Human Computation

北京大学 软件与微电子学院
School of Software and Microelectronics, Peking University

Huiping Sun(孙惠平)
sunhp@ss.pku.edu.cn

作业讲解

Why Information Is Hard: An Economic Perspective

Why Information Is Hard

- Abstract
- Introduction
- Network Externalities
- Competitive Applications and Corporate Warfare
- Information Warfare - Offense and Defence
- Distinguishing Good from Bad
- Conclusions

ATM DDOS

Microsoft Philosophy

Passport

Air TPM

Common Criteria

Network Externalities

Asymmetric Information

Moral Hazard

Adverse Selection

Liability Dumping

Tragedy of the Commons

Human Computation

上次课程内容回顾

1 可用安全概念

- 可用性定义
- 可用性评估
- 可用安全起源
- SOUPS
- 可用安全定义

2 理解可用安全

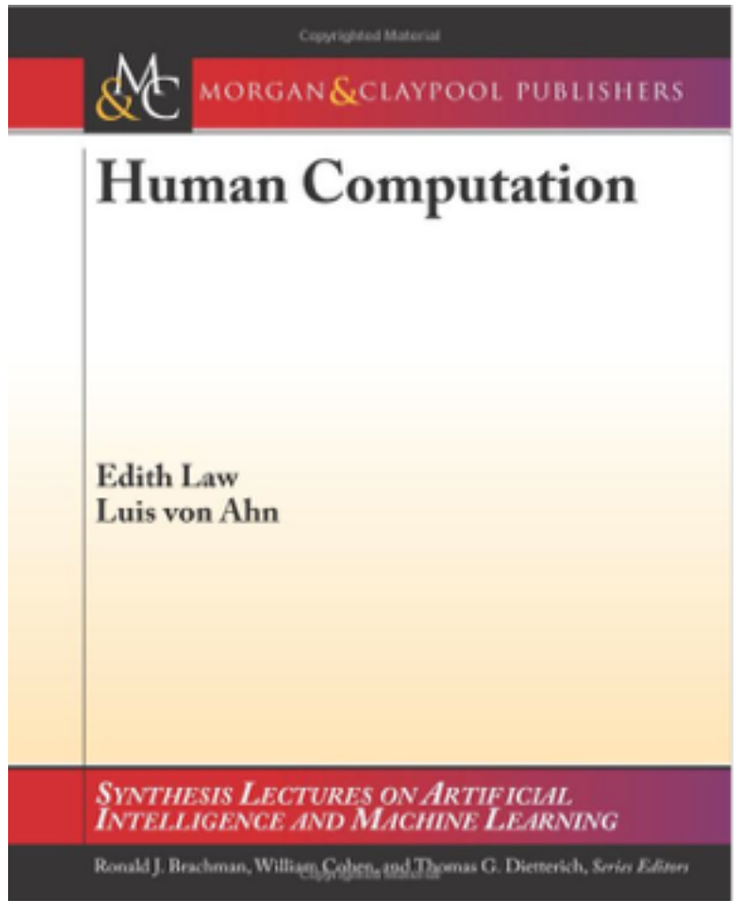
- 为什么需要
- 面临挑战
- 目标
- 安全不可见
- 可理解

3 可用安全例子

- 网络钓鱼
- 证书
- 口令
- PatternLock
- PrivacyBird

参考书

- *Introduction*
- *Human Computation Algorithms*
- *Aggregating Outputs*
- *Task Routing*
- *Understanding Workers and Requesters*
- *The Arts of Asking Questions*
- *Conclusions*



旗舰会议

HCOMP 2018
The sixth AAAI Conference on Human Computation and Crowdsourcing
July 5–8, 2018
Zurich, Switzerland

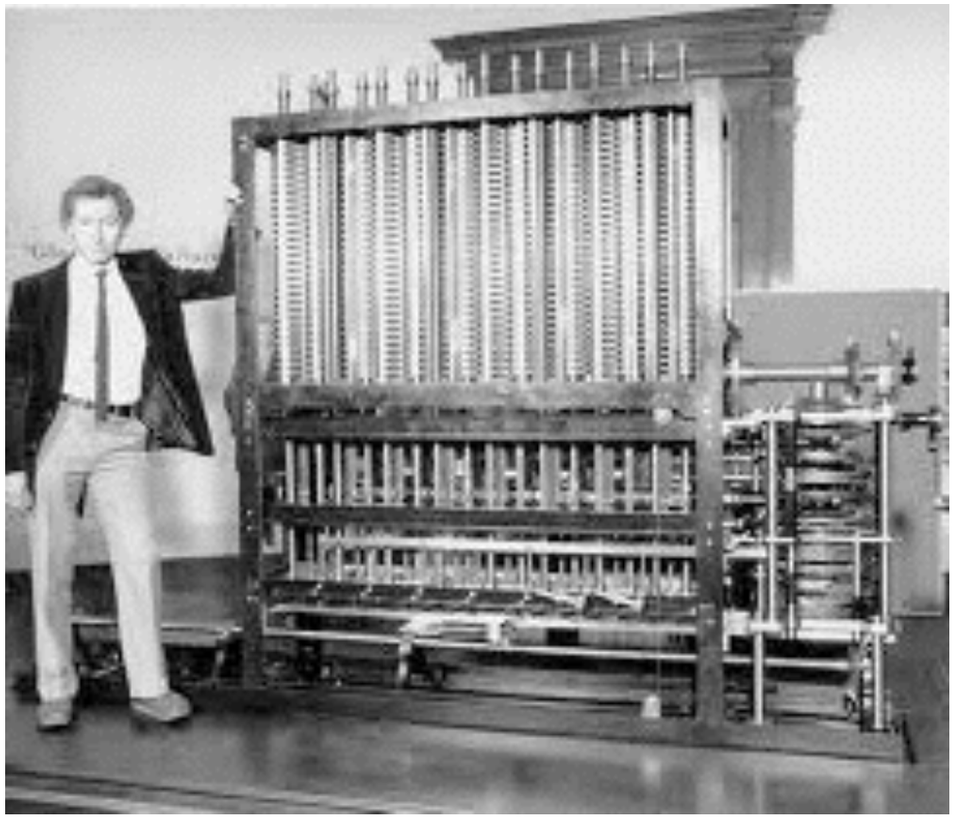
Human Computation
概念

Computation历史

计算公式 → 任务分解

↓

操作指南 → 结果合并



Human Computation

- 计算

  ✳使用算法映射输入到输出的过程

- Human Computation 2005

  ✳人来执行的计算 ⟶ **人工智能难题**

思考

现在有哪些Human Computation应用?

Human Computation产生需要什么基础?

人工智能难题

- 现在依然存在许多人工智能难题
  - ✴人很容易解决
  - ✴但是复杂的计算机算法很难解决

- 常见的人工智能难题
  - ✴感知（目标识别、分类）
  - ✴自然语言分析（观点分析、翻译）
  - ✴认知（计划、推理）

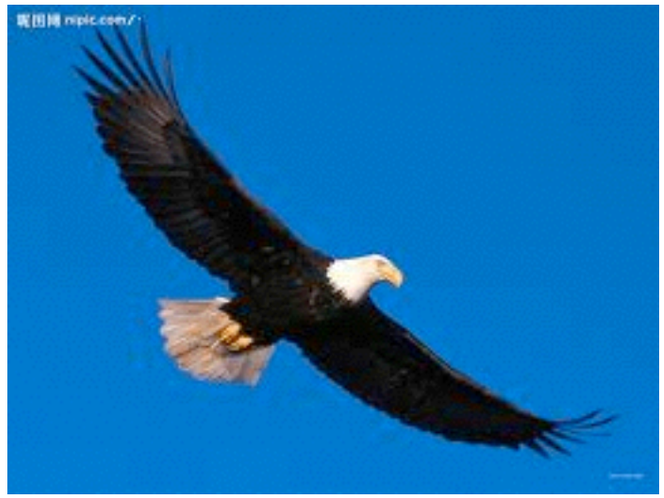
OCR难题

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.

The New-York State Yacht Squadron, on its annual cruise to Newport, came into the harbor yesterday afternoon. The following are the names of the boats that came to anchor here: *Jessie*, *Geraldine*, *Evelyn*, *Annie*, *Mannering*, *Julia*, *Bonita*, *Magic*, *Widgeon*, *Rambler*, *Fleur-de-Lis*, *Henrietta*, *Sea-Drift* and *Maria*, with the steamer *America* as a tender. On anchoring, each boat fired a gun, according to custom. The reports were heard distinctly in the city, causing considerable inquiry as to "what was up," and quite a number of sanguine individuals came into our office to inquire if the guns were not annunciatory signals of the successful laying of the Atlantic Cable. We invariably replied in the negative. The squadron will leave to-day for Newport. The yachts *Washington* and *Rambler*, of this city, start with it, with parties of New-Haven people.

Human Computation

图像识别难题



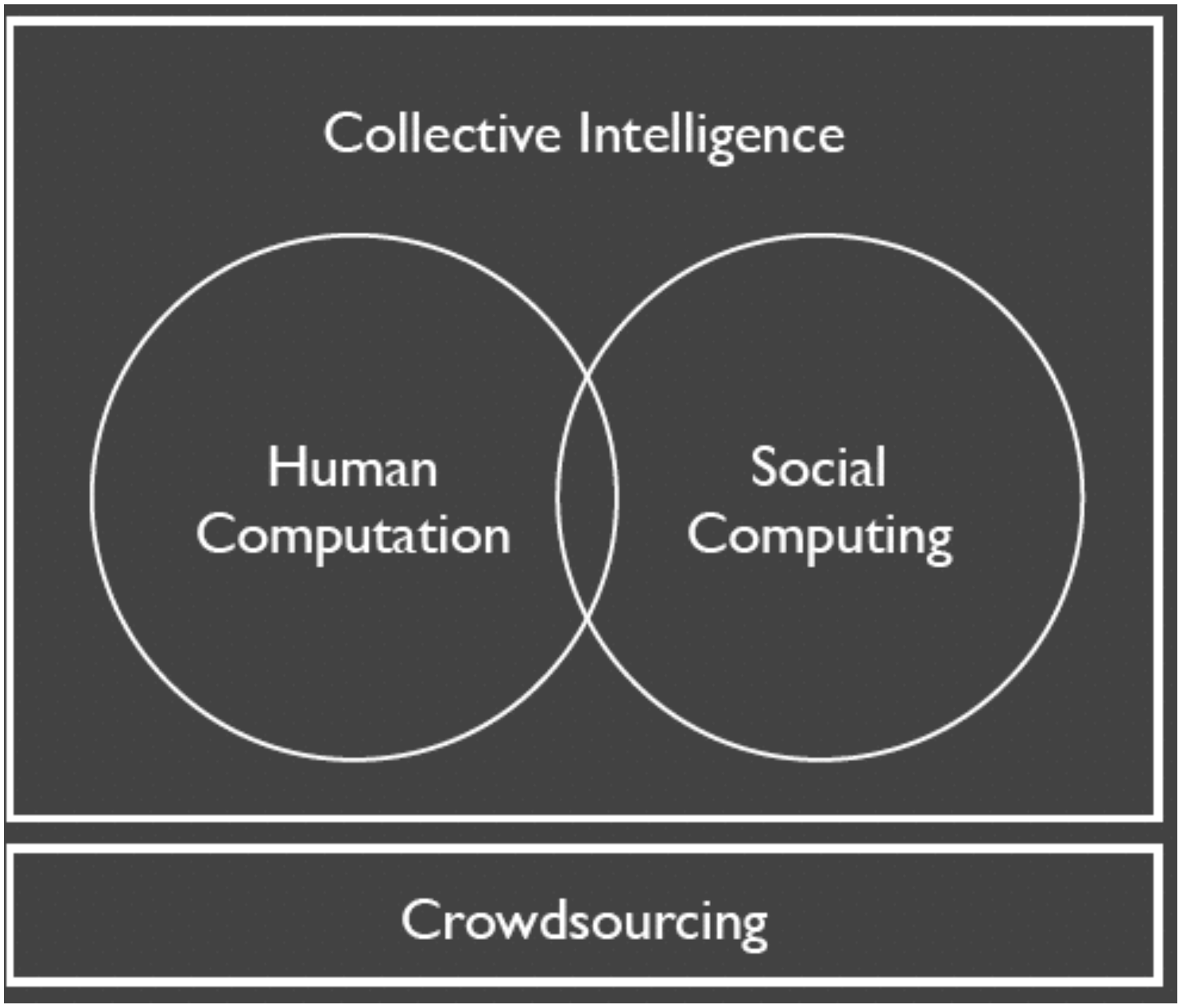
Tag系统　**Q&A系统**

相关概念

动物
机器

社会
行为

工具
方法

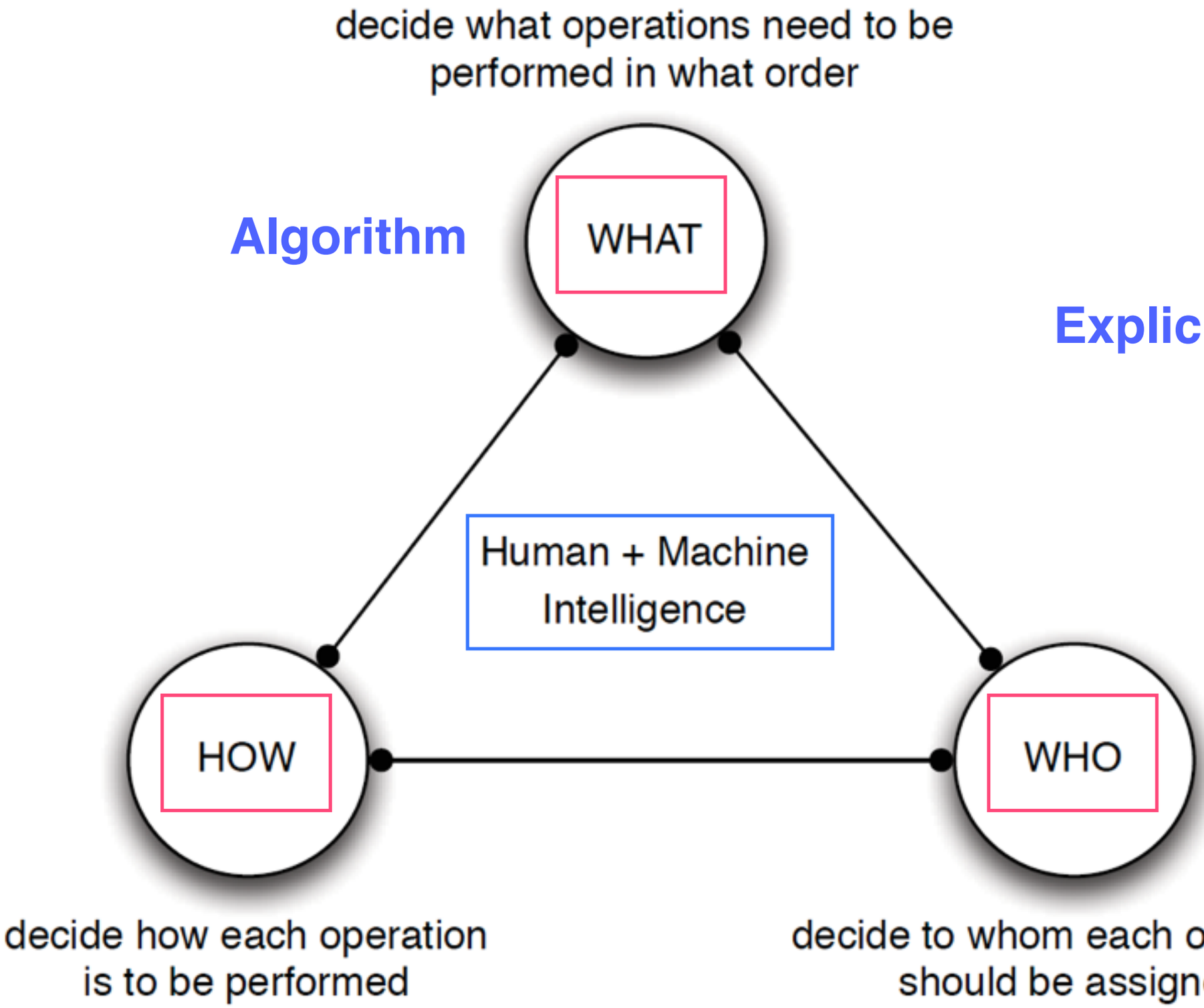


群体
智能

社会
计算

众包

三个层面


decide what operations need to be performed in what order
Algorithm
WHAT
Explicit Control
Human + Machine Intelligence
HOW
WHO
decide how each operation is to be performed
decide to whom each operation should be assigned

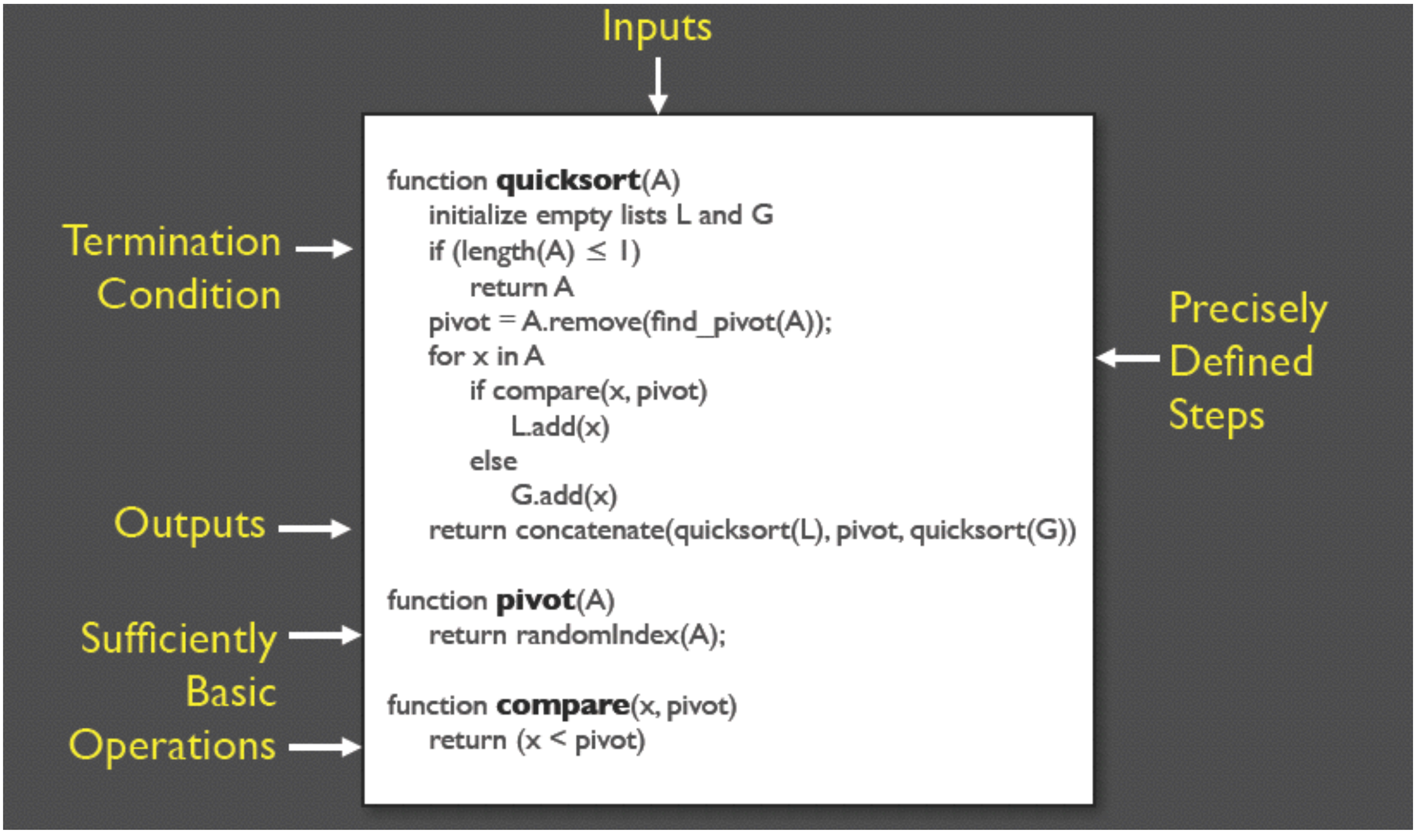
人工智能难题分类

- 识别：图像识别、语音识别, …
- 语言：NLP、翻译, …
- 计算密集问题：象棋、围棋、NP问题、生物学基因测序,…
- 自愿计算：SETI@Home, …
- 其余：医疗诊断、文本编辑、计划…

Human Computation

算法

算法


Inputs
Termination Condition
Outputs
Sufficiently Basic Operations
Precisely Defined Steps
function quicksort(A)
initialize empty lists L and G
if (length(A) ≤ 1)
return A
pivot = A.remove(find_pivot(A));
for x in A
if compare(x, pivot)
L.add(x)
else
G.add(x)
return concatenate(quicksort(L), pivot, quicksort(G))
function pivot(A)
return randomIndex(A);
function compare(x, pivot)
return (x < pivot)

Human Computation

```
function quicksort(A)
    initialize empty lists L and G
    if (length(A) ≤ 1)
        return A
    pivot = A.remove(find_pivot(A));
    for x in A
        if compare(x, pivot)
            L.add(x)
        else
            G.add(x)
    return concatenate(quicksort(L), pivot, quicksort(G))

function pivot(A)
    return randomIndex(A);

function compare(x, pivot)
    return human_compare(x, pivot)
```

Mechanical Turk Task

Instructions

You are shown two images. You must select the image that is more indicative of suspicious activities.

Task

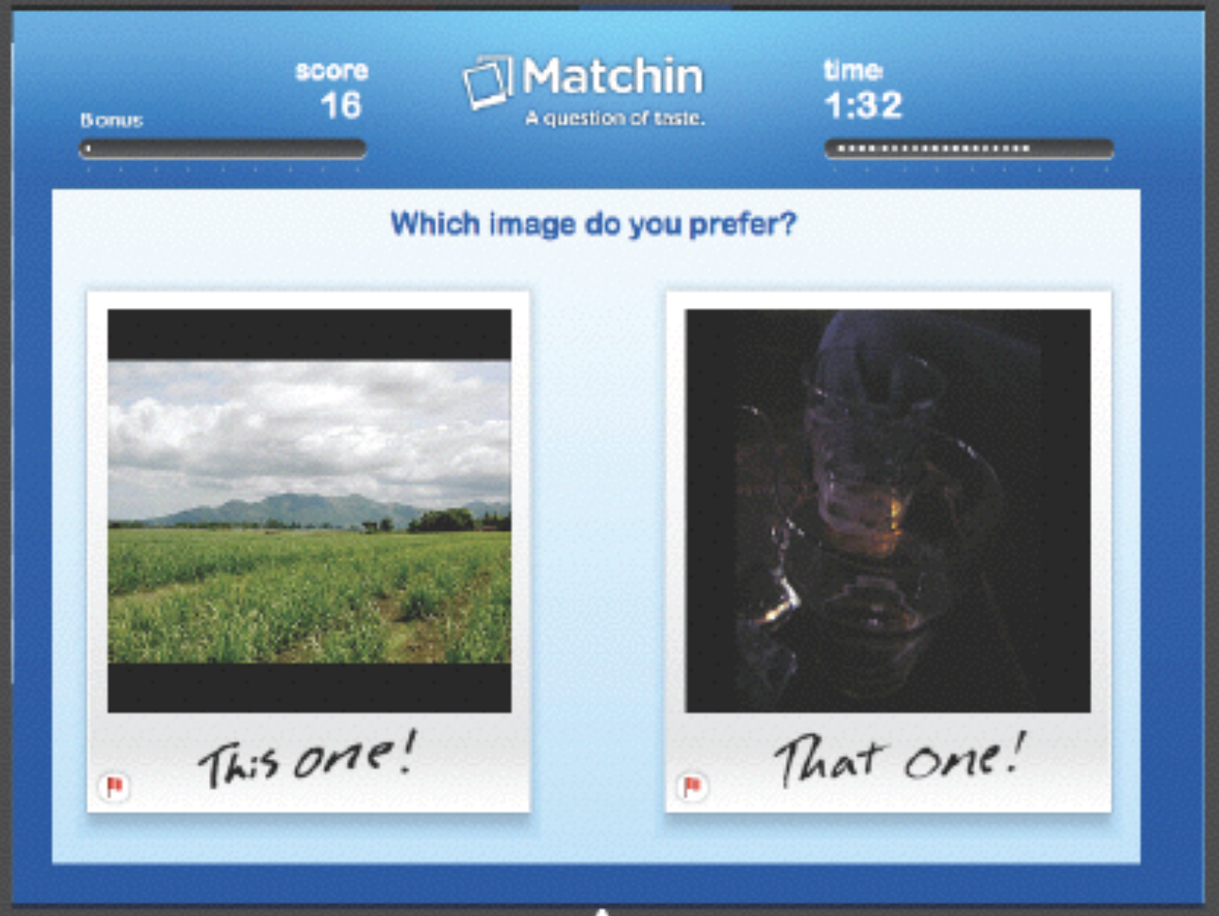
Imagine that you are a security guard and you are monitoring two places. Someone informed you that there are suspicious activities in one of the places, but you were not told which one. Which place will you attend to?

Submit

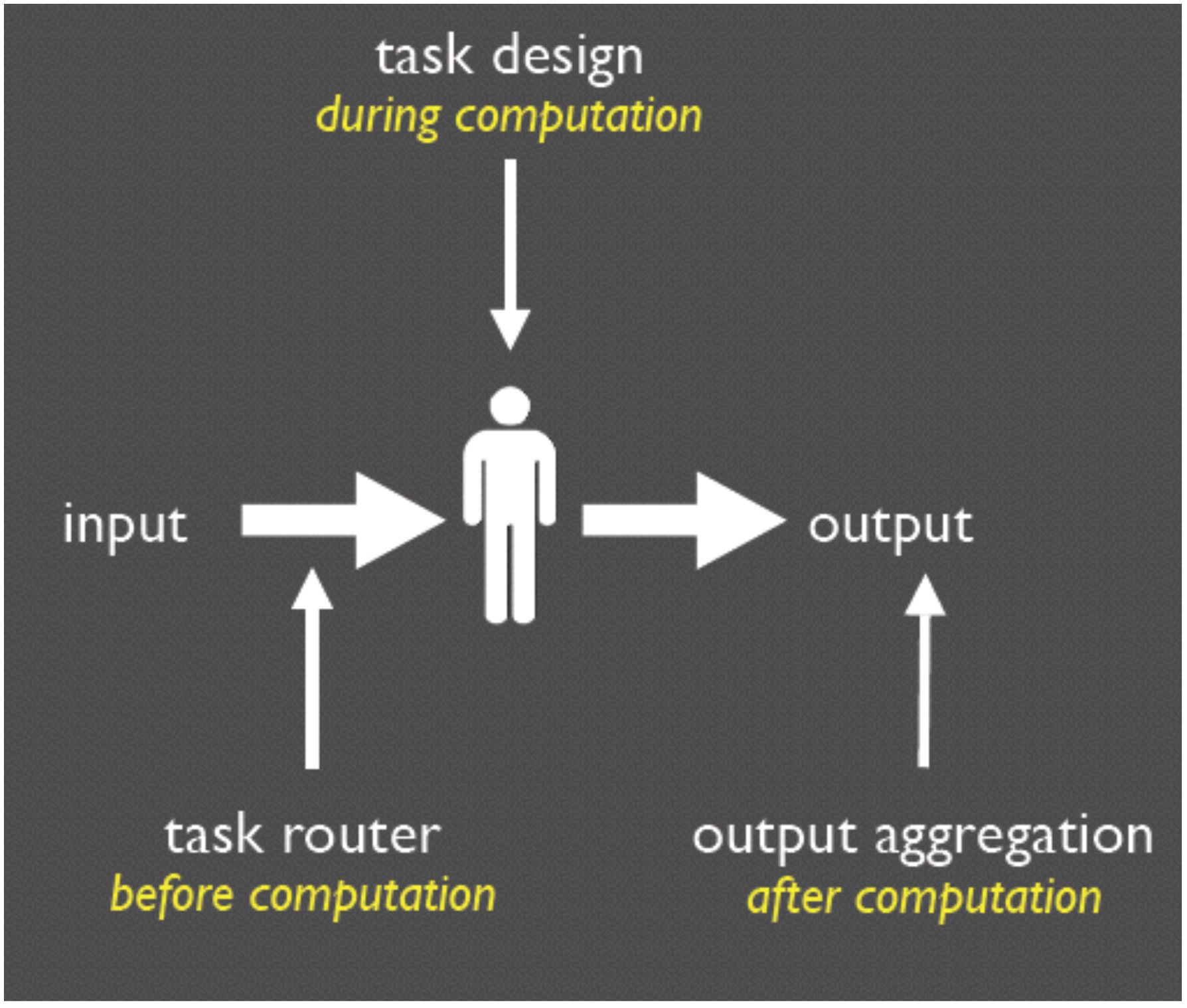
Human Computation

```
function quicksort(A)
    initialize empty lists L and G
    if (length(A) ≤ 1)
        return A
    pivot = A.remove(find_pivot(A));
    for x in A
        if compare(x, pivot)
            L.add(x)
        else
            G.add(x)
    return concatenate(quicksort(L), pivot, quicksort(G))

function pivot(A)
    return randomIndex(A);

function compare(x, pivot)
    return human_compare(x, pivot)
```

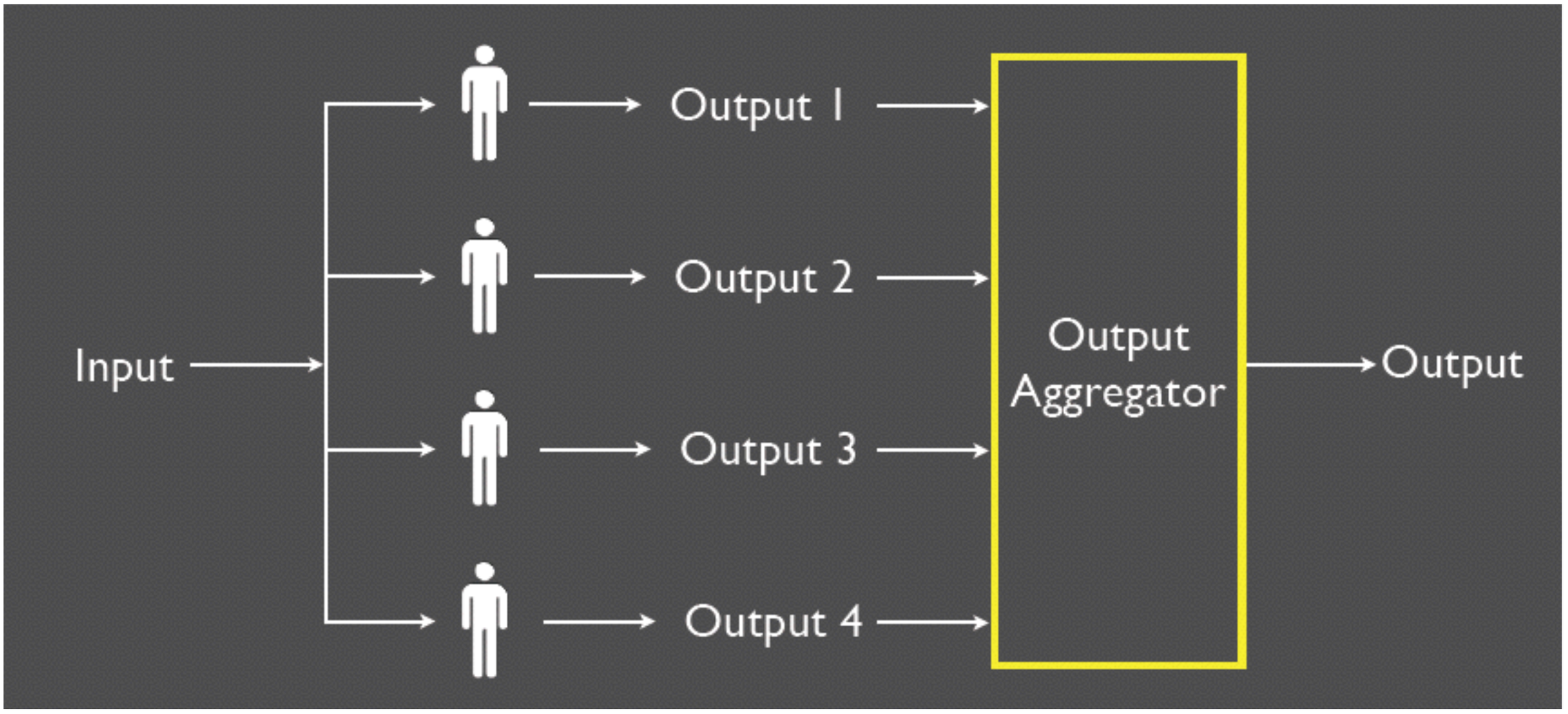
Games with a Purpose



算法正确性


task design
during computation
input
output
task router
before computation
output aggregation
after computation

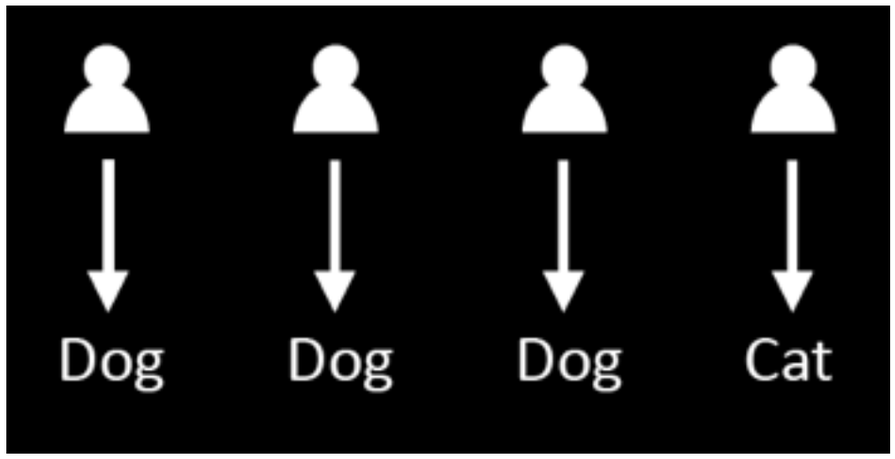
结果汇集

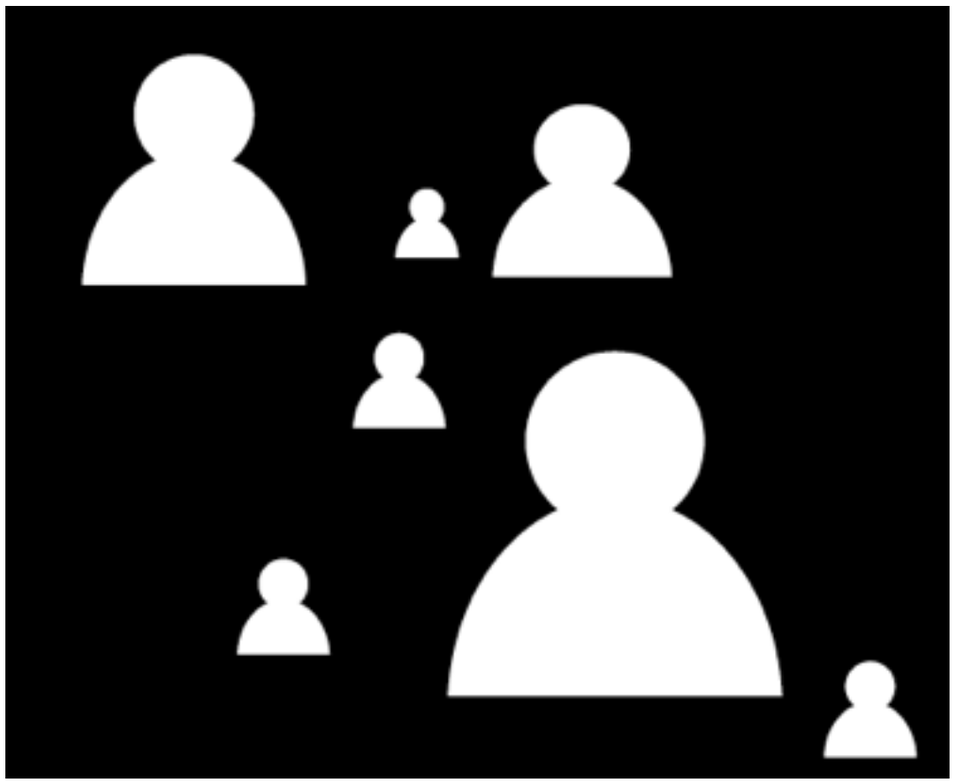

Input
Output 1
Output 2
Output 3
Output 4
Output
Aggregator
Output

参与动机

- Money
- Access
- Game
- Volunteer
- Learning

正确性





Tbe	Kiag	himself	directed	tlie	operations
The	King	birnsell	directed	tlie	operations
Tbe	King	bimself	directed	tlie	operations
The	King	lala		the	
		himself		the	

Human Computation
例子

ESP: Web图形识别

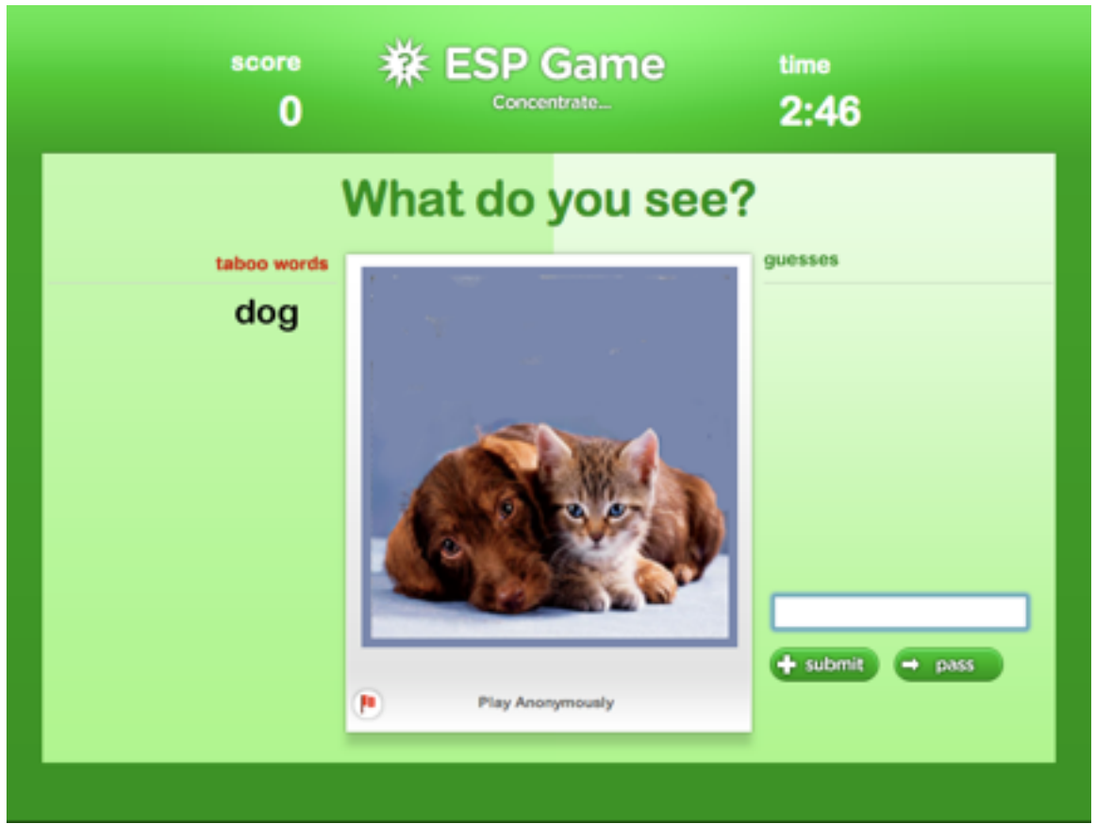
- Web上的图像识别是一个主要的技术挑战
- 大量图片存在，但是文本描述很少，自动识别很不准确
- 人来做标记是一个无奈的选择

游戏

- 每周现在有20亿用户玩在线游戏
- 21岁美国人万游戏时间，平均一生5年

ESP游戏

- 两个用户同时独立的标记一个图片
- 如果标记一致会得到奖励



Player 1 guesses: purse
Player 1 guesses: bag
Player 1 guesses: brown

Success! Agreement on “purse”

Player 2 guesses: handbag

Player 2 guesses: purse
Success! Agreement on “purse”

Citizen Science

- 科学是数据密集型的
  - ✴气候
  - ✴物种分布
  - ✴星体轨迹
- 科学家数量少
- 非科学家来收集和解释数据

Galaxy Zoo

Galaxy Zoo: Hubble
www.galaxyzoo.org
Galaxy Zoo is a ZOONIVERSE project
...just like MOON ZOO
GALAXY ZOO
HUBBLE
Home
The Story So Far
How To Take Part
Classify Galaxies
Explore Galaxies
The Science
FAQ
Forum
Blog
Contact Us
Pictures
You aren't logged in so we can't record your results!
Welcome to Galaxy Zoo, where you can help astronomers explore the Universe
Galaxy Zoo: Hubble uses gorgeous imagery of hundreds of thousands of galaxies drawn from NASA's Hubble Space Telescope archive. To understand how these galaxies, and our own, formed we need your help to classify them according to their shapes — a task at which your brain is better than even the most advanced computer. If you're quick, you may even be the first person in history to see each of the galaxies you're asked to classify.
More than 250,000 people have taken part in Galaxy Zoo so far, producing
Classifier Log In
Click here to log in
Register
Forgotten Password?
Explore galaxies
Enter a search term
Search

Human Computation

eBird

eBird News and Features

ebird.org/content/ebird/

eBird

Home | About eBird | Submit Observations | View and Explore Data | My eBird

Sign In | Register as a New User

Translate to: English

Welcome to eBird

eBird is as birder does!

eBird News and Features

eBirder of the Month -- Mary Gustafson
November 12, 2011

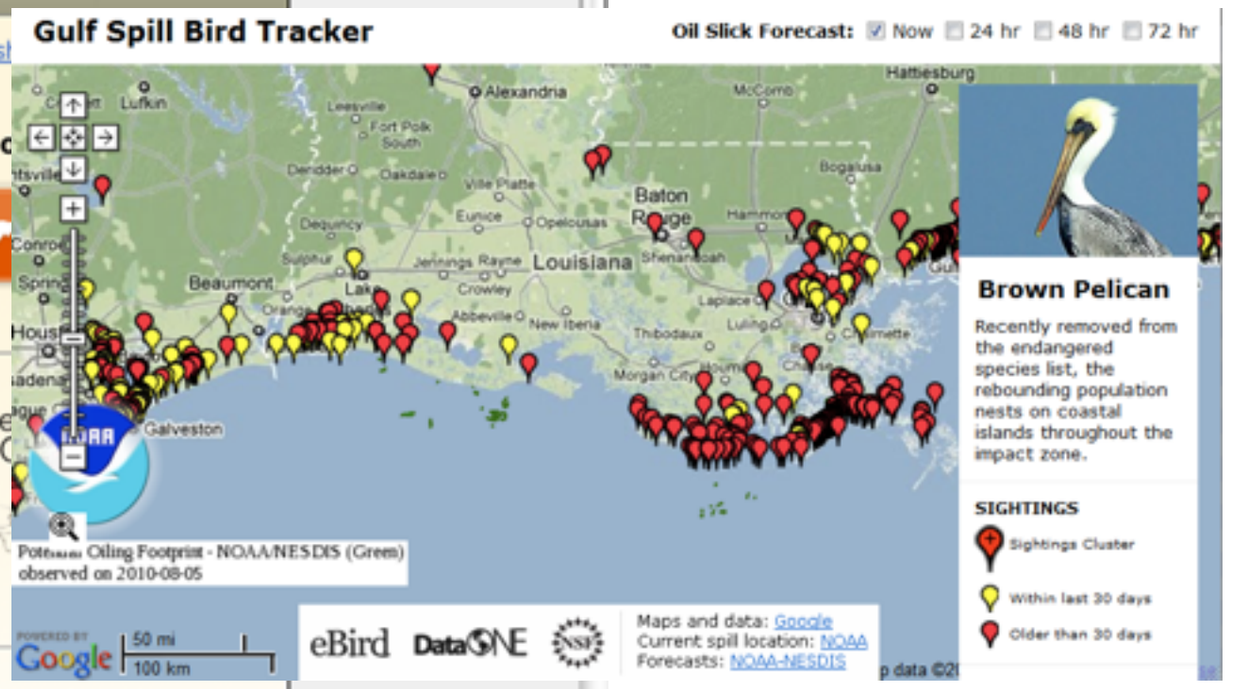
The Rio Grande Valley Birding Festival is taking place this weekend, which means that our current November eBirder of the Month will be buzzing around Harlingen corralling teams of leaders, making sure field trips get off on time, and generally making sure that the 2011 festival comes off as another huge success. Sadly, all this means that Mary won't be getting out in the field as much as she would like. So it seems like an appropriate time to recognize her not only for her hard work as field trip chair for this festival, but also for her unflagging commitment to eBird, both as a user and a reviewer. Moreover, Mary has solidified the Rio Grande Valley Birding festival's commitment to eBird by ensuring that all field trip data are entered into the system, and that all festival participants become aware of the project. This adds up to hundreds of new eBirders each year thanks to this festival! Thanks Mary, and be sure to get out (e)Birding again when things settle down on Monday!

Possible hybrid Harlan's Red-tailed Hawk x Rough-legged Hawk in Kansas
November 07, 2011

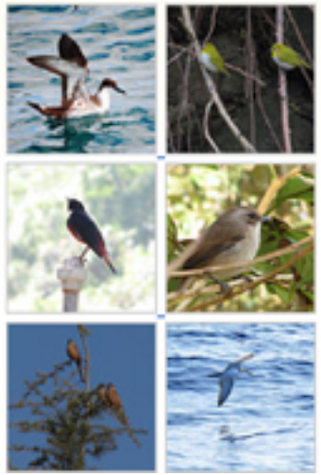
In North America, Red-tailed and Rough-legged Hawks

Gulf Spill Bird Tracker

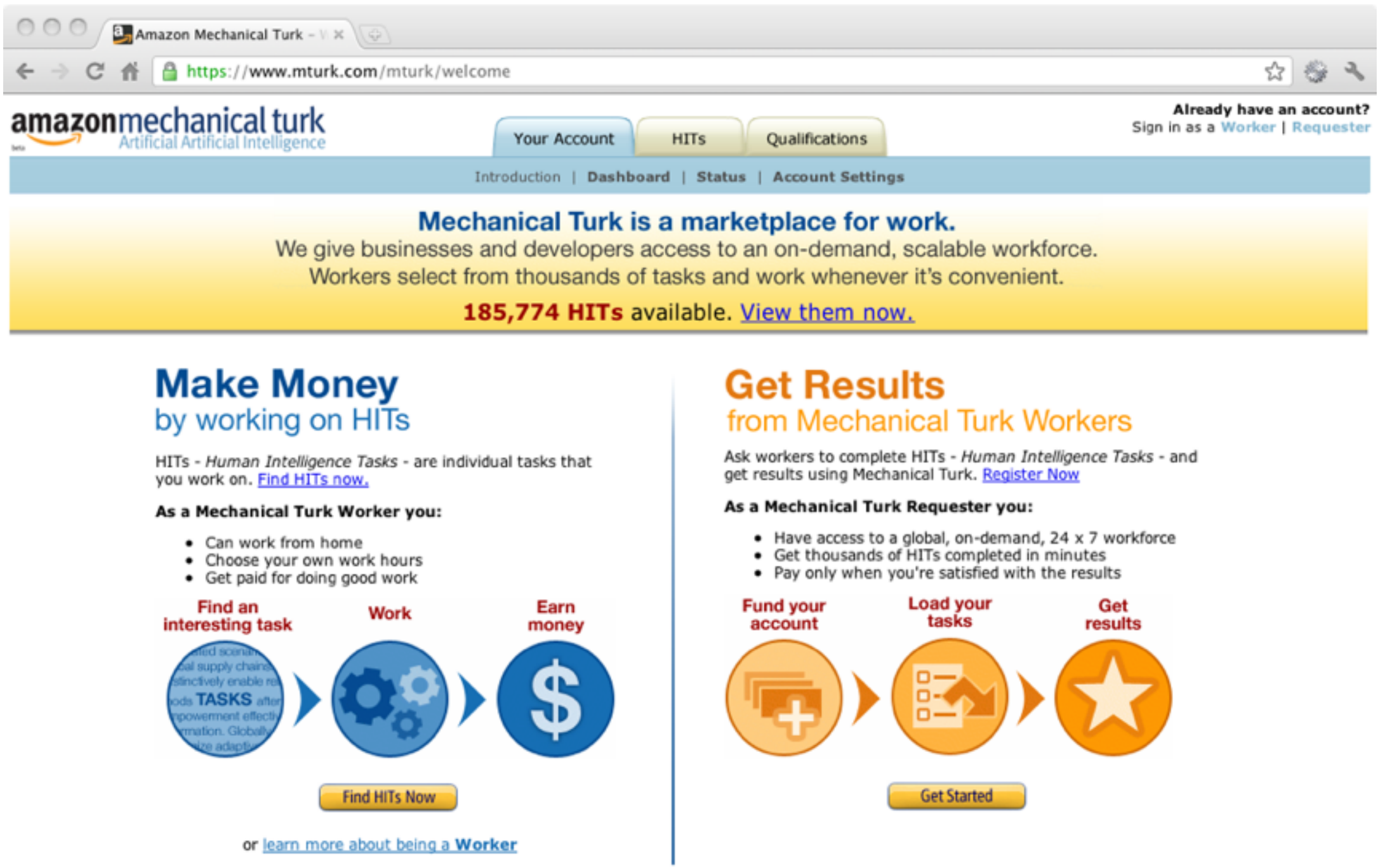
Oil Slick Forecast: Now 24 hr 48 hr 72 hr



eBird Rarity Photos Pool



Amazon Mechanical Turk



Amazon Mechanical Turk

amazonmechanical turk beta
Artificial Artificial Intelligence

Your Account | HITs | Qualifications

253,412 HITs available now

All HITs | **HITs Available To You** | **HITs Assigned To You**

Find HITs **containing** **that pay at least $** 0.00

All HITs

1-10 of 32004 Results

Sort by: HITs Available (most first) GO!

Show all details | Hide all details

Classify Arabic Tweets Dialects SEE REVISED HIT

Requester: Chris Callison-Burch **HIT Expiration Date:** Aug 4, 2013 (9 weeks 6 days) **Reward:** $0.00
Time Allotted: 60 minutes **HITs Available:** 23530

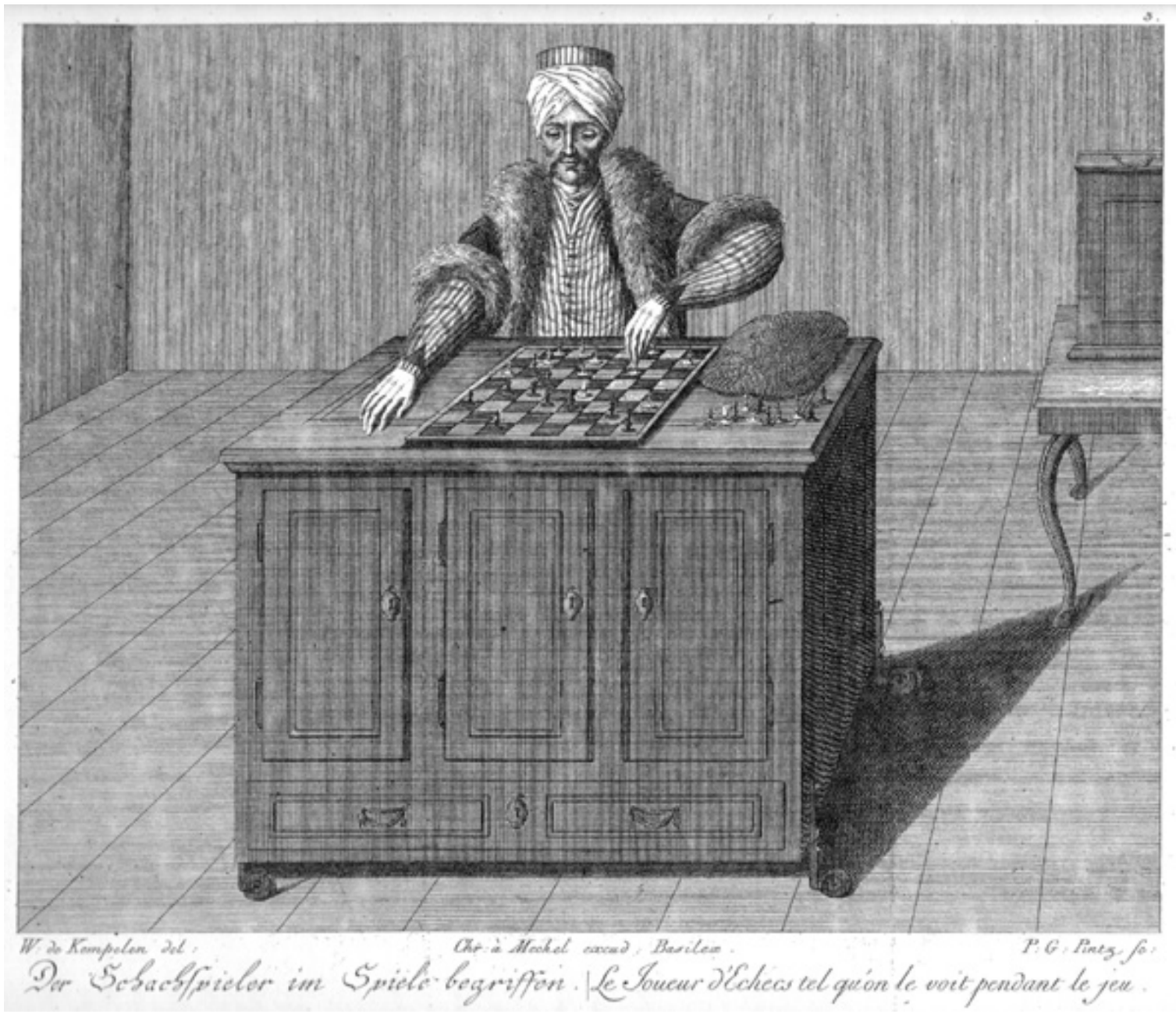
Identify whether the phrase/keywords belongs to the category provided?

Requester: InterestProfiler **HIT Expiration Date:** Jun 21, 2013 (3 weeks 4 days) **Reward:** $0.01
Time Allotted: 10 minutes **HITs Available:** 17969

Search: Keywords on Google.com (US)

Requester: CrowdSource **HIT Expiration Date:** May 27, 2014 (52 weeks) **Reward:** $0.08
Time Allotted: 16 minutes **HITs Available:** 14969

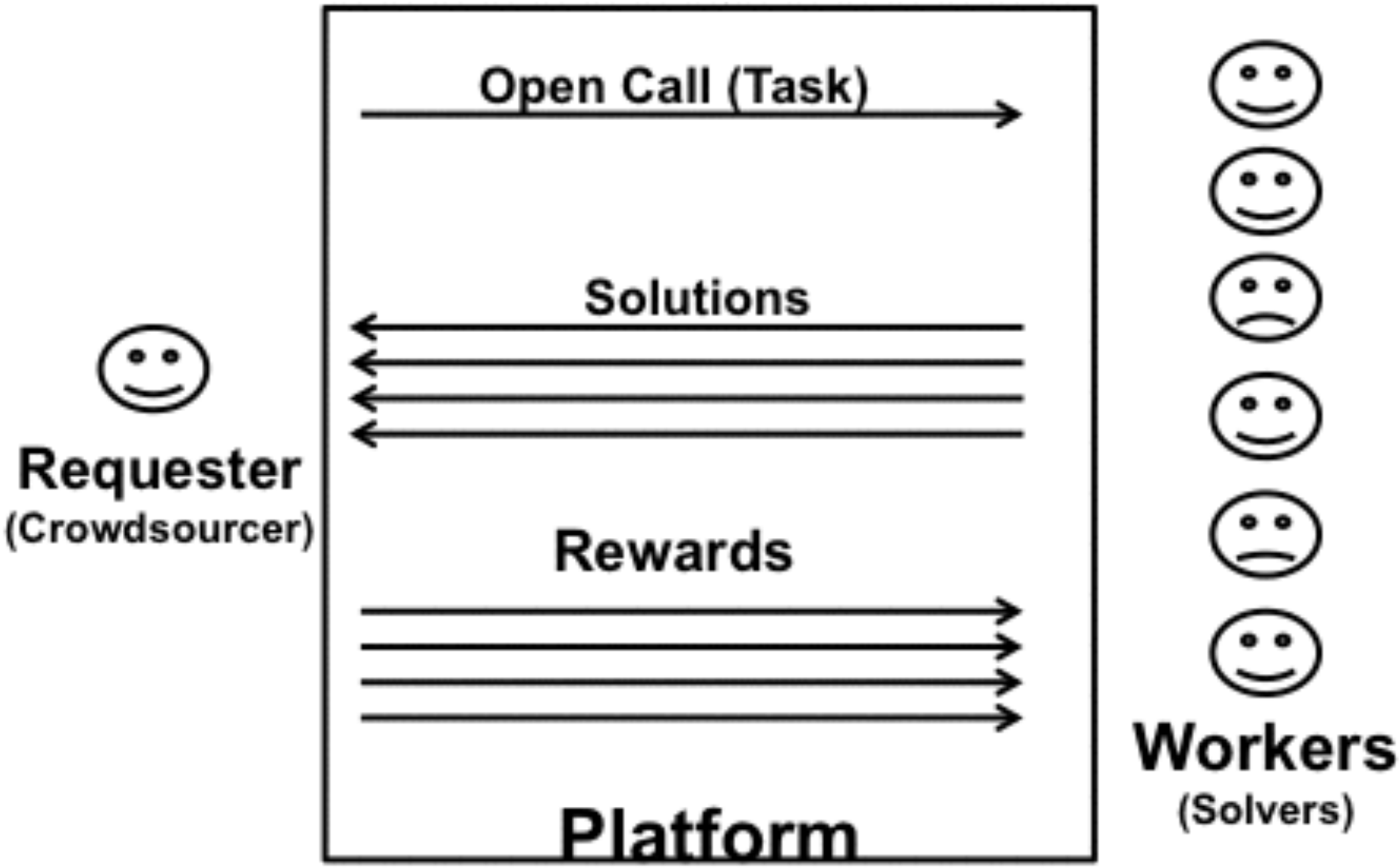
Mechanical Turk



W. de Kempelen del. Chr. à Mechel excud. Basileæ. P. G. Pintz, sc.

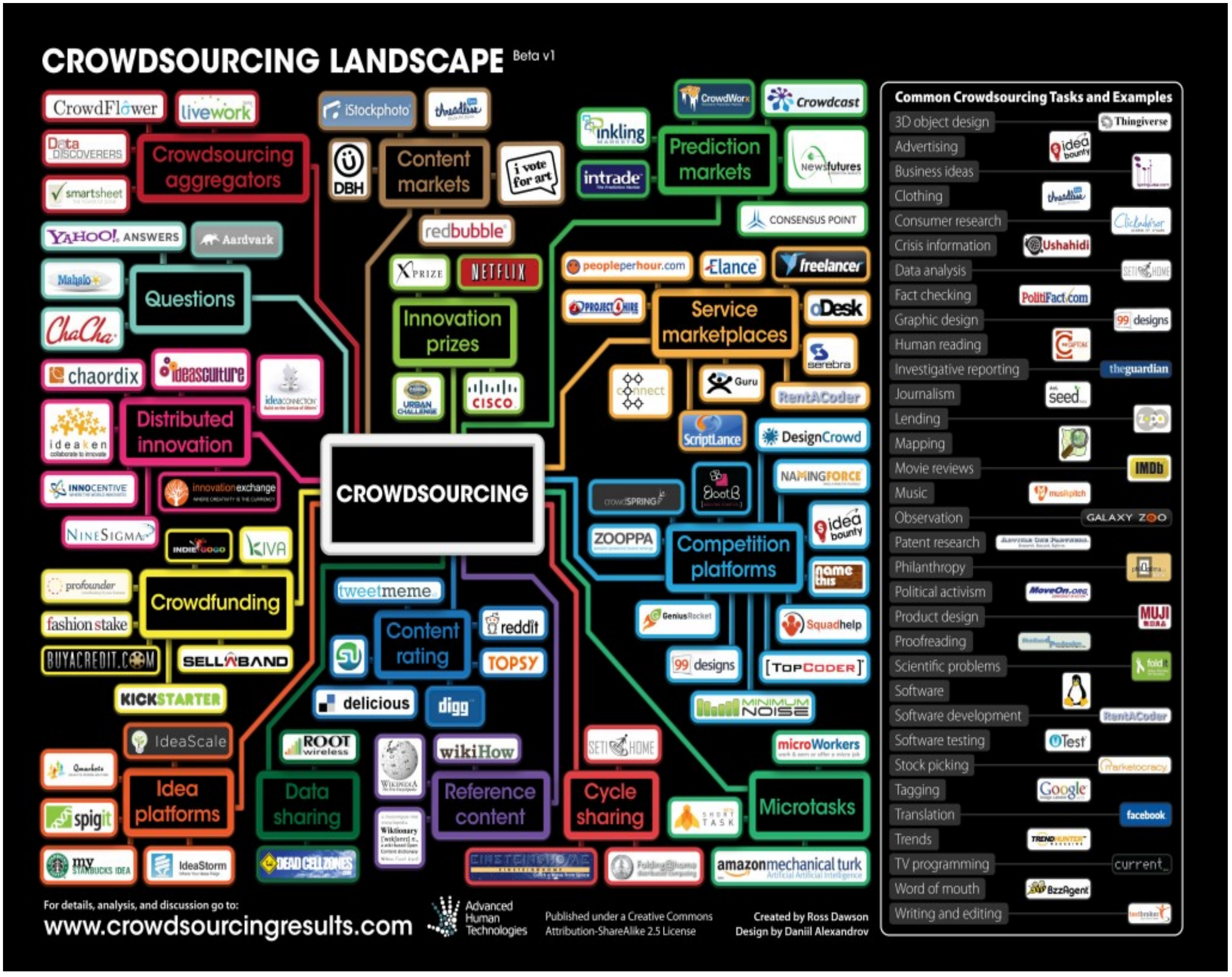
Der Schachspieler im Spiele begriffen. / Le Joueur d'Echecs tel qu'on le voit pendant le jeu.

众包

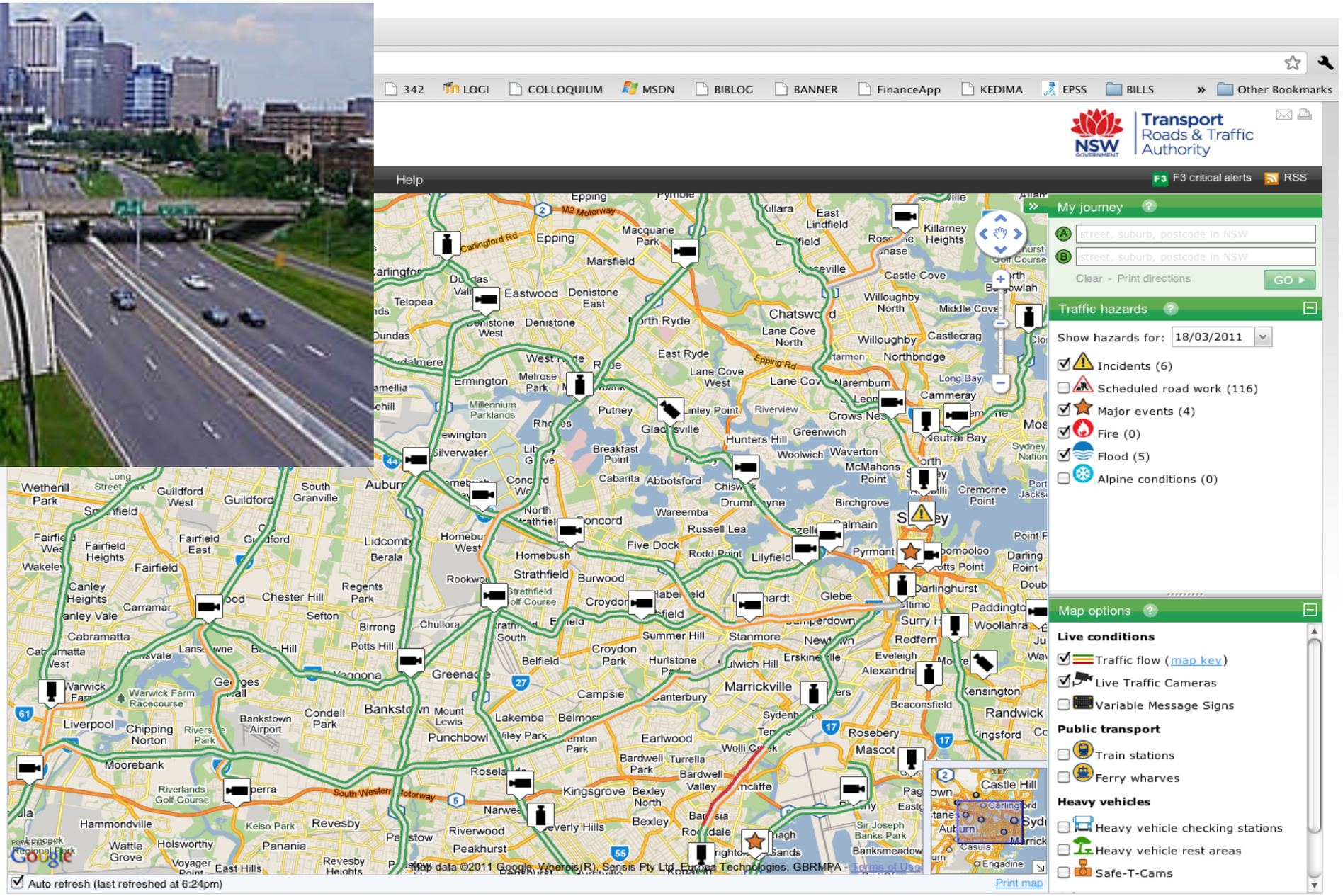

Open Call (Task)
Solutions
Rewards
Platform
Requester
(Crowdsourcer)
Workers
(Solvers)

Human Computation

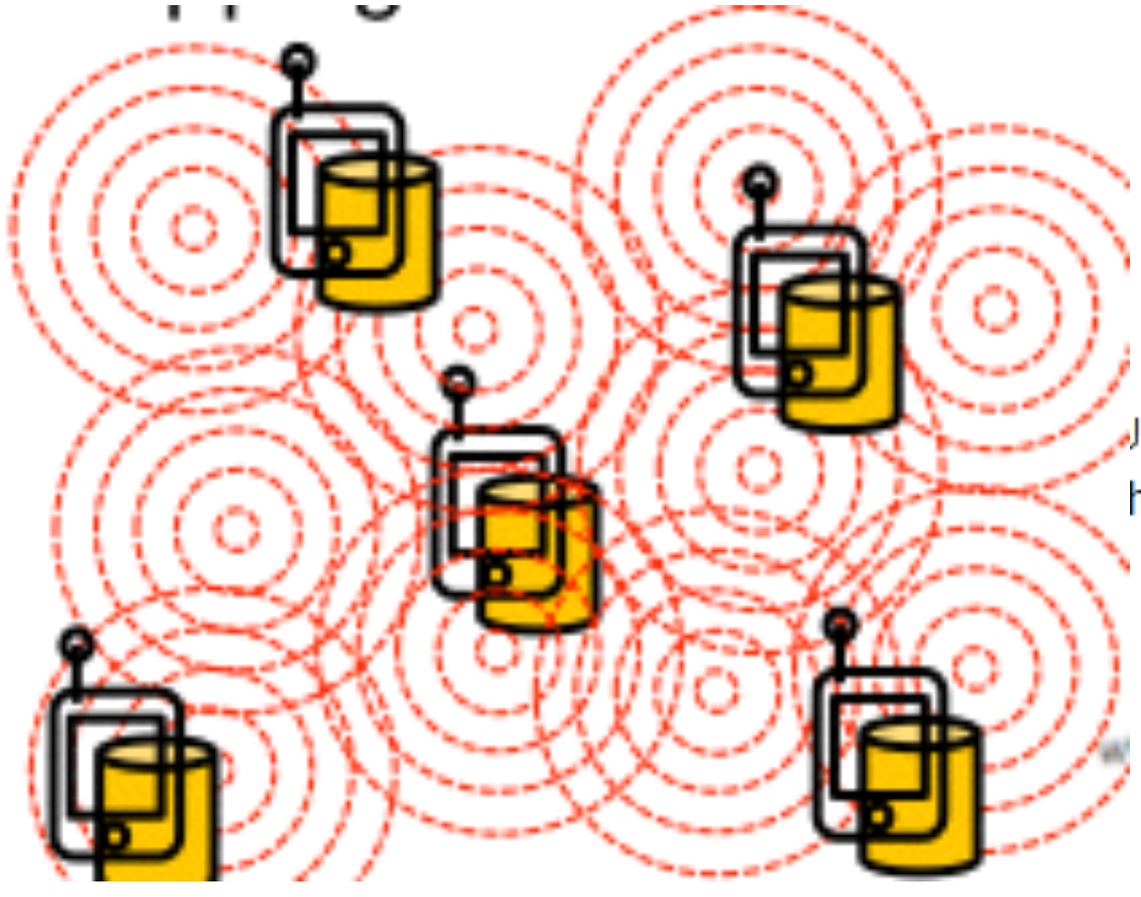
众包

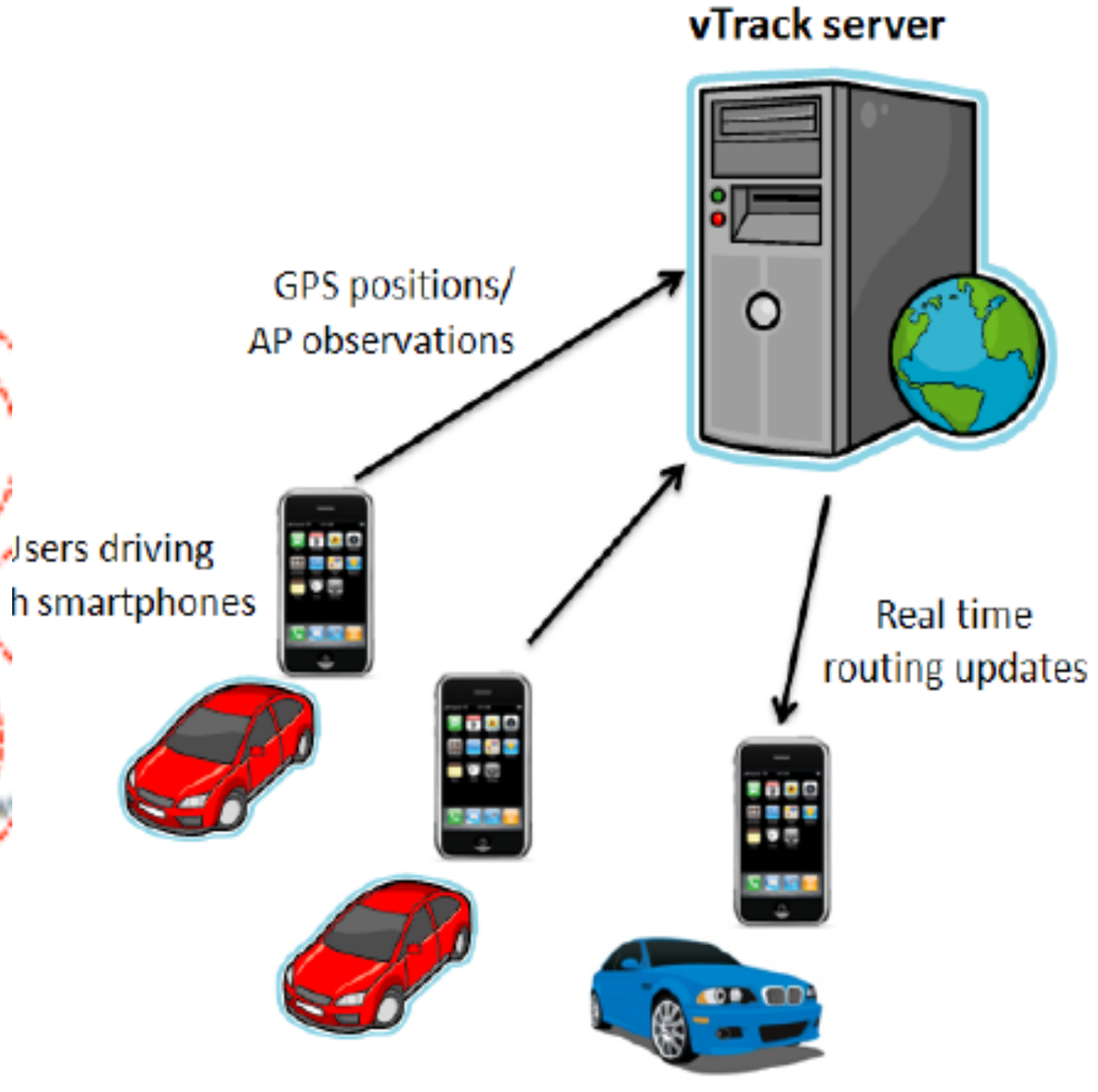


道路监控

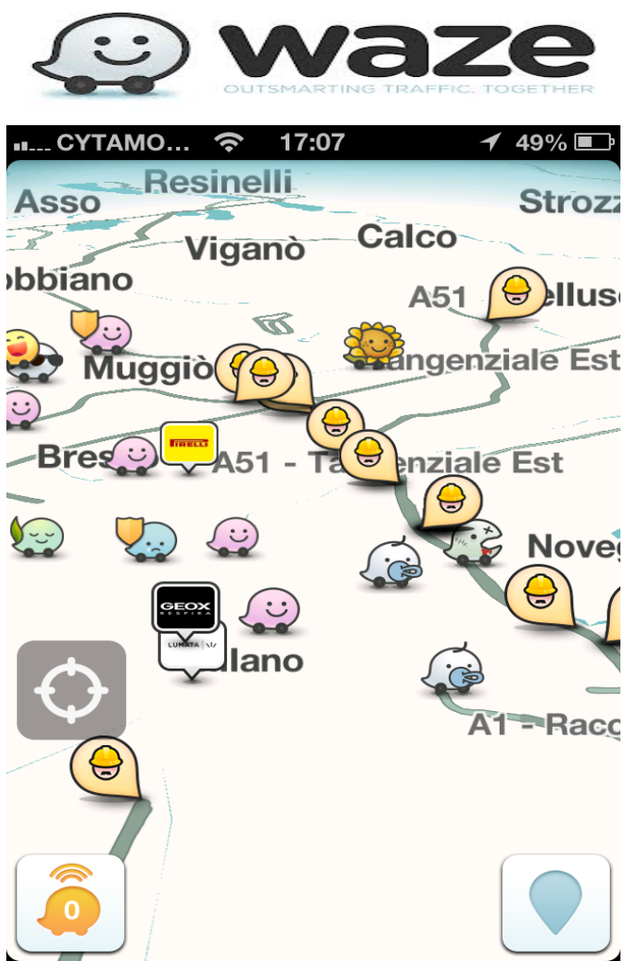

342
LOGI
COLLOQUIUM
MSDN
BIBLOG
BANNER
FinanceApp
KEDIMA
EPSS
BILLS
Other Bookmarks
Transport Roads & Traffic Authority
NSW GOVERNMENT
Help
F3 F3 critical alerts
RSS
My journey
street, suburb, postcode in NSW
Clear - Print directions
GO
Traffic hazards
Show hazards for: 18/03/2011
Incidents (6)
Scheduled road work (116)
Major events (4)
Fire (0)
Flood (5)
Alpine conditions (0)
Map options
Live conditions
Traffic flow (map key)
Live Traffic Cameras
Variable Message Signs
Public transport
Train stations
Ferry wharves
Heavy vehicles
Heavy vehicle checking stations
Heavy vehicle rest areas
Safe-T-Cams
Auto refresh (last refreshed at 6:24pm)
Print map
Map data ©2011 Google, Whereis(R), Sensis Pty Ltd, Europa Technologies, GBRMPA - Terms of Use
Privacy | Copyright | Terms | Sitemap
131700 Report a traffic incident • 132701 RTA traffic information line • © Roads and Traffic Authority (NSW) 2010. All rights reserved.

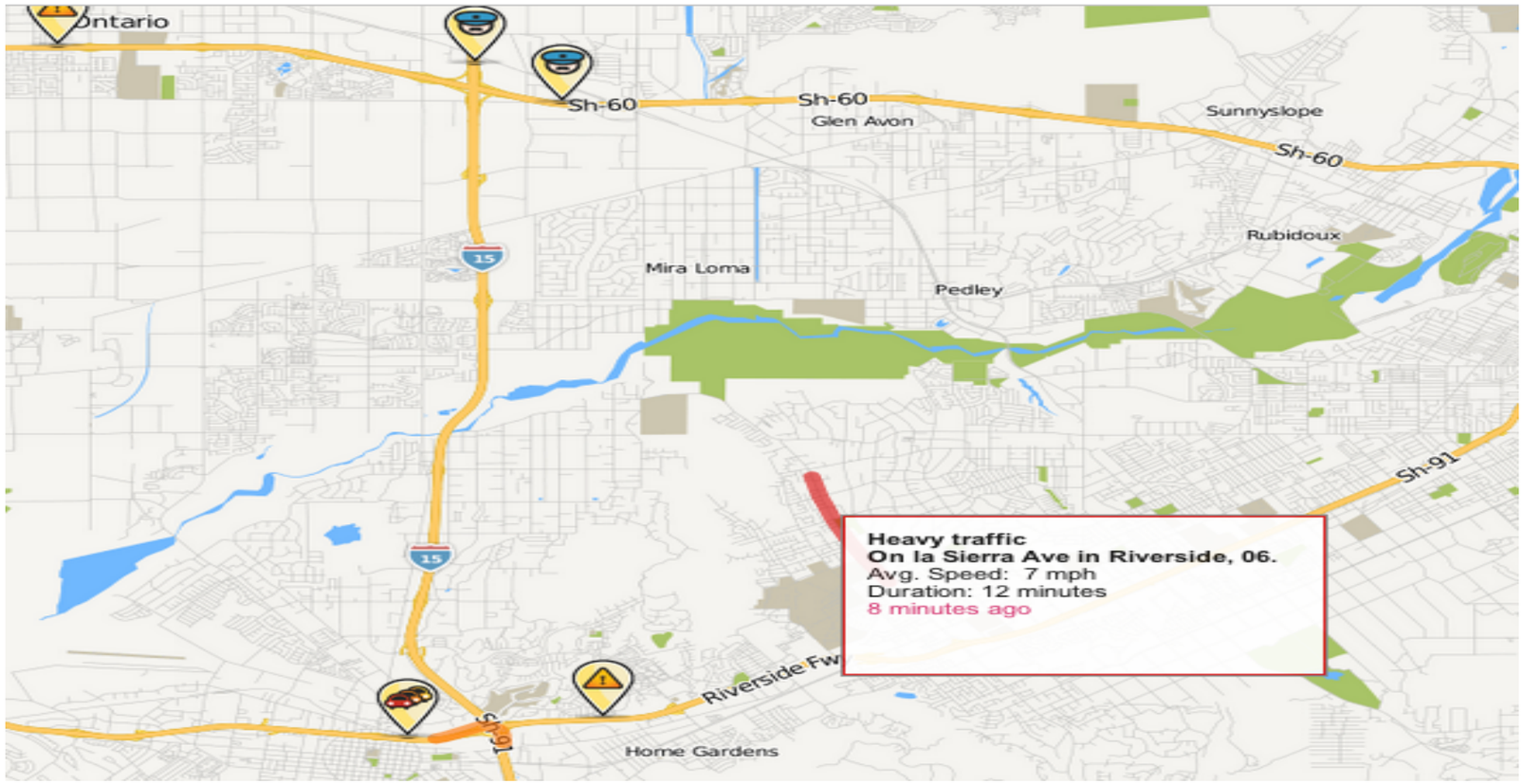
道路监控




vTrack server
GPS positions/
AP observations
Users driving
h smartphones
Real time
routing updates

道路监控


waze
OUTSMARTING TRAFFIC, TOGETHER
CYTAMO...
17:07
49%
Resinelli
Asso
Viganò
Calco
A51
Muggiò
A51 - Tangenziale Est
A1 - Racc


Ontario
Sh-60
Glen Avon
Sunnyslope
Rubidoux
Mira Loma
Pedley
Sh-91
Riverside Fw
Home Gardens
Heavy traffic
On la Sierra Ave in Riverside, 06.
Avg. Speed: 7 mph
Duration: 12 minutes
8 minutes ago

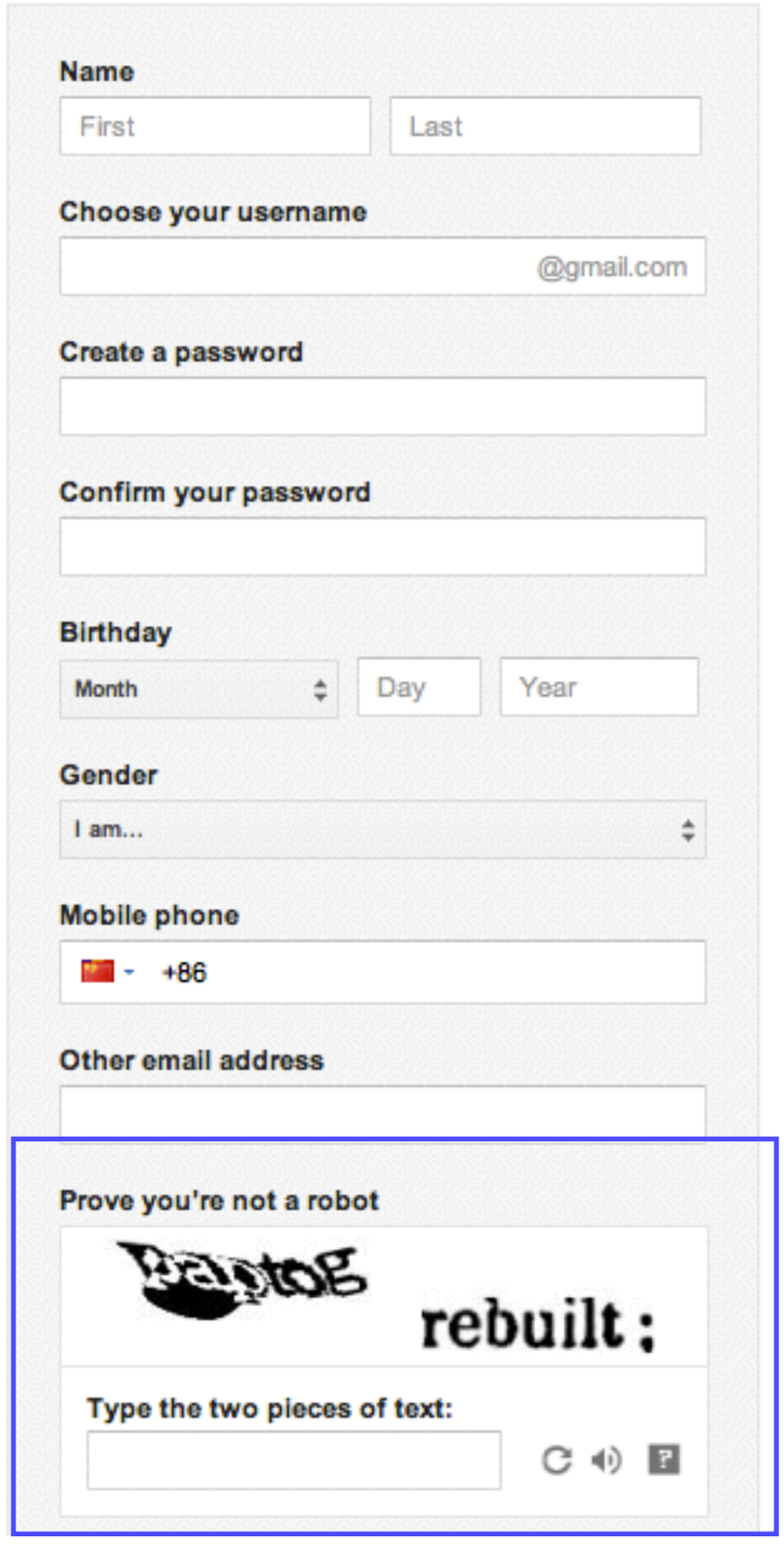
CAPTCHA

CAPTCHA定义

- CAPTCHA
  - Completely
  - Automated
  - Public
  - Turning test to tell
  - Computers and
  - Humans
  - Apart

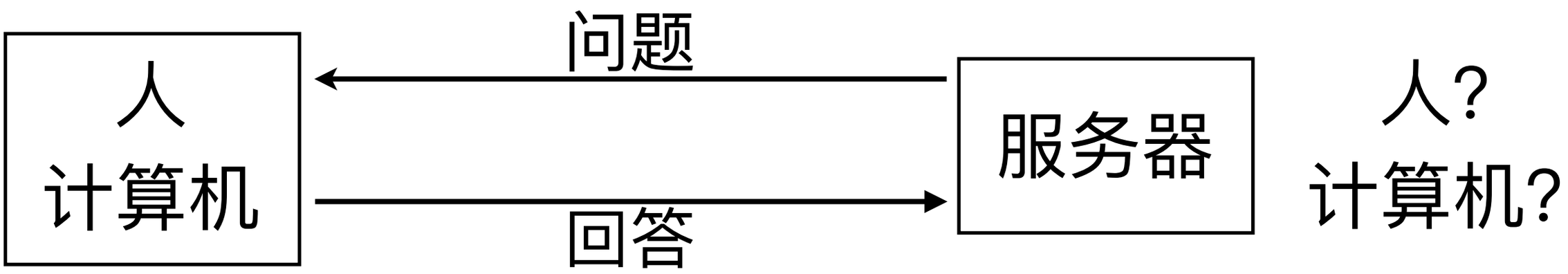
http://www.captcha.net/

http://en.wikipedia.org/wiki/CAPTCHA



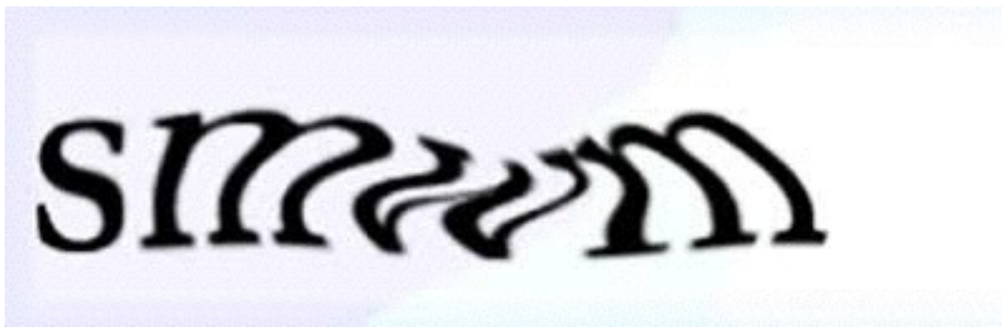
CAPTCHA定义

- A program that generates and grades tests that are human solvable but beyond the capabilities of current computer programs
- 全自动区分计算机和人类的图灵测试
- 验证码，一种区分用户是计算机和人的公共全自动程序



CAPTCHA历史

- 1996, Moni Naor
- 1997年, AltaVista
- Yahoo邮件是第一个大规模采用CAPTCHA的网站
- 2000年，von Ahn和Blum设计出第一个抵御OCR的CAPTCHA
- 2007年，reCAPTCHA
- 2007年，NoCAPTCHA
- … …



HIPs:
Human Interactive Proof

OCR:
Opitcal Character Recognition

CAPTCHA应用

- 判断人还是机器

✴保护网站用户注册，提供免费服务给用户而不是bots

✴保护搜索引擎，防止评论spam

✴确保在线调查更可信，防止虚假投票

✴防止spam制作者很容易的获得email地址

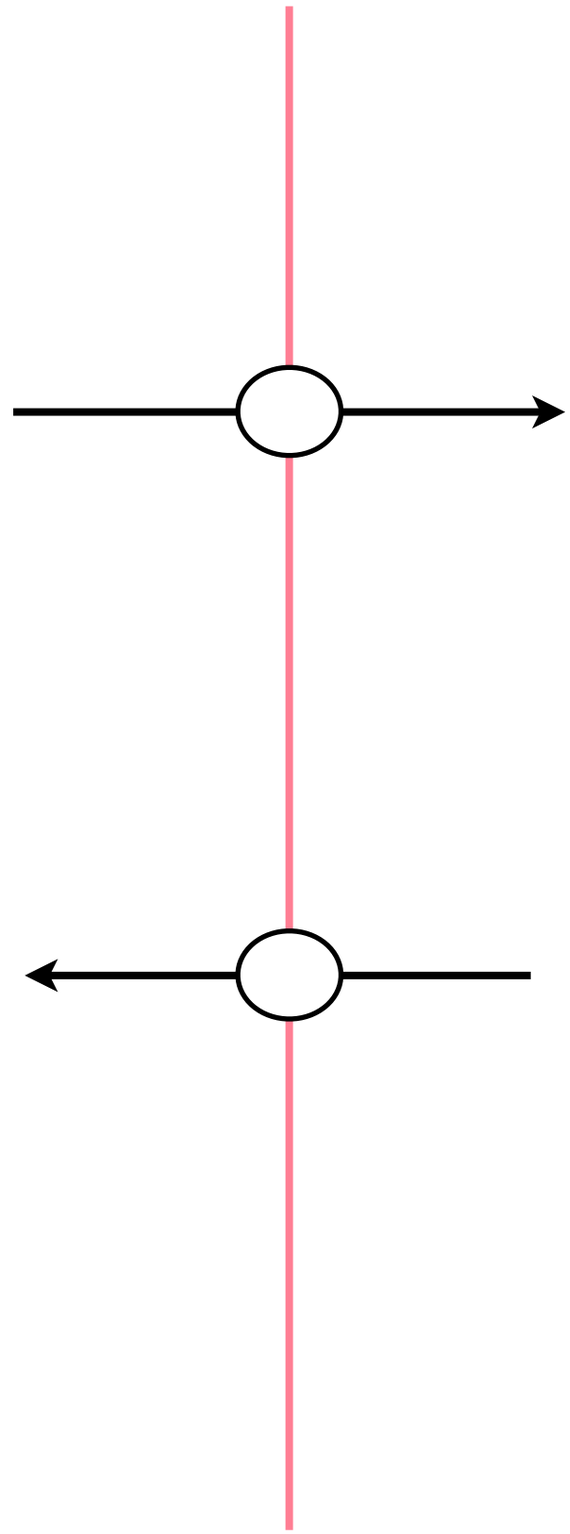
✴防止针对口令系统的字典攻击

✴抵御bots

CAPTCHA vs AI

- 设计CAPTCHA
  - Email
  - BBS
  - Blog
  - SNS
  - Security
  - ... ...

AI难题



- 攻击CAPTCHA
  - 垃圾信息生产者
  - 僵尸网络掌握者
  - 打码等其余攻击者
  - 自动化测试人员
  - ... ...

机器学习

文本 vs 图片

- 请进入如下图片的文字

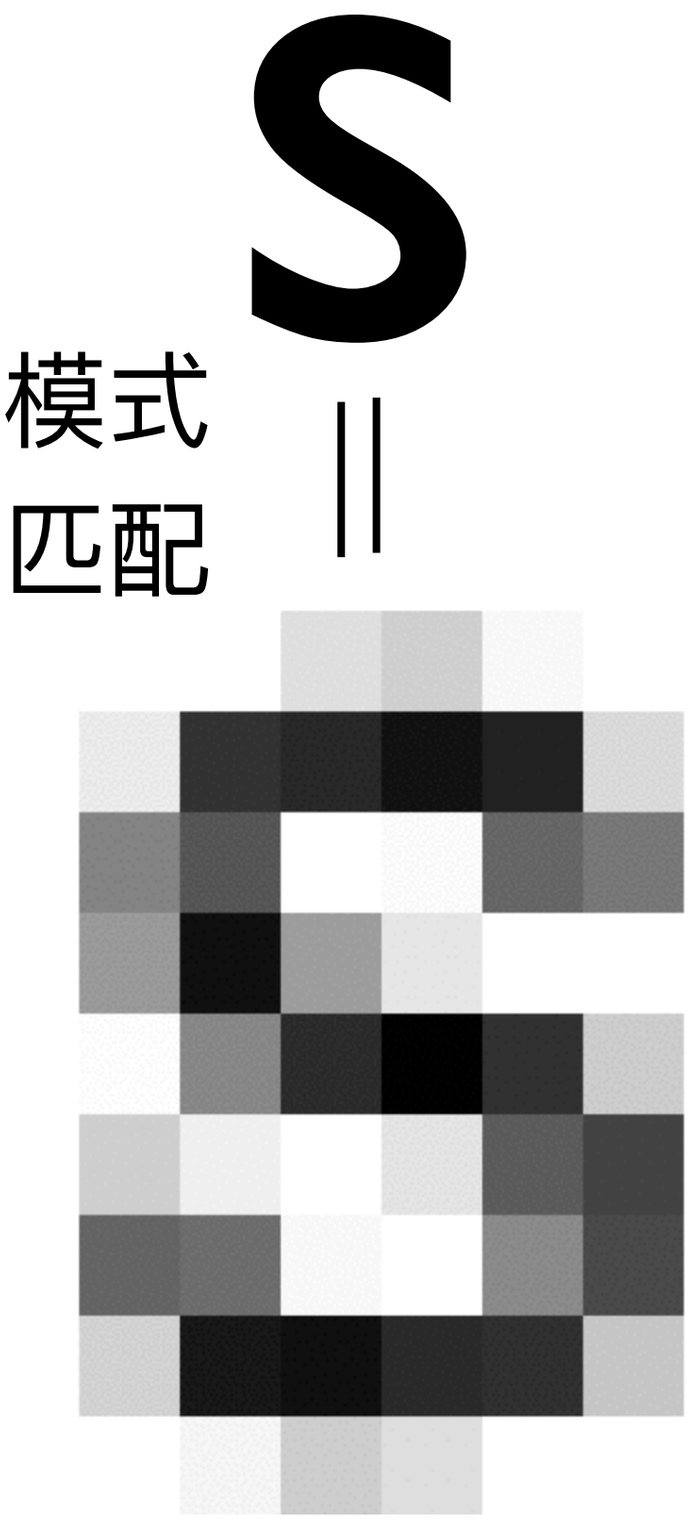
sunhp@ss.pku.edu.cn

- 文字如下：

sunhp@ss.pku.edu.cn

- 像素点
- 图片：二维的像素点
- 计算机不能识别图片中包括什么文本
- 计算机能识别图片中哪些像素相同不相同

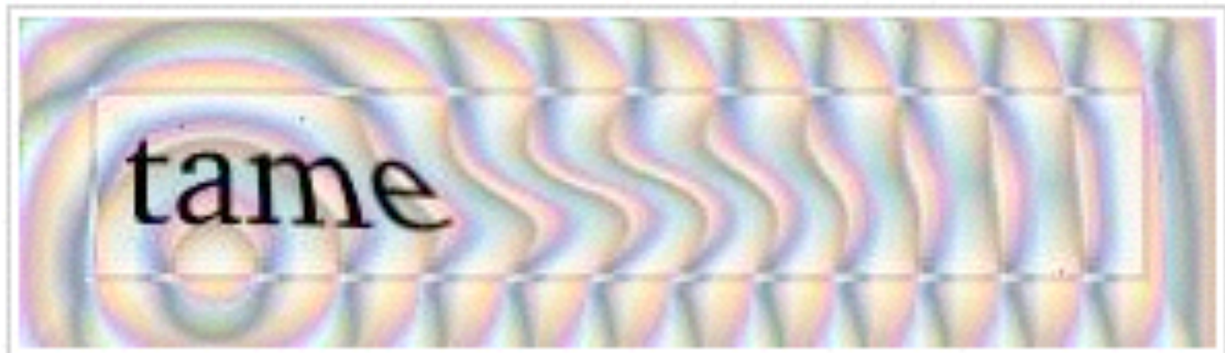
模式匹配



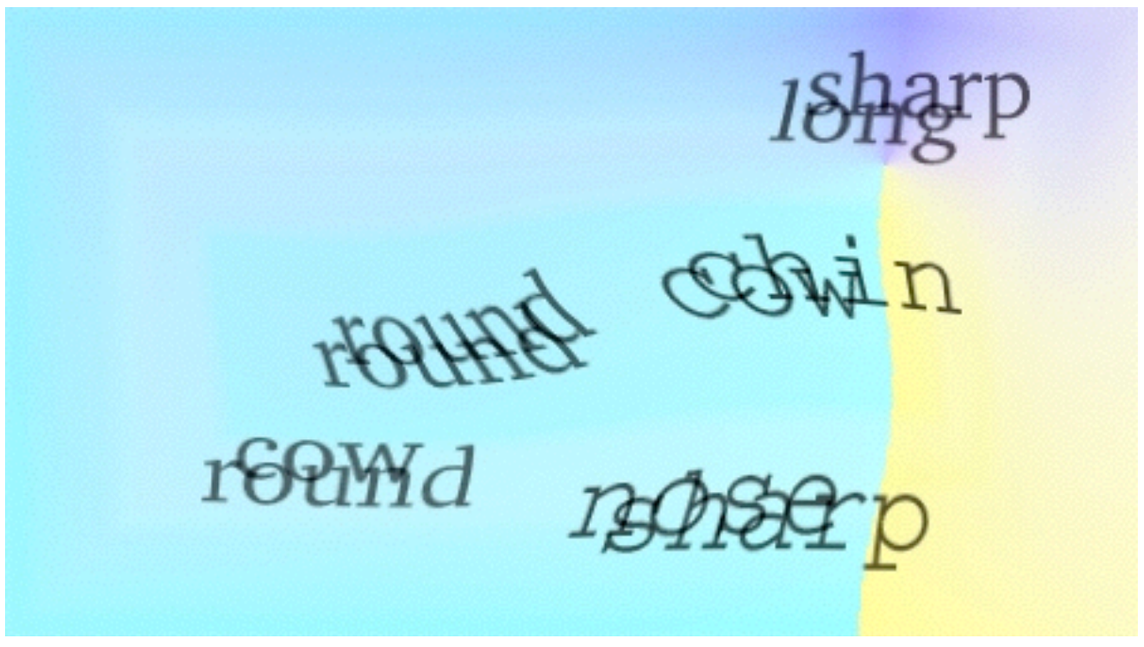
GIMPY / EZ-GIMPY

EZ-GIMPY

Enter the word as it is shown in the box below.



GIMPY



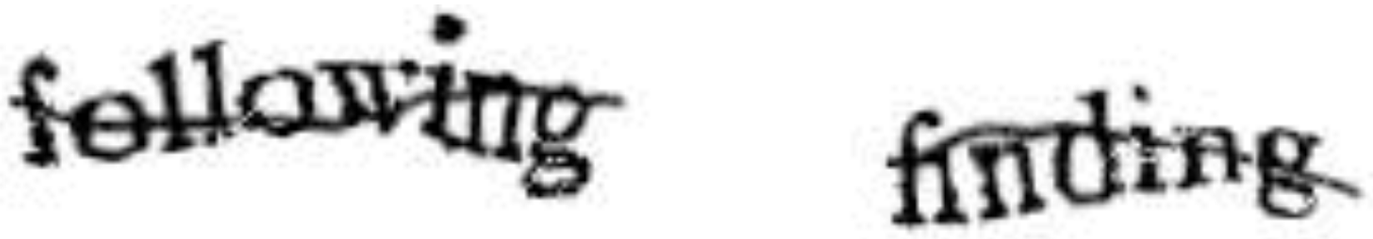
http://www.cs.sfu.ca/~mori/research/gimpy/

文本CAPTCHA原理

Yahoo's EZ-Gimpy

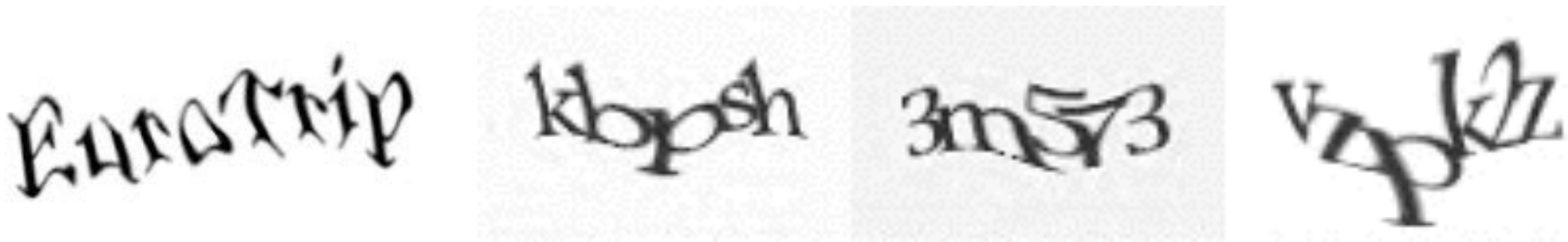
假设：人能读扭曲的文本，但计算机不能

一般情况下要求用户输入扭曲图像中的文字



增加一条曲线使得图像分割更困难

将符号彼此拥挤在一起，但是人也比较难于识别

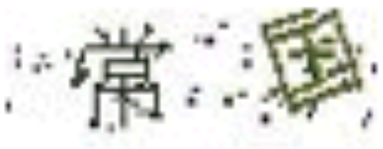


常见文本CAPTCHA

谷歌的验证码

QQ的验证码

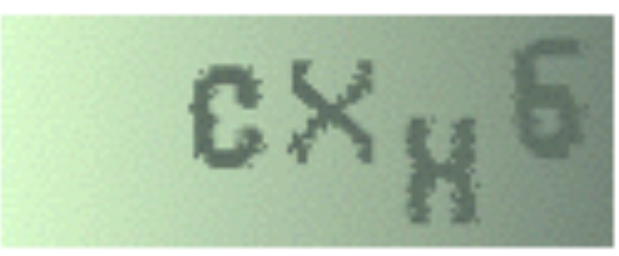
新浪的验证码



网易的验证码

人人网的验证码

天涯的验证码



Discuz的验证码

北大软微论坛

Human Computation

人为攻击CAPTCHA

任务名称	每码价格	更新周期	结算状态	赚友推荐	操作
银河答题（超简单）	28	1天	05月07日金币已发	165	继续任务
阳光打码（日赚30）	15	1天	05月07日金币已发	1435	立即参与
打码兔（日赚50元）	17	1天	05月07日金币已发	2625	继续任务
知码打码（强力推荐）活动	24	1天	05月07日金币已发	3311	立即参与
点图打码（夜班力荐）	28	1天	05月07日金币已发	26	立即参与
图像打码（积分结算）	15	1天	05月07日金币已发	371	立即参与
极速打码（工号加前缀）	16	1天	05月07日金币已发	476	立即参与
超速打码（奖励超高）活动	16	1天	05月07日金币已发	1036	立即参与
发财打码（积分结算）	14	1天	05月07日金币已发	237	立即参与
大众打码（积分结算）	10	1天	05月07日金币已发	399	立即参与

云打码 Yundama 神马都是浮云

DEATH BY CAPTCHA

FASTEST DISCOUNT CAPTCHA SOLVERS

DECAPTCHER.COM

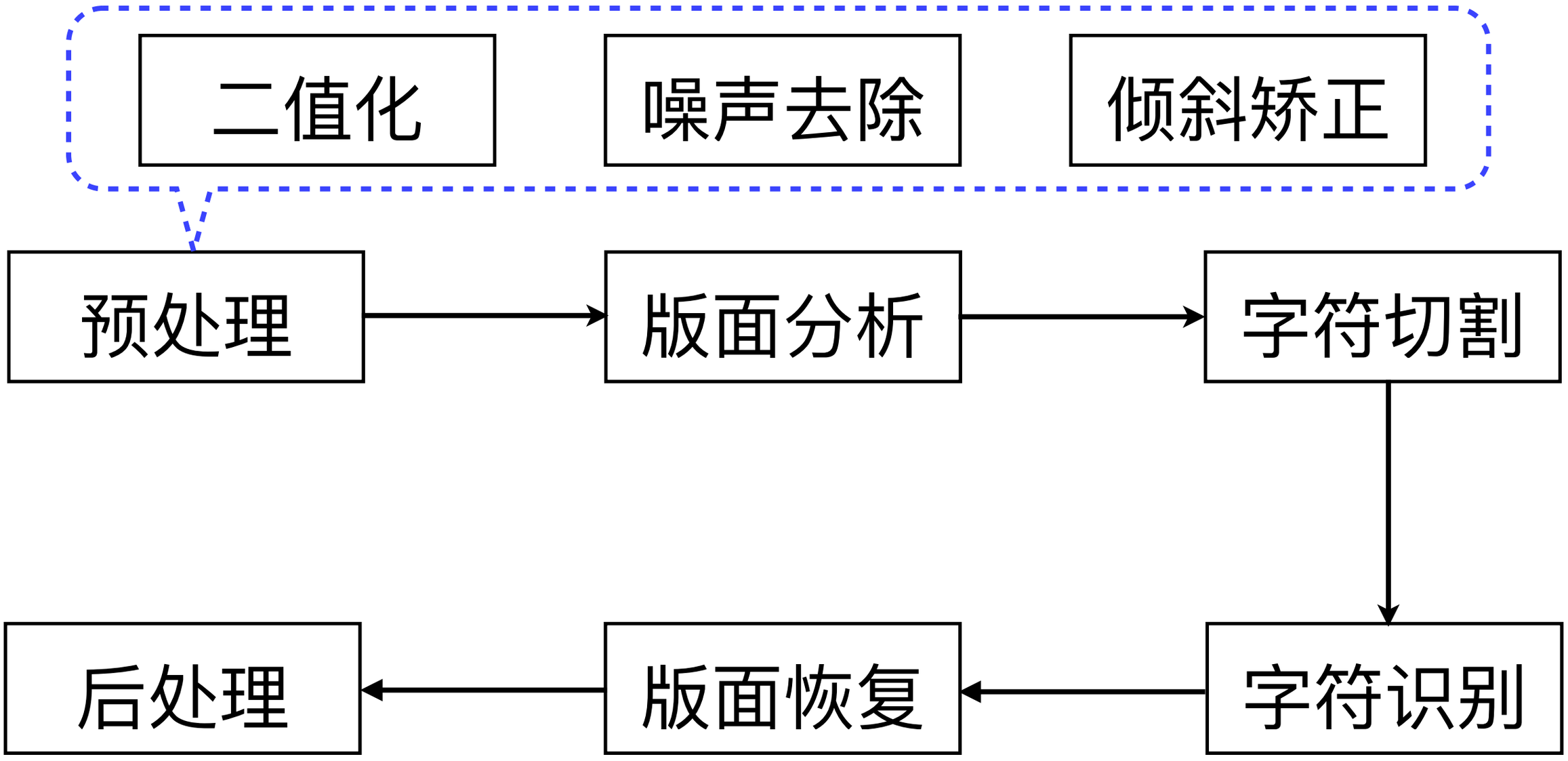
DeCaptcher

Bypass CAPTCHA

5k CAPTCHAs	Only US$6.95
10k CAPTCHAs	Only US$13.90
25k CAPTCHAs	Only US$34.75
50k CAPTCHAs	Only US$69.50
100k CAPTCHAs	Only US$139.00

OCR

- OCR：Optical Character Recognition，光学字符识别



http://en.wikipedia.org/wiki/Optical_character_recognition

常用技术

a．百度使用的验证码

b．天涯论坛使用的验证码

a．开心网使用的验证码

b．谷歌使用的验证码

粘连 扭曲

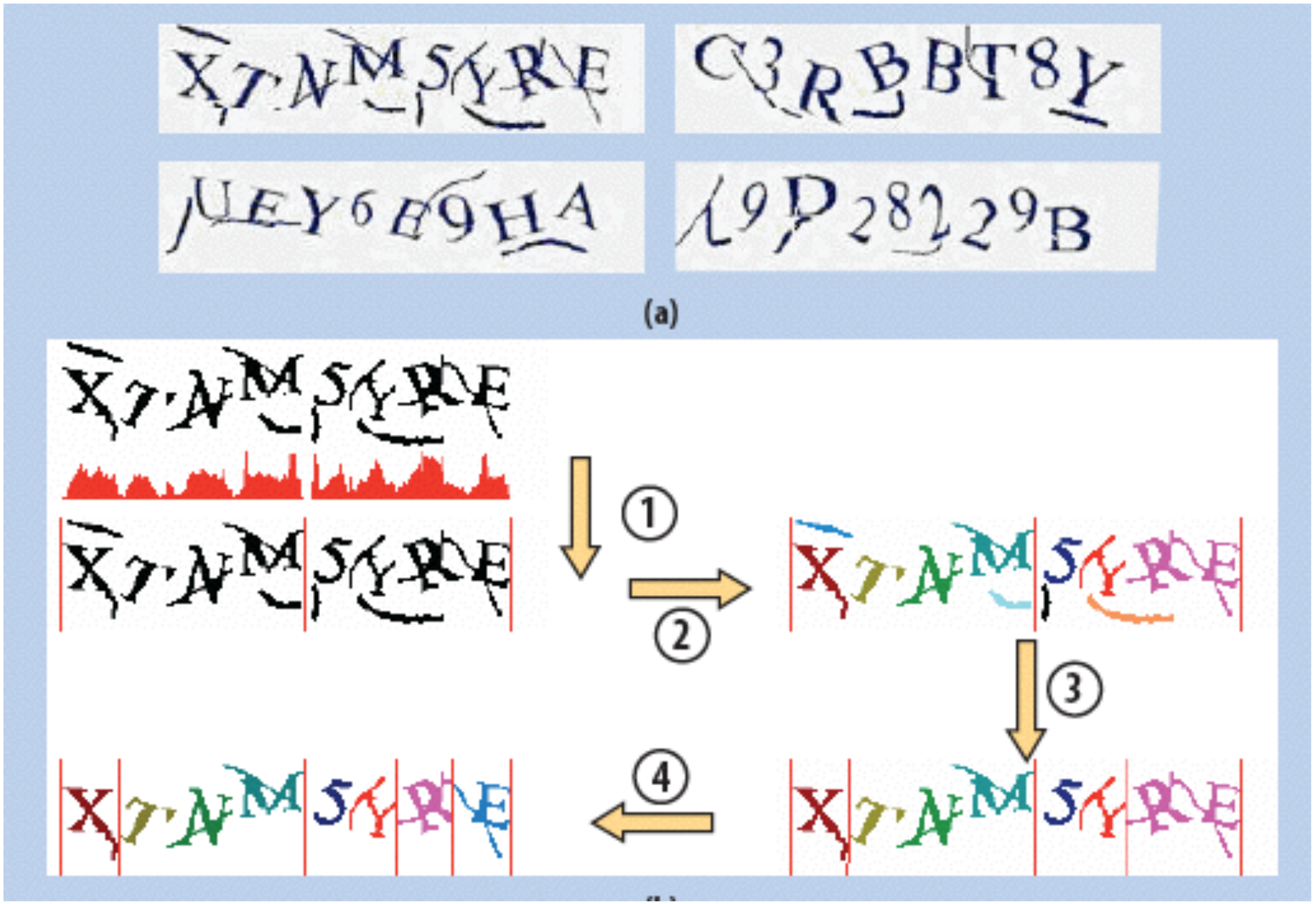
干扰线 其余

a．QQ使用的验证码

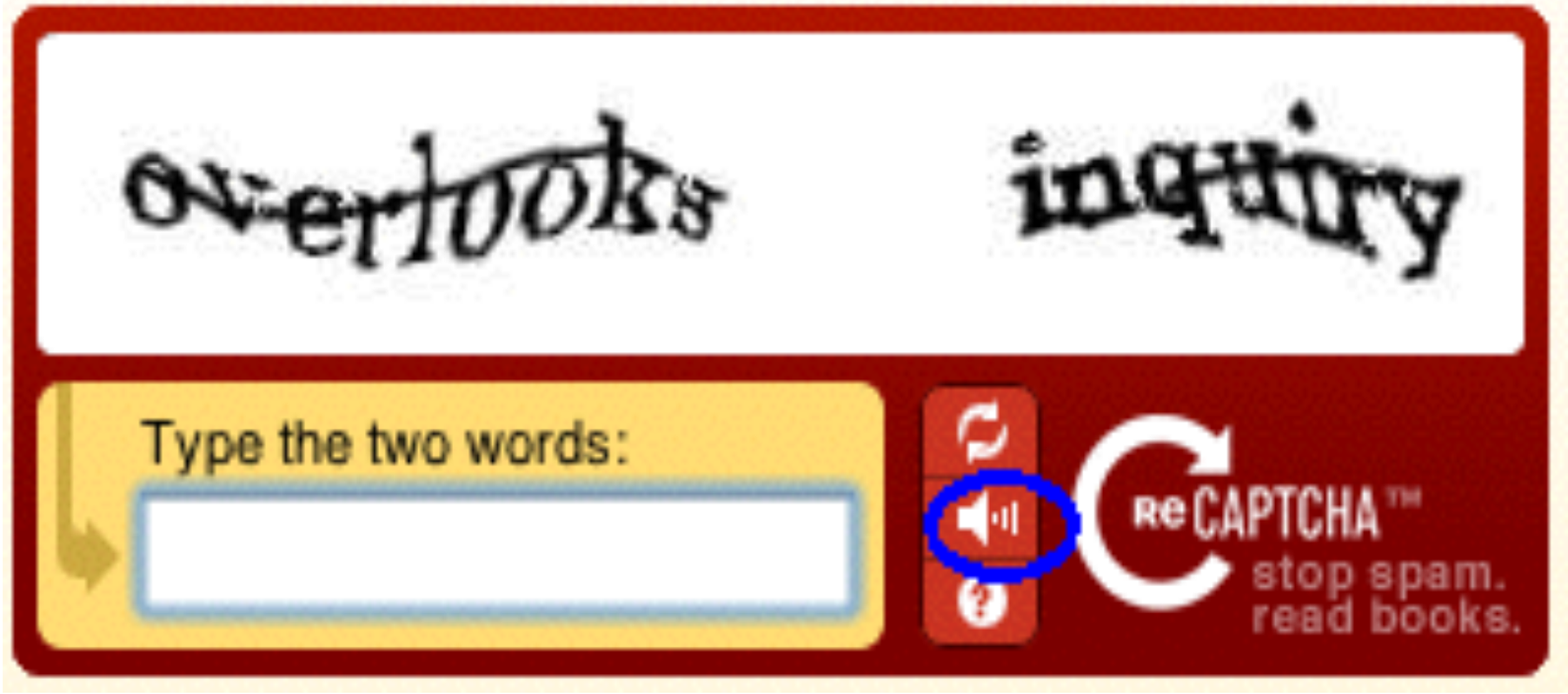
b．新浪网使用的验证码

- 不定字符
- 不定字体和大小
- 背景干扰

攻击例子



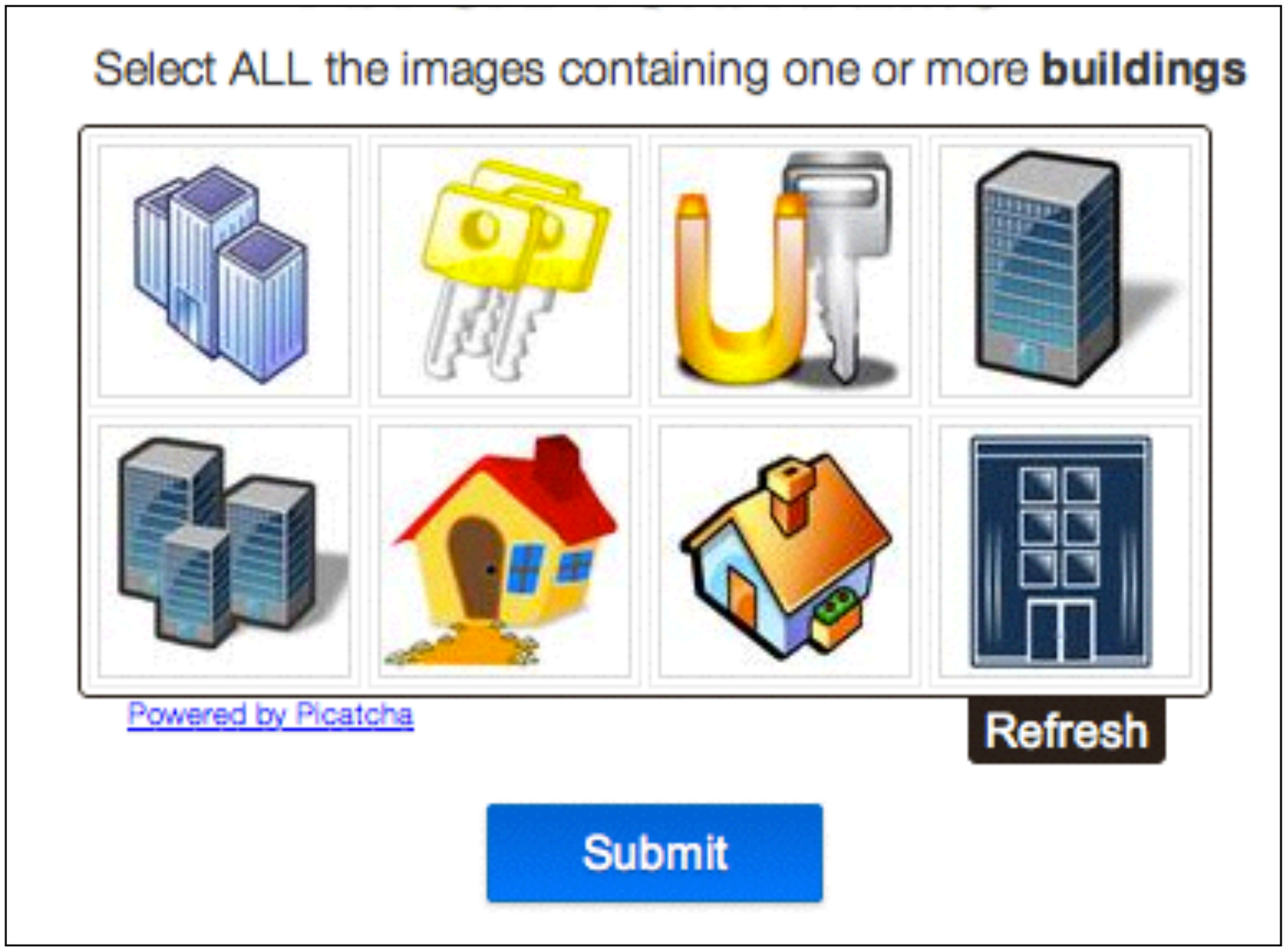
声音CAPTCHA


overlooks
inquiry
Type the two words:
reCAPTCHA™
stop spam.
read books.

分类CAPTCHA

Please click on all the images that show cats:
adopt me
adopt me
adopt me
adopt me
adopt me
adopt me
adopt me
adopt me
adopt me
adopt me
adopt me
adopt me
Score Test
You're a human!

分类CAPTCHA


Select ALL the images containing one or more **buildings**
Powered by Picatcha
Refresh
Submit

Human Computation

逻辑CAPTCHA

拼图CAPTCHA

基于拼图

NuCAPCTHA

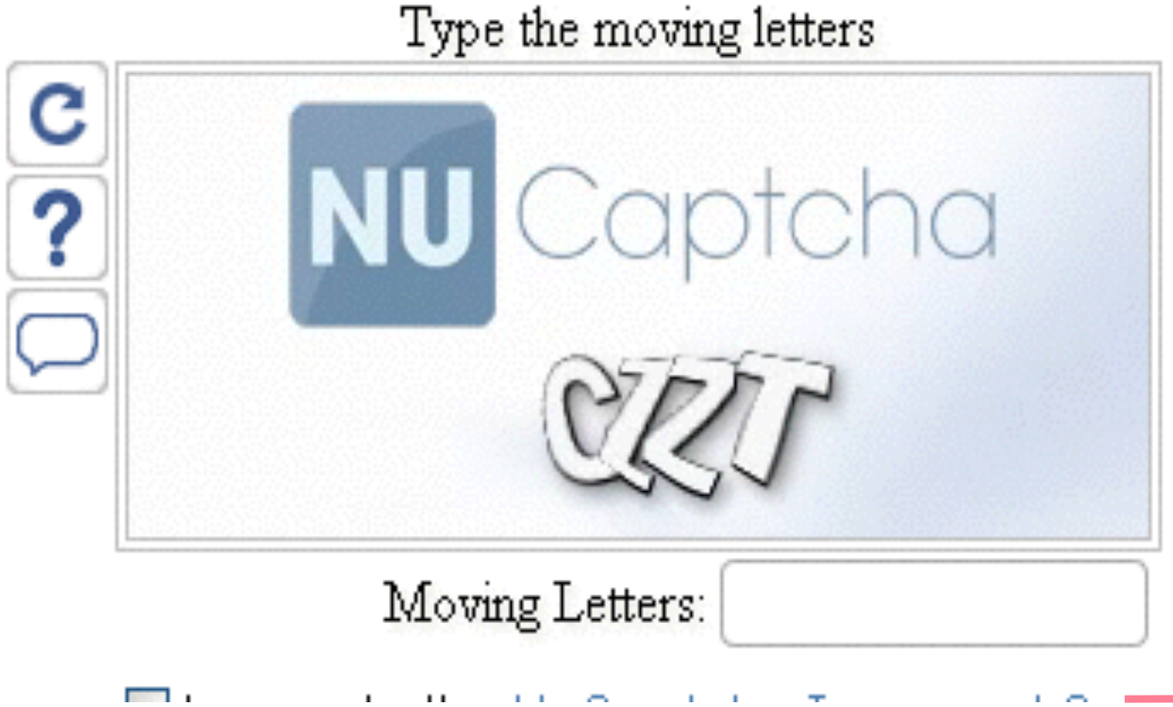
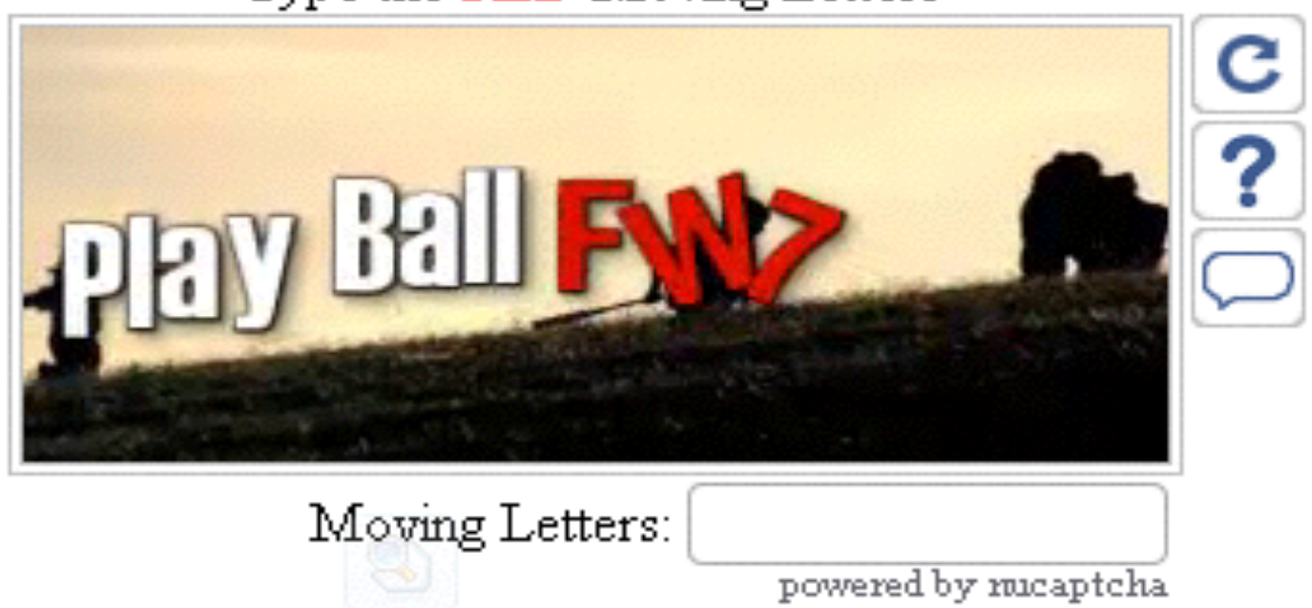


Figure 1: Sample screenshot non animated security settings for NuCaptcha

用户行为分析系统



Jump to this Ad / Change the image / Switch to a text type

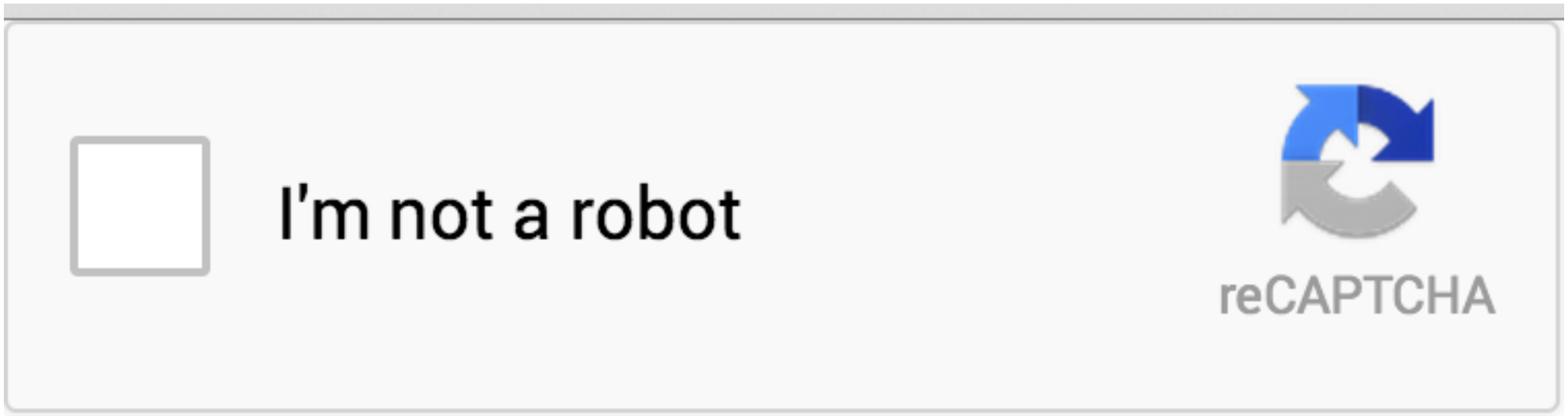
Confirm

Send

Mineye

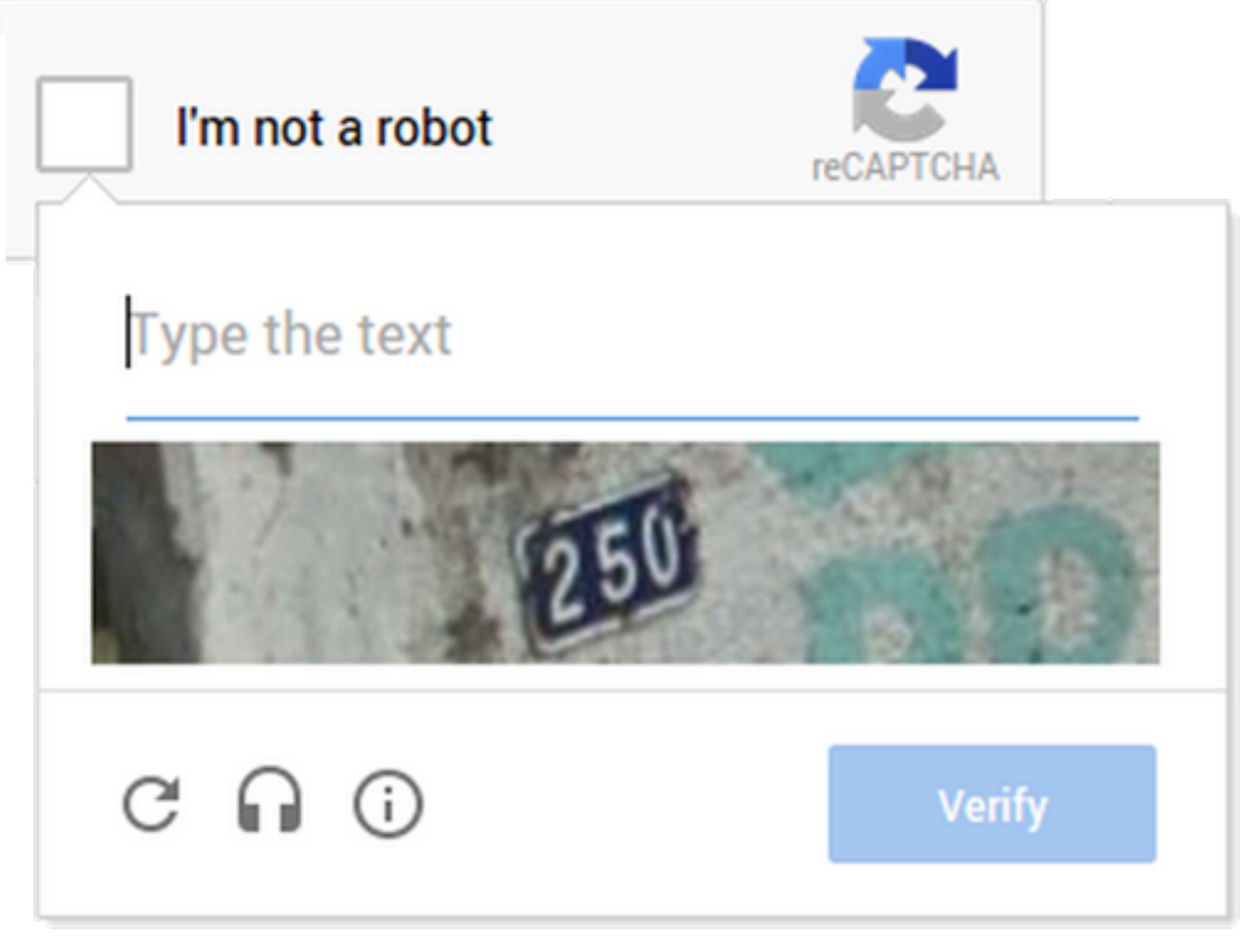
Submit

NoCAPTCHA


I'm not a robot
reCAPTCHA

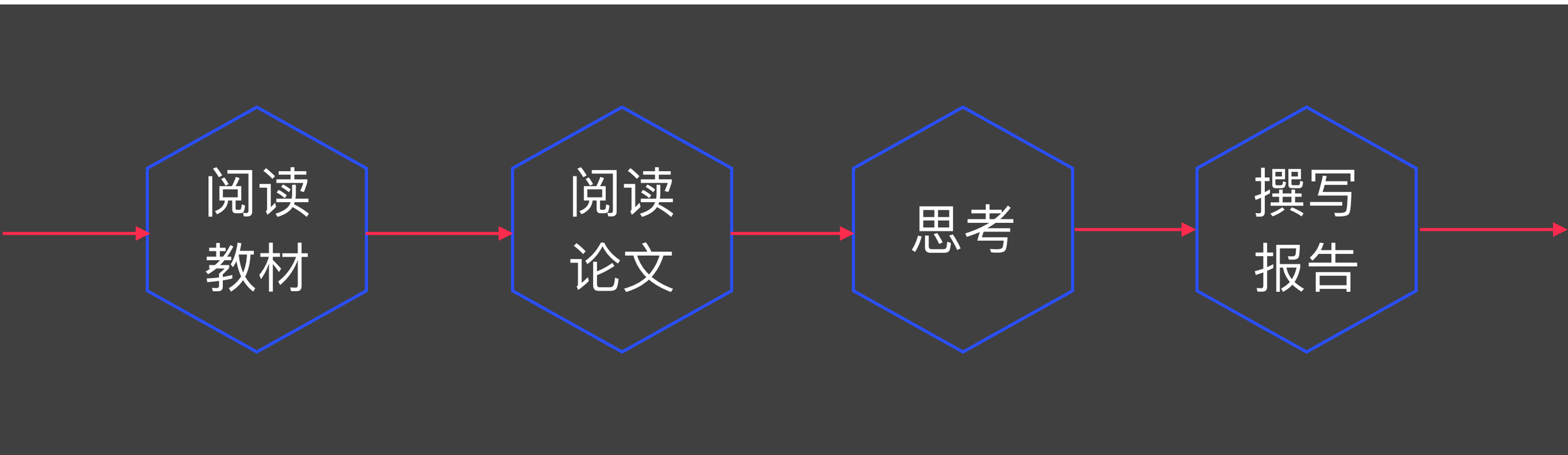
I'm not a robot
reCAPTCHA
Select all images below that match this one:
Verify

I'm not a robot
reCAPTCHA
Select all images below that match this one:
Verify


I'm not a robot
reCAPTCHA
Type the text
250
Verify

提问时间！

课后作业



Homework

课后作业

要求阅读上周分配的论文，写阅读报告

1、文章概述

2、主要收获

3、存在疑问

4、所思所感

5、一篇论文

周五晚上12点前提交

谢谢！

Huiping Sun

sunhp@ss.pku.edu.cn

https://huipingsun.github.io