. from business card to text)
- Etc.

Problems in Human Computation & Crowd-sourcing

Problems

- Incentive model
- Behaviourial & cognitive aspects of the users
- Decomposition of a task
- Quality assurance
- Method of aggregation
- Task routing

Problems - Incentives

How to motivate people to do the work?

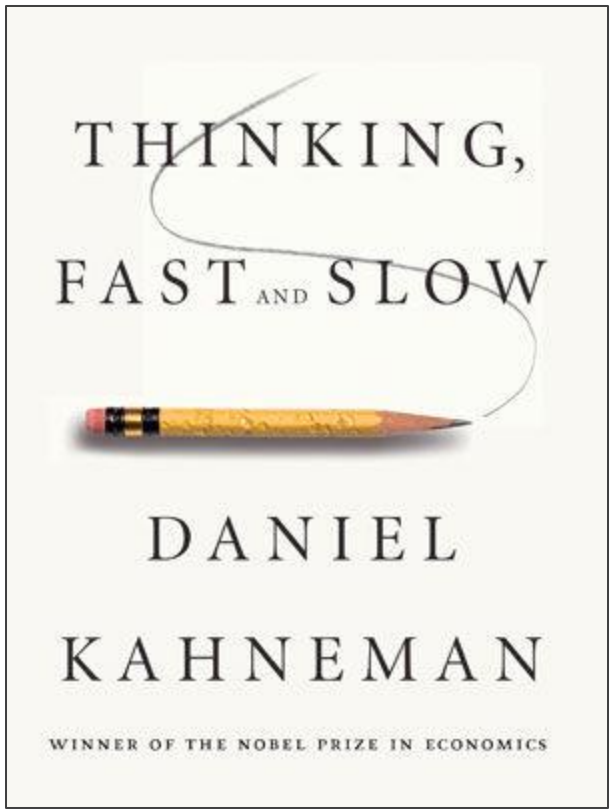
- Human computation and crowd-sourcing tasks usually have a good and meaningful purpose
- However, people do not work for nothing
- How can we motivate or incentivise people to take up the work?
  - ***Money $$$***
  - ***Entertainment (games, e.g. ESP Game, Tag-a-tune)***
  - ***Educational (something to learn, e.g., Duolingo)***
  - ***Credits***
  - ***Altruism (e.g., free food)***

Problems – Human Errors/Biases

- Computers do not make mistakes if a program or an algorithm is designed and implemented correctly
- However, ***humans*** make ***mistakes***
- Output of human computation can be ***noisy*** and ***inaccurate***, even if they are legitimate workers (not spammers)

Problems – Human Errors/Biases

- Many psychology experiments show that human can be ***systematically biased*** by how a question is presented and what information is included
- ***Cognitive Biases***:
  - Anchoring
  - Priming
  - Sequential context biases
  - Intensities of answers are connected



Problems – Human Errors/Biases

Anchoring (cognitive bias)

E.g. What is your best guess of the percentage of African nations in the UN?

- Participants are asked to spin a wheel, and write down the number they see
- People's guesses:
  - Mean answer = 25% for those who wrote down 10
  - Mean answer = 45% for those who wrote down 65

Problems – Human Errors/Biases

Anchoring (cognitive bias)

Students shown "1 x 2 x 3 x 4 x 5 x 6 x 7 x 8" made a median estimate of 512, while students shown "8 x 7 x 6 x 5 x 4 x 3 x 2 x 1" made a median estimate of 2,250

- 40,320 being the correct answer

Problems – Human Errors/Biases

Priming (psychology)

- According to a study at the New York University, what people read can affect their walking speed
- People are divided into two groups, one group is asked to manipulate words about the ***elderly*** (e.g. *wrinkle*, *forgetful*, *bald*, *gray*, *Florida*, …)
- People in the above group walk down the corridor more ***slowly*** than the others.

Other examples include place of a voting booth, priming people with money-related objects, …

Problems – Quality Assurance

- Humans do not only make mistakes, but some may intentionally create '***spams***' (e.g. just to earn the money)
- How can we ***maintain quality*** of the output in crowd-sourcing?
- **Common approaches**:
  1. **Verification** (insert tests with ground truth occasionally, ref. reCAPTCHA)
  2. **Voting** (send the same tasks to multiple workers and obtain majority votes)
  3. **Filtering** (remove worst or outlier answers)
  4. **Track user's reputation** (user ranking, black lists)

Problems – Quality Assurance

Inter-rater/Inter-annotator Agreement

- Let's say you are working on a ***classification task***, trying to classify datapoints into '**Yes**' and '**No**'
- You need some ***ground truth data***, so you ask two of your friends to label some data for you
- It is possible that they give you ***different*** answers, for example:
  - Friend A: 001=Yes, 002=No, 003=Yes, …
  - Friend B: 001=No, 002=No, 003=Yes, …
- At the end, how do you judge the ***quality of this labelling***?

Problems – Quality Assurance

Inter-rater/Inter-annotator Agreement

- We can use a statistical measure called Cohen's Kappa

$$\kappa = \frac{p(a) - p(e)}{1 - p(e)}$$

$p(a)$ is the proportion of agreement among two raters

$p(e)$ is the hypothetical probability of chance agreement

- The measure κ quantifies the agreement between 2 raters who each classify N items into C mutually exclusive categories
  - Complete agreement yields $\kappa = 1$, else 0

Problems – Quality Assurance

Inter-rater/Inter-annotator Agreement - Example

- Let's say you asked each of them to *label N=50 data points*:
- P(a) = (18 + 20) / 50 = 0.76
- P(A says 'Yes') = (18 + 5) / 50 = 0.46
  P(B says 'Yes') = (18 + 7) / 50 = 0.50
- Therefore,
  P(e)
  = P(Both say 'Yes') + P(Both say 'No')
  = (0.46 x 0.50) + (0.54 x 0.50) = 0.5
- Kappa = 0.52

		B: Yes	B: No
A	Yes	18	5
	No	7	20

$$\kappa = \frac{p(a) - p(e)}{1 - p(e)}$$

Problems – Quality Assurance

Inter-rater/Inter-annotator Agreement

- Cohen's Kappa is applicable when there are 2 raters
- When there are more raters, use Fleiss' Kappa (http://en.wikipedia.org/wiki/Fleiss%27_kappa)

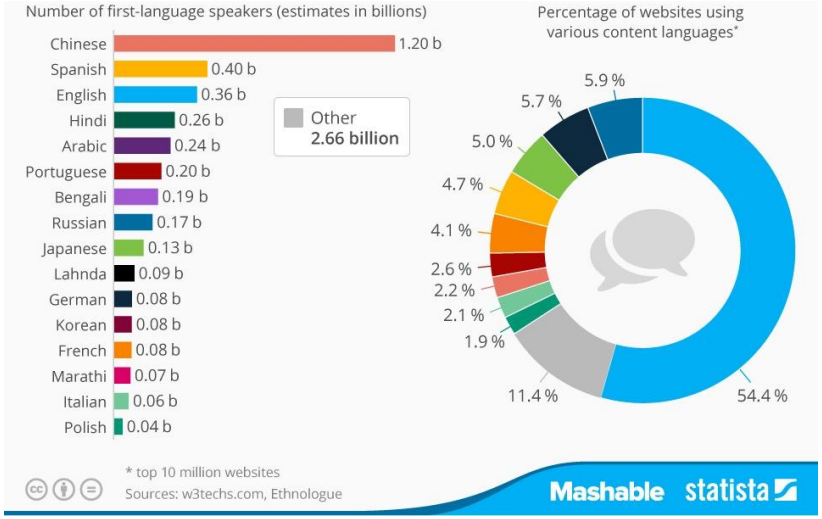
Luis von Ahn: Massive-scale Online Collaboration

https://www.youtube.com/watch?v=-Ht4qiDRZE8

Summary

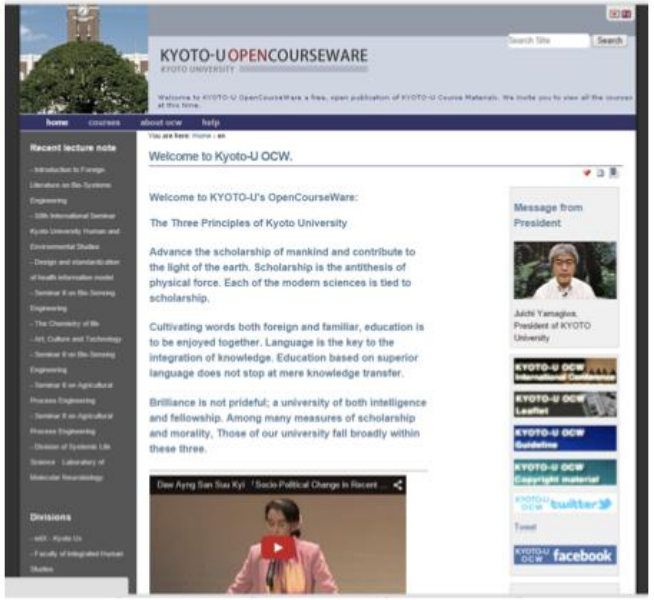
Information Policy and Ethics

- ICT infrastructure
- Japan's ICT policy
- Ethical issues of ICT
- Information security and cyber crimes



Information Education

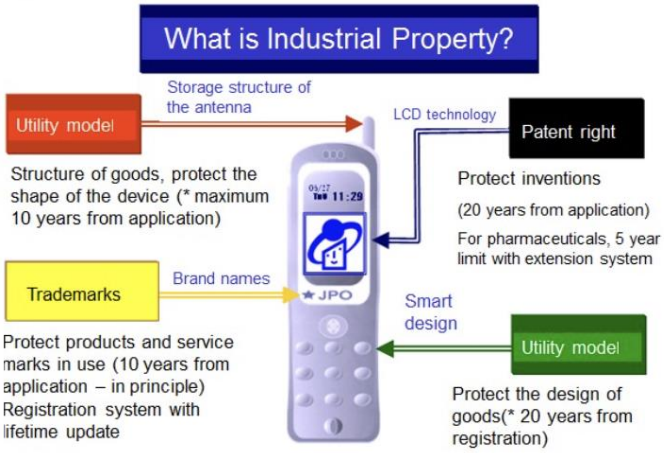
- e-learning
- OpenCourseWare (OCW) and MOOC
- Information literacy
- Etc.

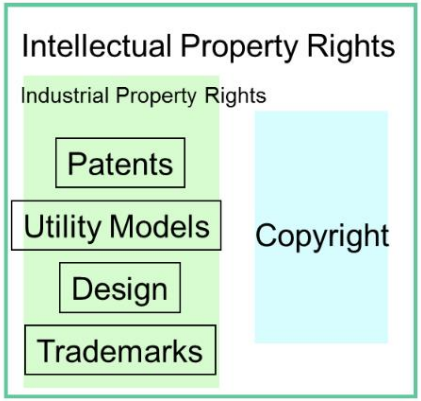


Information Law

- Intellectual property rights and patents, university and patents
- Copyrights
- Personal information protection
- Etc.

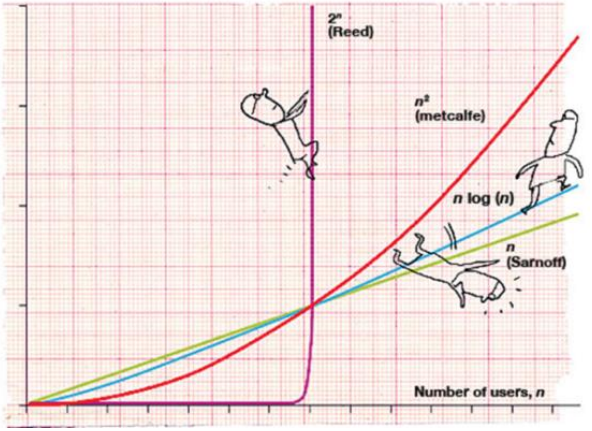
•Can be implemented industrially?	Industrial usability, usefulness
•Is it is new or not?	Novelty
•Could it have been thought of simply?	Non-obviousness, progressivity
•Has it been previously applied for?	Previously applied principle
•Is it not an anti-social invention?	Sociability ?
•Are the contents of the invention adequately explained in the specifications document?	Feasibility

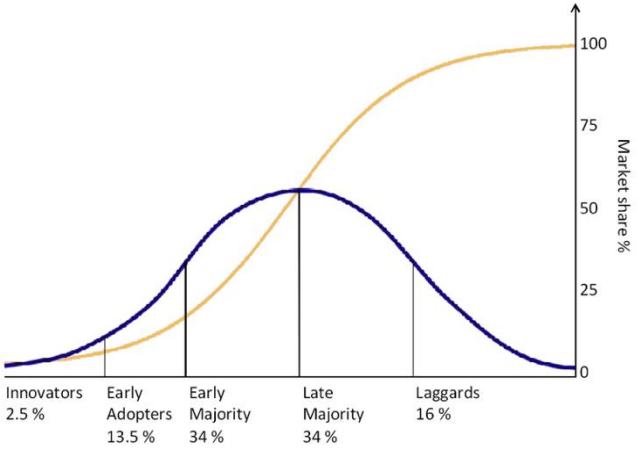


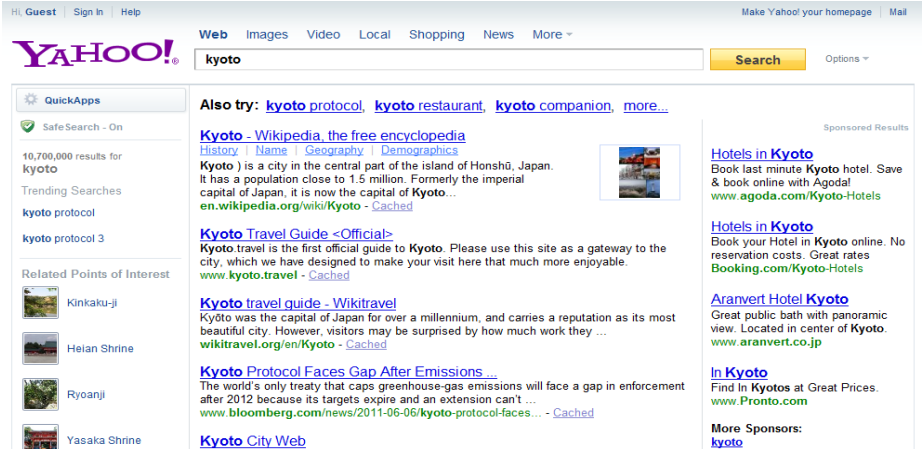


Information Economy

- Impact of Internet on economy
- Externality, network externality
- Consumer search theory
- Information asymmetry
- Online advertising

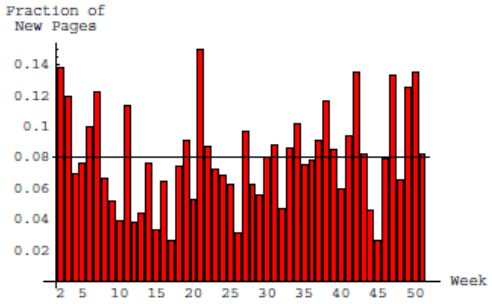


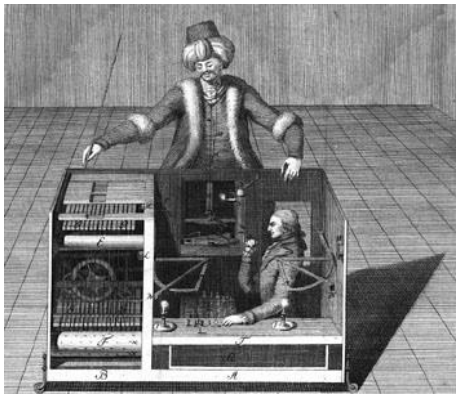




Other Topics

- Evolution of information society
- Information archiving
- Social media analysis
- Cloud computing
- Crowdsourcing and Human computation





Class Questionnaire

- “Questionnaire System” in “Common Portal for All Students
  https://student.iimc.kyoto-u.ac.jp/index.html

- Period: July 3 (Monday) to August 22 (Tuesday)

These slides may only be used with class enrollees. Distribution of these slides is prohibited. Graduate School of Informatics, Kyoto University

Feedback

- No specified class for feedback
- Feel free to email the instructor if you would like to discuss any of the topics of the lecture

Rafik Hadfi

Department of Social Informatics

Kyoto University

Email: rafik.hadfi@i.kyoto-u.ac.jp

These slides may only be used with class enrollees. Distribution of these slides is prohibited. Graduate School of Informatics, Kyoto University