

# Crowdsourcing



# Overview

- Crowdsourcing = Crowd + Outsourcing
  - Forced cooperation
  - Volunteer cooperation
  - Incented cooperation
  - Paid microtasks

# reCAPTCHA



- CAPTCHA

- Completely Automated Public Turing test to tell Computers and Humans Apart
- Challenge response test to prevent bots

- reCAPTCHA

- Originally CMU project, acquired by Google
- Helps to digitize books, newspapers, old time radio
- 200 million CAPTCHAs / day



Submit

The words above come from scanned books.  
By typing them, you help to digitize old texts.

# CAPTCHA

- Accessibility
  - Audio version
- Defeating CAPTCHA
  - Use computer vision/machine learning
  - Replay to humans to solve
    - Sweat shop somewhere, \$4/day
  - Replay on high volume site



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





**WIKIPEDIA**  
The Free Encyclopedia

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia

- Interaction
  - Help
  - About Wikipedia
  - Community portal
  - Recent changes
  - Contact Wikipedia

- Toolbox
- Print/export

- Languages
  - Acèh
  - Afrikaans

Log in / create account

Article [Talk](#)

Read [View source](#) [View history](#)

Search

# Wikipedia

From Wikipedia, the free encyclopedia



*For Wikipedia's non-encyclopedic visitor introduction, see [Wikipedia:About](#).*

*For Wikipedia's formal organizational structure, see [Wikipedia:Formal organization](#).*

**Wikipedia** (/wɪkɪˈpiːdiə/ or /wɪkiˈpiːdiə *WIK-i-PEE-dee-ə*) is a free, collaborative, multilingual Internet encyclopedia supported by the non-profit Wikimedia Foundation. Its 20 million articles (over 3.8 million in English alone) have been written collaboratively by volunteers around the world. Almost all of its articles can be edited by anyone with access to the site,<sup>[3]</sup> and it has about 100,000 regularly active contributors.<sup>[4]</sup> As of January 2012, there are editions of Wikipedia in 283 languages. It has become the largest and most popular general reference work on the Internet,<sup>[5][6][7][8]</sup> ranking sixth globally among all websites on Alexa and having an estimated 365 million readers worldwide.<sup>[5][9]</sup> It is estimated that Wikipedia receives 2.7 billion monthly pageviews from the United States alone.<sup>[10]</sup>

Wikipedia was launched in January 2001 by Jimmy Wales and Larry Sanger.<sup>[11]</sup> Sanger coined the name *Wikipedia*,<sup>[12]</sup> which is a portmanteau of *wiki* (a technology for creating websites collaboratively, from the Hawaiian word *wiki*, meaning "quick")<sup>[13]</sup> and *encyclopedia*.

Wikipedia's departure from the expert-driven style of encyclopedia building and the presence of a large body of unacademic content has received ample attention in

## Wikipedia



## WIKIPEDIA

The logo of Wikipedia, a globe featuring glyphs from many different writing systems

**Screenshot** [\[show\]](#)

**URL** [wikipedia.org](http://wikipedia.org)

**Slogan** The Free Encyclopedia

**Commercial?** No [Improve this page](#)



- Online astronomy project
  - Volunteers classify galaxies, citizen science
- 2007 version
  - Sloan Digital Sky Survey, 1M image galaxies
  - Classify: elliptical or spiral, clockwise/anti-clockwise
    - 24 hours after launch, 70,000 classification / hour
    - 50M classifications / year from 150K people
    - Multiple volunteers as good as professional astronomers





- Galaxy Zoo 2

- 250K brightest galaxies from Galaxy Zoo
- More detailed classification, shape/intensity, oddities
- 60M classifications

- Galaxy Zoo Hubble

- Images from NASA Hubble telescope
- Many more questions
  - Is it smooth with no sign of a disk?
  - How rounded is it?
  - Could this be a disk viewed edge-on?



# Crowdsourced games

- ESP game

- Pairs of players try and guess same word for image
- "Labeling Images with a Computer Game", Luis von Ahn and Laura Dabbish, CHI 2004



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"

**Figure 1. Partners agreeing on an image. Neither of them can see the other's guesses.**

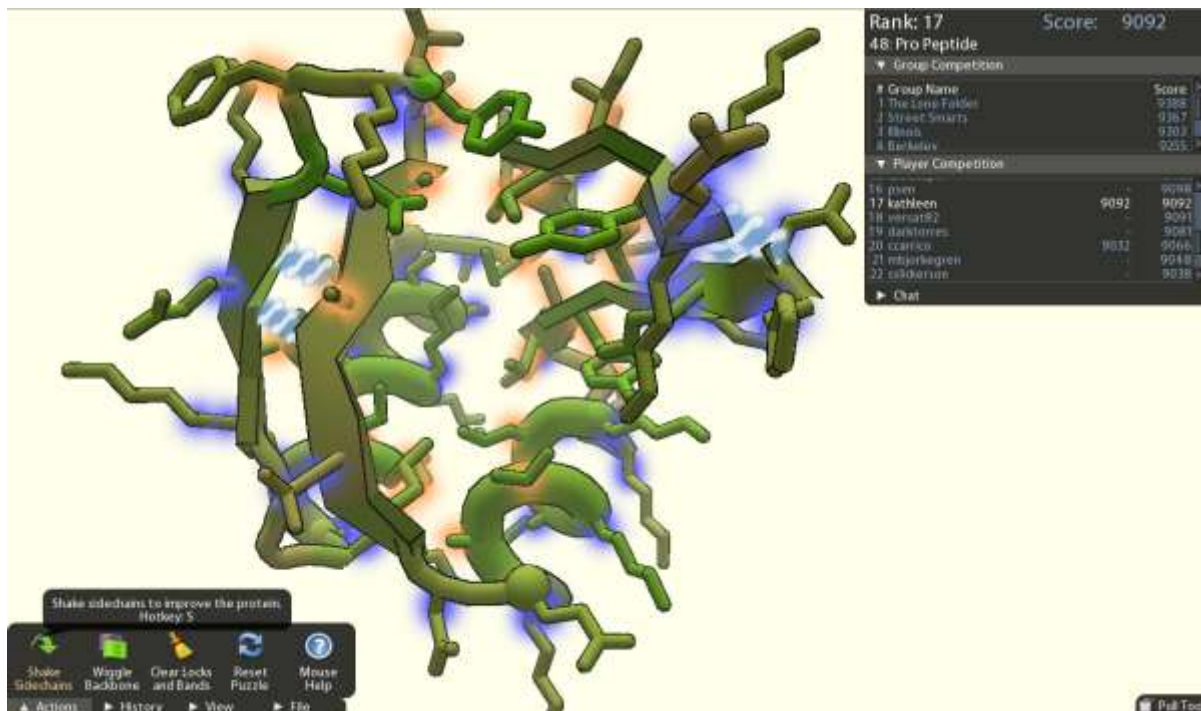


**Figure 2. The ESP Game. Players try to "agree" on as many images as they can in 2.5 minutes. The thermometer at the bottom measures how many images partners have agreed on.**



# Crowdsourced games

- Foldit
- EteRNA
  - Games related to folding of RNA molecules



# A real-world task

- 2009 DARPA network challenge
  - Defense Advanced Research Projects Agency
  - 40<sup>th</sup> anniversary of the Internet
  - \$40,000 first team to locate 10 moored, 8 foot, red, weather balloons
  - 10 previously undisclosed locations
  - 4,000 teams competed





# MIT team's strategy

- Multi-level marketing

- \$2000 to whoever sends correct coordinates
- \$1000 to whoever invited them
- \$500 whoever invited the person who invited them
- \$250 ...

- Mobilizing large groups requires the right incentive

- Georgia Tech

- Promised to donate proceeds to charity
- 2<sup>nd</sup> place

[http://www.youtube.com/watch?v=6Ga\\_EJWLzHA](http://www.youtube.com/watch?v=6Ga_EJWLzHA)

# Netflix prize

- \$1M prize
- Improve algorithm, predict ratings given past ratings
  - 10% improvement over Netflix's algorithm
  - Oct 2006, won in Sept 2009
- Data:
  - Training: 100M ratings, 480K users, 18K movies
  - <user, movie, data of grade, grade>
  - Quiz set (1.4M), Test set (1.4M)



# Paid crowdsourcing

amazonmechanical turk  
Artificial Artificial Intelligence

text

Your Account

HITs

Qualifications

Introduction | Dashboard | Status | Account Settings

## Mechanical Turk is a marketplace for work.

We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient.

**417,411 HITs** available. [View them now.](#)

**Make Money**  
by working on HITs

**Get Results**  
from Mechanical Turk Workers

CrowdFlower

Why CrowdFlower?

Products

Solutions

Customers

About

Blog

The world's largest enterprise crowdsourcing platform.

With access to more than 1.5 Million Contributors worldwide, our technology platform offers quality-assured crowdsourcing at unprecedented scale.

PICK A SOLUTION ▼



# Amazon Mechanical Turk

- Human Intelligence Task (HIT)
  - Workers and requestors
  - Web-based
    - HTML + JavaScript if hosted entirely on MTurk
  - Price per HIT
  - Number of workers per HIT
  - Qualifications for workers
  - Accept/reject work

<http://www.behind-the-enemy-lines.com/2010/03/new-demographics-of-mechanical-turk.html>

[http://waxy.org/2008/11/the\\_faces\\_of\\_mechanical\\_turk/](http://waxy.org/2008/11/the_faces_of_mechanical_turk/)

