

Information and Society-E2

Crowdsourcing and Human Computation

Rafik Hadfi

Department of Social Informatics

Kyoto University

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

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?

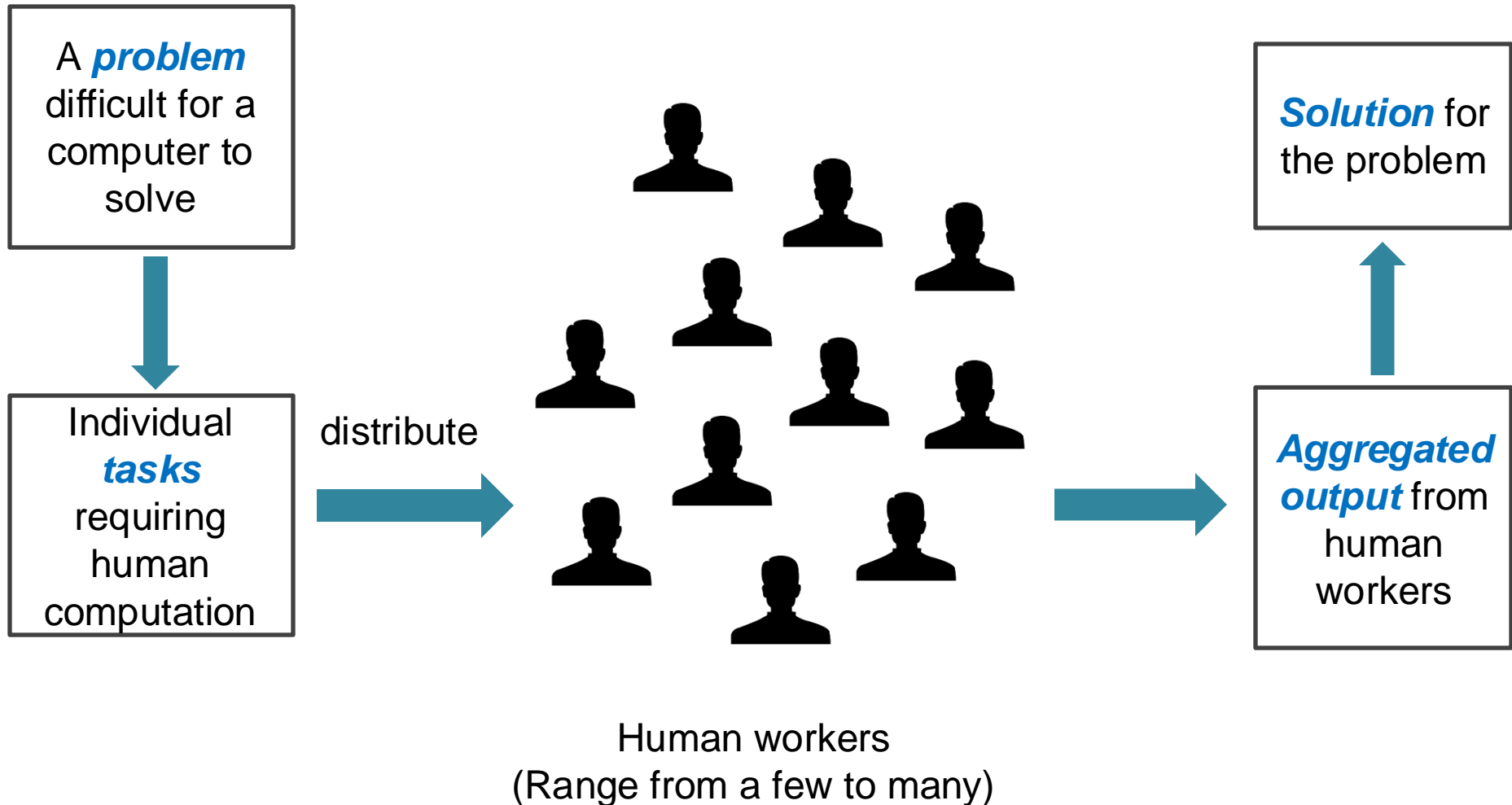
- *The process of receiving some data input, processing it, and producing some output*

Human Computation

- Computation carried out by human beings
- A new research area, continuously evolving

Human Computation

What does a human computation system look like?



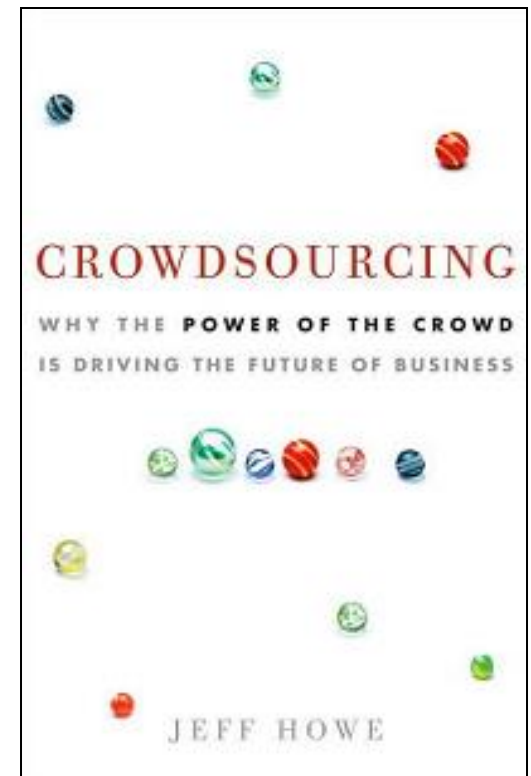
Crowd-sourcing

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.”

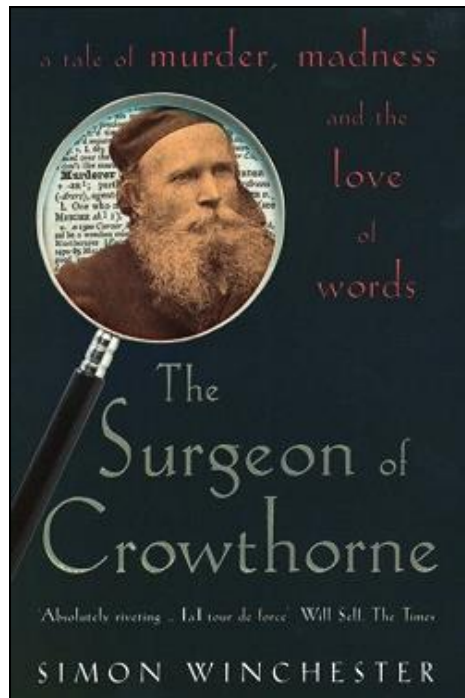
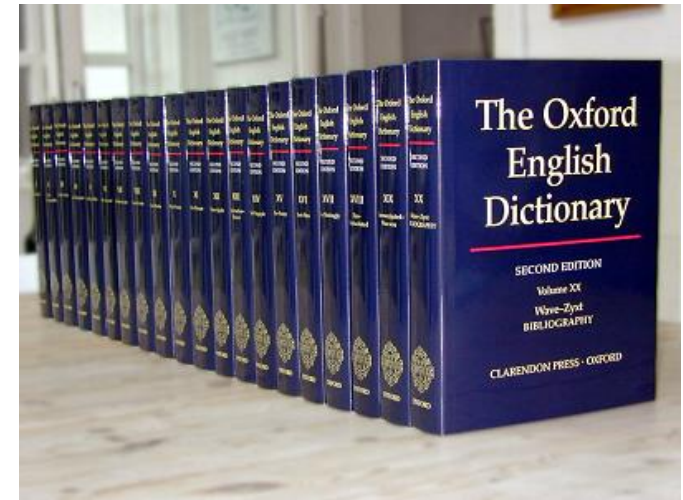
Jeff Howe (Contributing Editor of Wired Magazine)
(<http://crowdsourcing.typepad.com/>)



Crowd-sourcing Examples

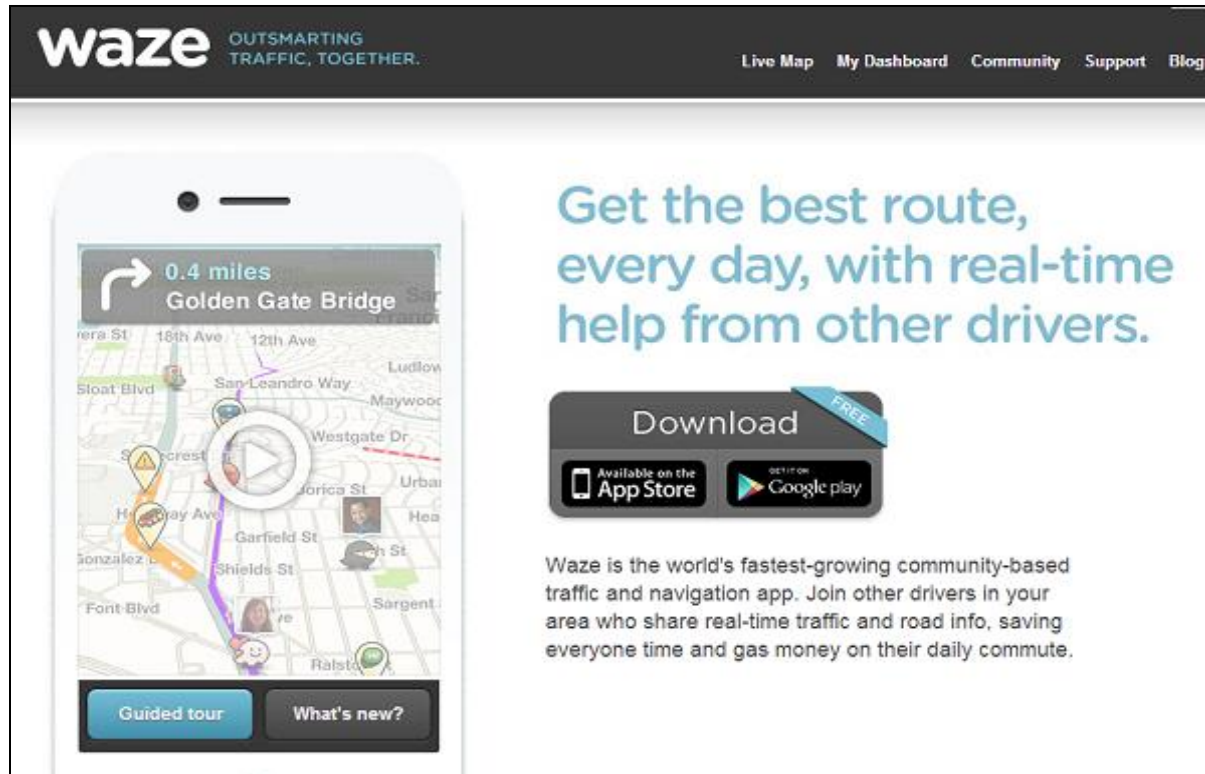
The Oxford English Dictionary (OED)

- 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/>)




Explore!

Your World!



Learn!

About Life!



Record!

Add Observations!

Welcome!


to **iNaturalist.org**, where you can **record** what you see in nature, **meet** other nature lovers, and **learn** about the natural world.

[Learn More »](#)

[Sign Up!](#)


Recent additions

A **Plants**
Observer: finatic
Date: Apr 6, 2013
Date added: Apr. 10, 2013
[ID Please!](#) [View »](#)




3 photos »

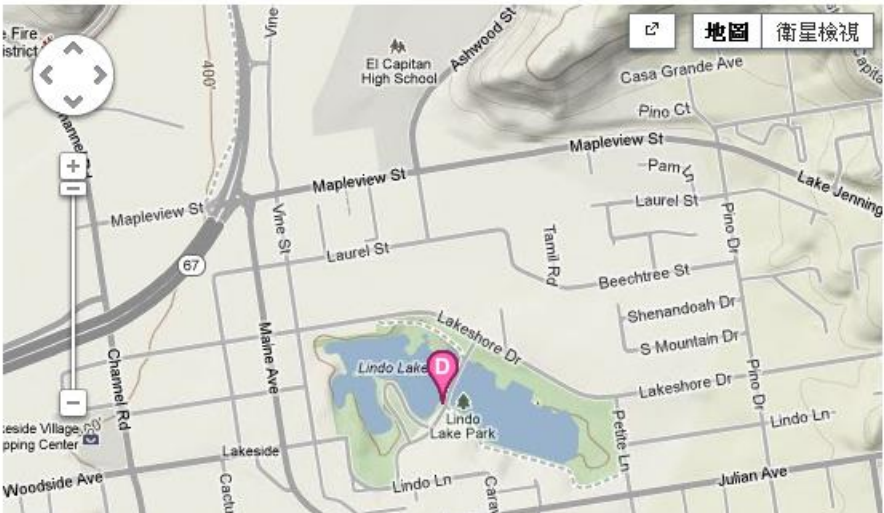
B **mustard**
Observer: finatic
Date: Apr 6, 2013
Date added: Apr. 10, 2013
[View »](#)



2 photos »

C **wild radish**
Observer: finatic
Date: Apr 6, 2013





Map showing Lindo Lake Park area with streets: Mapleview St, Vine St, Laurel St, Ashwood St, Casa Grande Ave, Pino Ct, Maplevue St, Pam St, Lake Jenning, Beechtree St, Shenandoah Dr, S Mountain Dr, Lakeshore Dr, Lindo Ln, Julian Ave, Woodside Ave, Cactus, Lakeside, Lindo Ln, Lindo Lake, Lindo Lake Park, El Capitan High School, Fire District, 400, 67, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000

Citizen Science

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



reCAPTCHA

Problem 1:

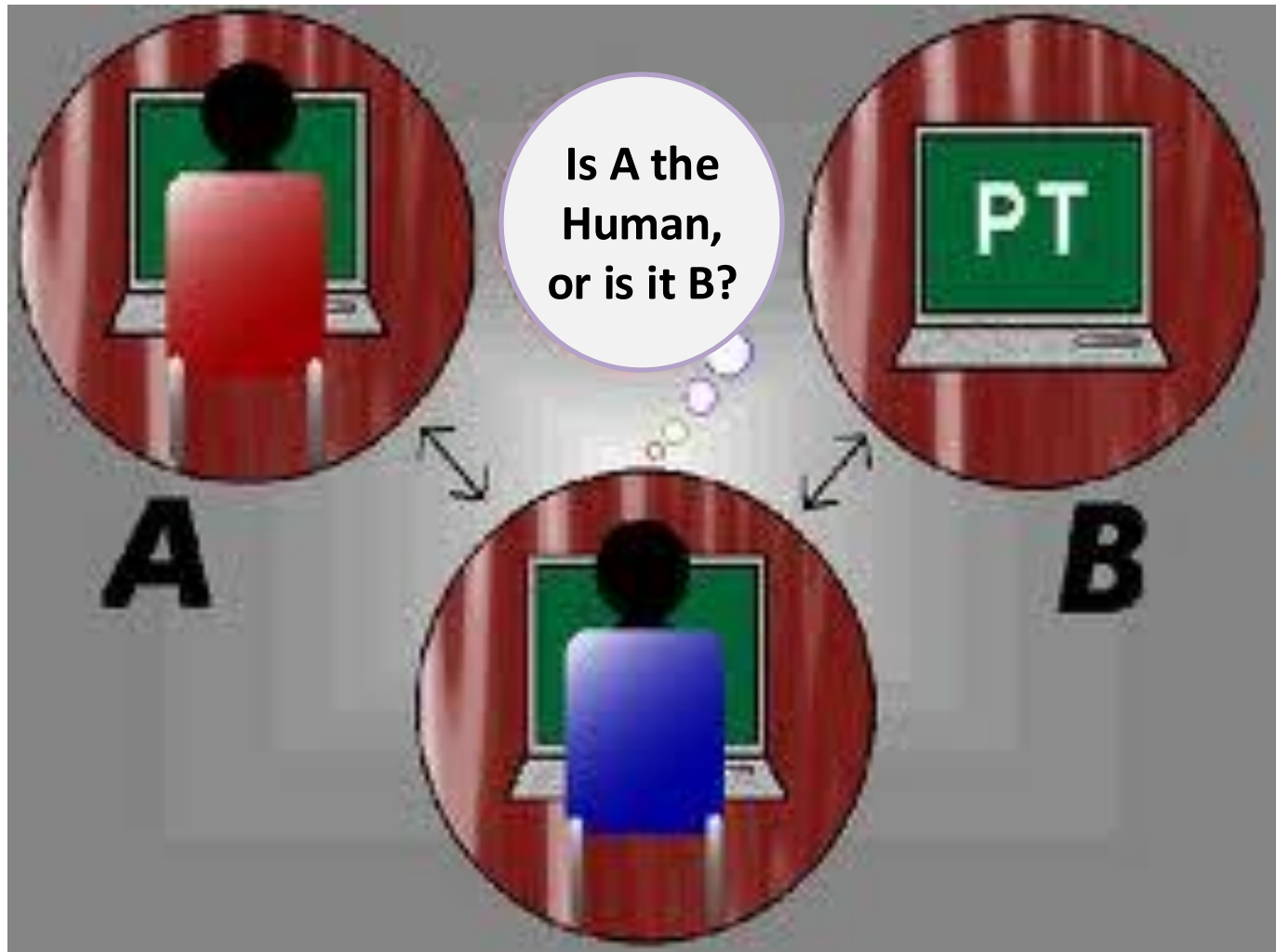
- Online forms (e.g. registrations, applications) can be filled in *programmatically* (using some automatic programs or scripts)
- **Spamming** becomes a severe problem for online forms on Websites
- How can we *verify that it is a human* that is filling in the form?



VS.



Turing Test and Reverse Turing Test



reCAPTCHA

What is this?

☐ 私はロボットではありません

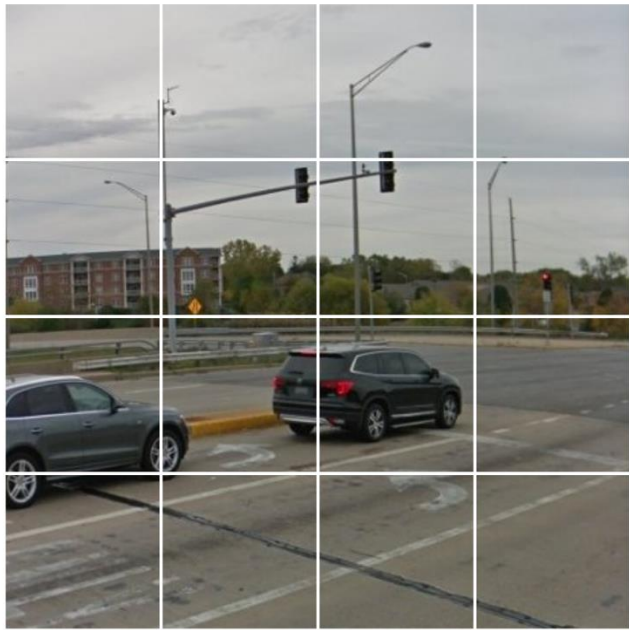
reCAPTCHA
プライバシー - 利用規約

overlooks inquiry

Type the two words:

reCAPTCHA™
stop spam.
read books.

Select all squares with
traffic lights
If there are none, click skip



⌂ 🔊 ⓘ SKIP

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 has strengthened that side of the question. It is signed by 65 Democrats, many of whom occupy respectable 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. Some 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.

Source Document
(Scanned Image)

OCR Transcription

The Hreckinridge and Lane Democrats, having taken courage at the recent eastern advises, are [xxxxxxxxxx] energetically for the campaign: Several prominent Democrats who at first favored DonoLea, are coming out for the other aide, apparently under the [xxxxxxxxxx] of Federal [xxxxxxxxxx]. An address to the National Democracy of [tifornia], urging the party to support HaeeslipsIDas, has recently been published, which manifestly bss strengthened that aide of the [xxxxxxxxxx]: 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, [xxxx] more are recipients of Federal patronage, and the others represent a mass of politicians giving the document [xxxx] [xxxxxx] mTheDcu8las Democrats are also active The Irish and German vote will mostly go with ths# branch of the party, but it is [xxxxxxxxxx] to [xxxxxxxxxx] [xxxxxx] [xxx] [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.

OCR Output
(Text)

reCAPTCHA

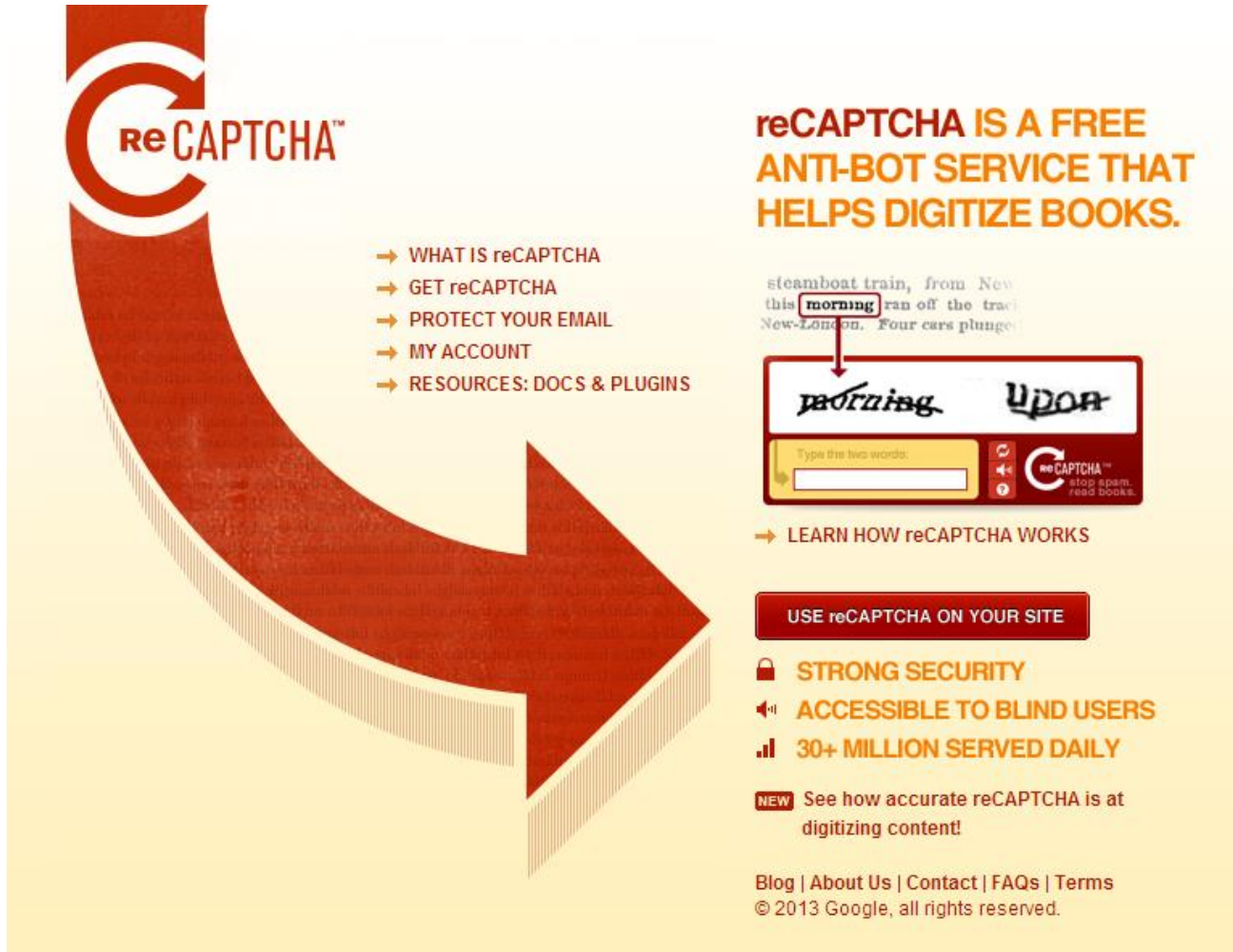
A word selected from the database, the correct answer of which is known
(*control word*)

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



The image shows a landing page for reCAPTCHA. On the left, a large red arrow curves from the top left towards the bottom right. At the top of this arrow is the reCAPTCHA logo, which consists of a red 'C' with a white 'r' inside, followed by the word 'CAPTCHA' in red. To the right of the arrow, there is a list of links: 'WHAT IS reCAPTCHA', 'GET reCAPTCHA', 'PROTECT YOUR EMAIL', 'MY ACCOUNT', and 'RESOURCES: DOCS & PLUGINS'. Further right, the text 'reCAPTCHA IS A FREE ANTI-BOT SERVICE THAT HELPS DIGITIZE BOOKS.' is displayed in orange. Below this, there is a snippet of text from a book: 'steamboat train, from New this morning ran off the track New-London. Four cars plunge'. A red box highlights the word 'morning'. Below the text snippet is a reCAPTCHA interface showing the words 'morning' and 'upon' in a distorted font. A text input field with the placeholder 'Type the two words:' is shown below the words. To the right of the input field are three icons: a red 'C' with a white 'r', a red 'C' with a white 'r', and a red 'C' with a white 'r'. Below the interface, there is a link 'LEARN HOW reCAPTCHA WORKS' and a button 'USE reCAPTCHA ON YOUR SITE'. Below the button are three features: 'STRONG SECURITY' with a lock icon, 'ACCESSIBLE TO BLIND USERS' with a speaker icon, and '30+ MILLION SERVED DAILY' with a bar chart icon. At the bottom, there is a 'NEW' badge followed by the text 'See how accurate reCAPTCHA is at digitizing content!'. At the very bottom, there are links 'Blog | About Us | Contact | FAQs | Terms' and the copyright notice '© 2013 Google, all rights reserved.'

reCAPTCHA™

- WHAT IS reCAPTCHA
- GET reCAPTCHA
- PROTECT YOUR EMAIL
- MY ACCOUNT
- RESOURCES: DOCS & PLUGINS

reCAPTCHA IS A FREE ANTI-BOT SERVICE THAT HELPS DIGITIZE BOOKS.

steamboat train, from New
this **morning** ran off the track
New-London. Four cars plunge

morning upon

Type the two words:

reCAPTCHA™
stop spam.
read books.

→ LEARN HOW reCAPTCHA WORKS

USE reCAPTCHA ON YOUR SITE

- STRONG SECURITY**
- ACCESSIBLE TO BLIND USERS**
- 30+ MILLION SERVED DAILY**

NEW See how accurate reCAPTCHA is at digitizing content!

[Blog](#) | [About Us](#) | [Contact](#) | [FAQs](#) | [Terms](#)
© 2013 Google, all rights reserved.

<http://www.google.com/recaptcha>

reCAPTCHA

CAPTCHA = “**C**ompletely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part”

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)

The screenshot shows the Gwap website interface. At the top is a navigation bar with links: gwap, ESP Game, Tag a Tune, Verbosity, Squigl, Matchin, Flipit, and PopVideo. Below this is a login section with input fields for email and password, a 'Sign In' button, a 'remember me' checkbox, and a 'forgot password?' link. A message states: 'We have partially resumed operations, this message will disappear when all services are online.'

The main content area has a dark blue background with a starry pattern. The headline reads 'Play the Games, Change the Web.' followed by the text 'When you play a game at Gwap, you aren't just having fun.' Below this are two buttons: 'Learn More' and 'Register'. To the left are three large, colorful circles labeled 'Gender Guesser', 'Prizes', and 'Best Images'. On the right, a spotlight effect highlights the 'ESP Game' section, which includes the text 'Concentrate...', 'What do you see in an image?', and 'Do you have ESP?'. Below this is a preview of the ESP Game interface, showing a picture of a tree and a list of words to guess: 'branch', 'trunk', 'leaves', 'tree', 'forest'. A 'PLAY NOW' button is next to the preview.

At the bottom, there is a 'Today's' section with a 'Top 10' list of scores:

Rank	Player	Score
6	guest243271	9,900
7	skuppins	8,225
8	rosie	7,875
9	Lottie	7,850
10	guest243248	7,650

At the very bottom are links for 'Blog', 'About', and 'Contact'.

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

The image shows a screenshot of the ESP Game interface, which is a green-themed web application. At the top, it displays the game title "ESP Game" with a star icon and the instruction "Concentrate...". The current score is 100 and the time is 2:21. The main question is "What do you see?". In the center is a photograph of four sheep in a field. To the left of the image are "taboo words" listed in red: "peace" and "lay". To the right are "guesses" listed in green: "sheeps..." and "sheep". Below the image is a text input field with a cursor, and two buttons: "+ submit" and "→ pass".

Callout 1 (Left):

Sheep
Animal
Grass
...

Callout 2 (Right):

Tree
White
Sheep
...

Callout 3 (Bottom):

'sheep' is a label for this image.

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

YAHOO! JAPAN 知恵袋 ログイン
IDでもっと便利に [新規取得]

[Yahoo! JAPAN - ヘルプ](#)

トップ カテゴリ ▼ 回答ひろば [? いますぐ利用登録!](#)

検索 [条件を指定して検索](#) [質問する](#)

YAHOO! ANSWERS

Search [Web Search](#)

HOME BROWSE CATEGORIES MY ACTIVITY ABOUT

Ask

What would you like to ask?

[Continue](#)

Answer

Share your knowledge, Help others and be an Expert

[Browse Open Questions](#)

Discover

The Best Answers chosen by the Community

[Browse Resolved Questions](#)

What are you looking for? [Search Answers](#)

[Home](#) [Search](#) [Print](#) [Tools](#) [Awards](#) [Help](#) [English](#)

Answers.com® | WikiAnswers®
The world's leading Q&A site

Ask Enter question or phrase... Search: ☒ All sources ☐ Community Q&A ☐ Reference topics

Answer [Go](#)

Browse: [Unanswered questions](#) | [New questions](#) | [New answers](#) | [Reference library](#)

answerbag
Every Question Deserves a Great Answer

Find your answer

Google Custom Search [Search](#)

[Ask a new question or post a poll.](#)

HOME CATEGORIES TOPICS QUESTIONS POLLS DEBATES FORUMS [Connect](#) | [Login](#) / [Register](#)

Ask a question

[Submit](#)

Crowd-sourcing Market Places

Amazon Mechanical Turk

Amazon Mechanical Turk

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

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work



Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Get started.](#)

As a Mechanical Turk Requester you:

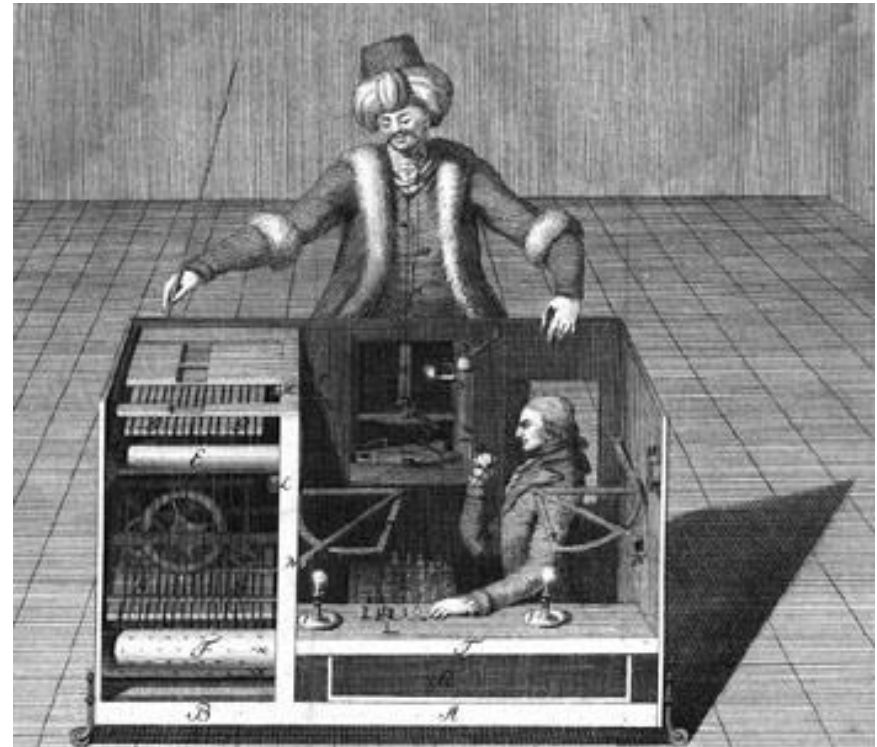
- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results



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 city of the user by reading a list of tweets from that user.
Please pay attention to the *locations, events, activities, organizations, businesses*, mentioned in the tweets. They may 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 in Miami, 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 girls. Orlando, Miami are mentioned. From the context, the user doesn't seem to live in Orlando.
- The Heat play the Philadelphia 76ers.(**hint:**The Heat is a professional basketball team located in Miami, FL. The 76ers is a team from Philadelphia, PA. Combined the hints above, the user is more likely from Miami.)
- 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.



Please **select/crop** company logo or image from the business card above. Click + Drag to select the company logo.

[Click to Zoom & Rotate Image](#)

Name [?](#)

Title Company

Email Website

Address: [?](#)

Address Line 1

[add line](#)

City State Zip Code

Phone: [click here if not a U.S. phone number](#) [?](#)

Work Ext.

Mobile

Fax

[add phone](#)

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?

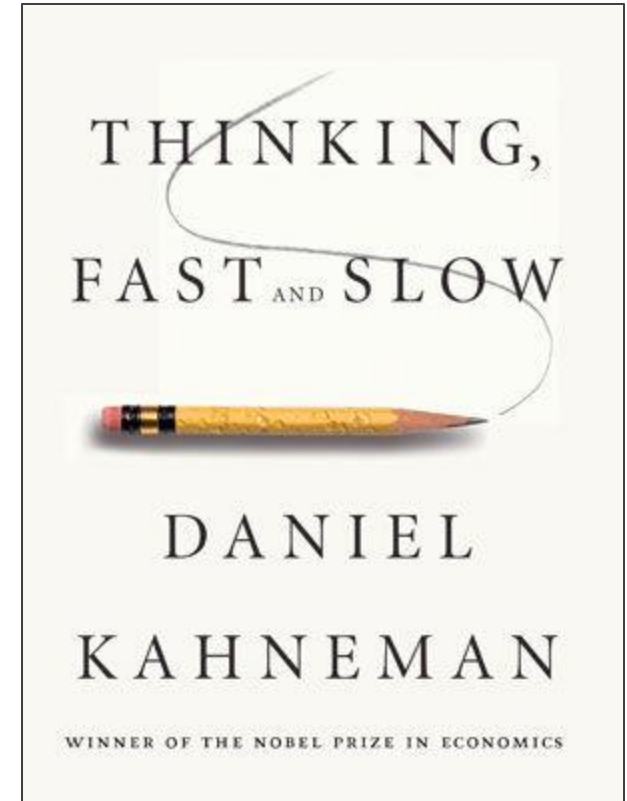
- Human computation and crowd-sourcing tasks usually have a **good** and **meaningful** purpose
- However, people do not work for nothing
- How can we **motivate** or **incentivise** people to take up the work?
 - **Money \$\$\$**
 - **Entertainment (games, e.g. ESP Game, Tag-a-tune)**
 - **Educational (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:
- $P(a) = (18 + 20) / 50 = 0.76$
- $P(A \text{ says 'Yes'}) = (18 + 5) / 50 = 0.46$
 $P(B \text{ says 'Yes'}) = (18 + 7) / 50 = 0.50$
- Therefore,
 $P(e)$
 $= P(\text{Both say 'Yes'}) + P(\text{Both say 'No'})$
 $= (0.46 \times 0.50) + (0.54 \times 0.50) = 0.5$
- Kappa = 0.52

		B	
		Yes	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)

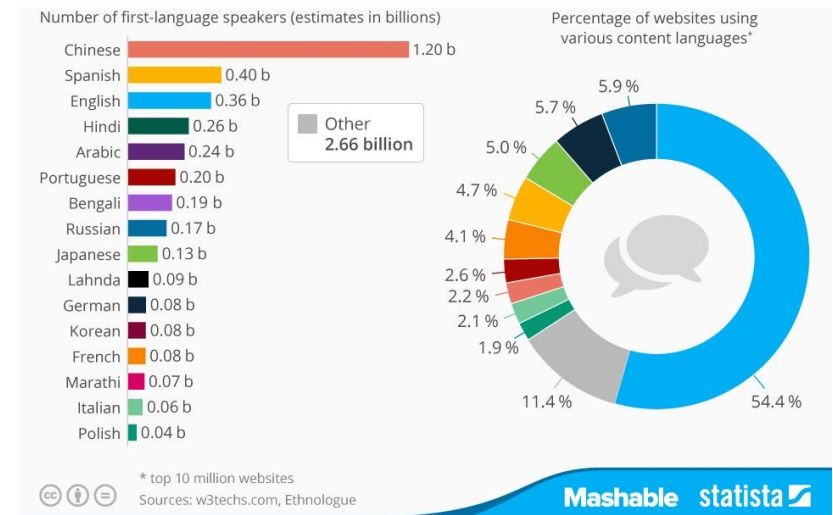
Luis von Ahn: Massive-scale Online Collaboration

<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.

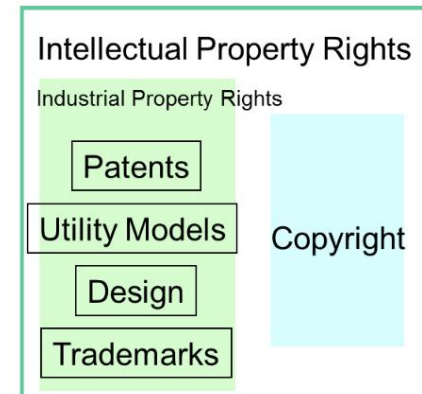
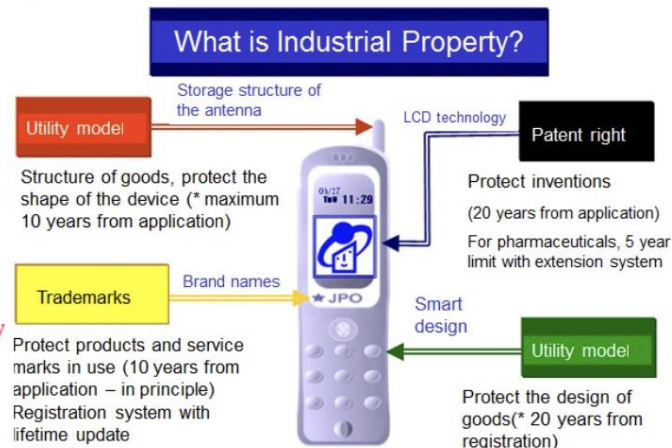


Information Law

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

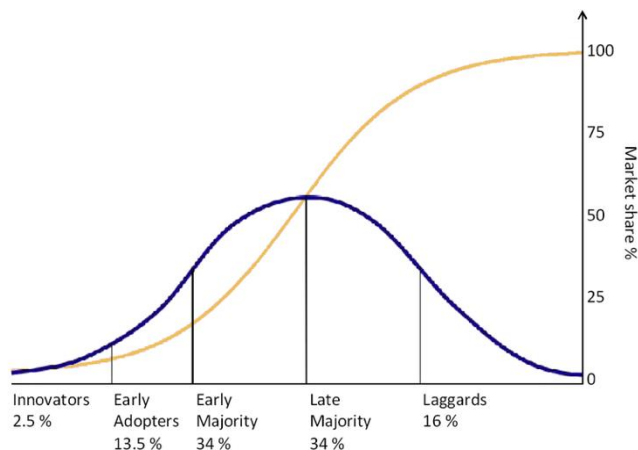
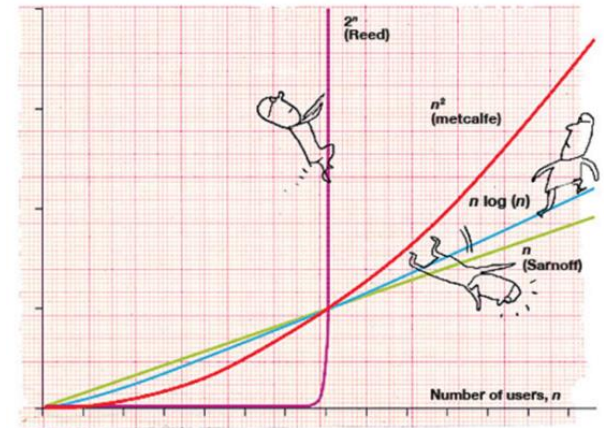
- Can be implemented industrially?
- Is it new or not?
- Could it have been thought of simply?
- Has it been previously applied for?
- Is it not an anti-social invention?
- Are the contents of the invention adequately explained in the specifications document?

Industrial usability, usefulness
Novelty
Non-obviousness, progressivity
Previously applied principle
Sociability ?
Feasibility



Information Economy

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



Hi, Guest | Sign In | Help

Web Images Video Local Shopping News More

YAHOO! kyoto Search Options

Also try: [kyoto protocol](#), [kyoto restaurant](#), [kyoto companion](#), more...

Kyoto - Wikipedia, the free encyclopedia
[History](#) | [Name](#) | [Geography](#) | [Demographics](#)
 Kyoto is a city in the central part of the island of Honshū, Japan. It has a population close to 1.5 million. Formerly the imperial capital of Japan, it is now the capital of Kyoto...
[en.wikipedia.org/wiki/Kyoto](#) - [Cached](#)

Kyoto Travel Guide <Official>
 Kyoto.travel is the first official guide to Kyoto. Please use this site as a gateway to the city, which we have designed to make your visit here that much more enjoyable.
[www.kyoto.travel](#) - [Cached](#)

Kyoto travel guide - Wikitravel
 Kyoto was the capital of Japan for over a millennium, and carries a reputation as its most beautiful city. However, visitors may be surprised by how much work they ...
[wikitravel.org/en/Kyoto](#) - [Cached](#)

Kyoto Protocol Faces Gap After Emissions...
 The world's only treaty that caps greenhouse-gas emissions will face a gap in enforcement after 2012 because its targets expire and an extension can't ...
[www.bloomberg.com/news/2011-06-06/kyoto-protocol-faces...](#) - [Cached](#)

Kyoto City Web

Hotels in Kyoto
 Book last minute Kyoto hotel. Save & book online with Agoda!
[www.agoda.com/Kyoto-Hotels](#)

Hotels in Kyoto
 Book your Hotel in Kyoto online. No reservation costs. Great rates
[Booking.com/Kyoto-Hotels](#)

Aranvert Hotel Kyoto
 Great public bath with panoramic view. Located in center of Kyoto.
[www.aranvert.co.jp](#)

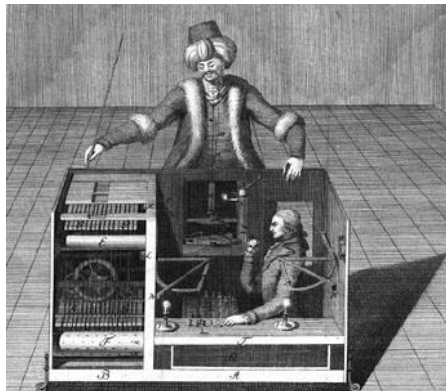
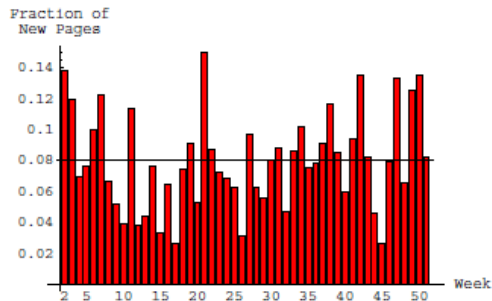
In Kyoto
 Find In Kyotos at Great Prices.
[www.Pronto.com](#)

More Sponsors:
[kyoto](#)

QuickApps
 Safe Search - On
 10,700,000 results for kyoto
 Trending Searches
 kyoto protocol
 kyoto protocol 3
 Related Points of Interest
 Kinkaku-ji
 Heian Shrine
 Ryoanji
 Yasaka Shrine

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)**

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