



NANYANG
TECHNOLOGICAL
UNIVERSITY
SINGAPORE

CE/CZ4052 Cloud Computing

Crowdsourcing in Cloud

Dr. Tan, Chee Wei

Email: cheewei.tan@ntu.edu.sg

Office: N4-02c-104



A number of slides due to Prof. Luis von Ahn at Carnegie Mellon University (www.cs.cmu.edu/~biglou/TechEd.ppt)

Human-assisted Computation: Crowdsourcing

One of the first SaaS on Amazon Cloud

For the first time in **human** history,
hundreds of millions of **people** can, via
the **Internet**, easily collaborate on the
same problem.

Human-In-Loop: Mechanical Turk

- The Turk, also known as the [Mechanical Turk](#) or Automaton Chess Player, was a fake chess-playing machine constructed in the late 18th century. From 1770 until its destruction by fire in 1854 it was exhibited by various owners as an automaton, though it was eventually revealed to be an elaborate hoax.
- Constructed and unveiled in 1770 by Wolfgang von Kempelen to impress the Empress Maria Theresa of Austria, the mechanism appeared to be able to play a strong game of chess against a human opponent, as well as perform the knight's tour, a puzzle that requires the player to move a knight to occupy every square of a chessboard exactly once. Even Napoleon (Emperor of France) played against it in 1809.



Human-In-Loop: Mechanical Turk



The interior of the machine was very complicated and designed to mislead anyone who observed it. When opened on the left, the front doors of the cabinet exposed a number of gears and cogs, similar to clockwork. The section was designed so that if the back doors of the cabinet were open at the same time one could see through the machine.

The other side of the cabinet did not house machinery; instead it contained a red cushion and some removable parts, as well as brass structures. This area was designed to provide a clear line of vision through the machine. Underneath the robes of the Ottoman model, two other doors were hidden.

These also exposed clockwork machinery and provided a similarly unobstructed view through the machine. The design allowed the presenter of the machine to open every available door to the public. In this way, he could maintain the illusion.

Human-In-Loop: Amazon Mechanical Turk

- In 2005, Amazon.com launched the Amazon Mechanical Turk (MTurk). The service was inspired, in part, by the way Kempelen's Mechanical Turk operated.
- MTurk provides companies with access to a marketplace of on-demand workers whenever a business needs them connecting to a variety of skill sets and capabilities via a global workforce.
- Organizations can harness the power of crowdsourcing via MTurk for a range of use cases, such as microwork, human insights and machine learning development.

amazon mechanicalturk™
Artificial Artificial Intelligence



[All HITs](#) | [HITs Available To You](#) | [HITs Assigned To You](#)

Search for containing that pay at least \$ for which you

All HITs

1-10 of 913 Results

Sort by:

GO!

[Show all details](#) | [Hide all details](#)

1 [2](#) [3](#) [4](#) [5](#) > [Next](#) >> [Last](#)

[Share your FAVORITES!](#)

[View a HIT in this group](#)

Requester: [mastercard](#)

HIT Expiration Date: Jul 4, 2009 (5 days 9 hours) **Reward:** \$0.10

Time Allotted: 30 minutes

HITs Available: 16480

[Favorites Second Round Edit](#)

[View a HIT in this group](#)

Requester: [mastercard](#)

HIT Expiration Date: Jun 30, 2009 (1 day 2 hours) **Reward:** \$0.03

Time Allotted: 60 minutes

HITs Available: 12680

[Read a news article and identify the director\(s\) that are leaving or joining the board.](#)

[View a HIT in this group](#)

Requester: [Katie Kong](#)

HIT Expiration Date: Jul 26, 2009 (3 weeks 6 days) **Reward:** \$0.01

Time Allotted: 60 minutes

HITs Available: 4860

[Insurance Co. Contact List](#)

[View a HIT in this group](#)

Requester: [Jeffrey Bloom](#)

HIT Expiration Date: Jul 8, 2009 (1 week 2 days) **Reward:** \$0.10

Time Allotted: 60 minutes

HITs Available: 3647

Human-assisted Computation: Crowdsourcing

The Story of **CAPTCHA** (Completely Automated Public Turing test to tell Computers and Humans Apart) and **reCAPTCHA**





https://edit.yahoo.com/config/eval_register?.intl=us&new=1&.don



Google



Yahoo! Registration



Verify Your Registration

* Enter the code shown: [More info](#)

This helps Yahoo! prevent automated registrations.



pc2d

Terms of Service

Please review the following terms and indicate your agreement below. [Printable Version](#)

1. ACCEPTANCE OF TERMS

Yahoo! Inc. ("Yahoo!") welcomes you. Yahoo! provides its service to you subject to the following Terms of Service ("TOS"), which may be

By clicking "I Agree" you agree and consent to (a) the Yahoo! [Terms of Service](#) and [Privacy Policy](#), and (b) receive required notices from Yahoo! electronically.



I Agree

I Do Not Agree

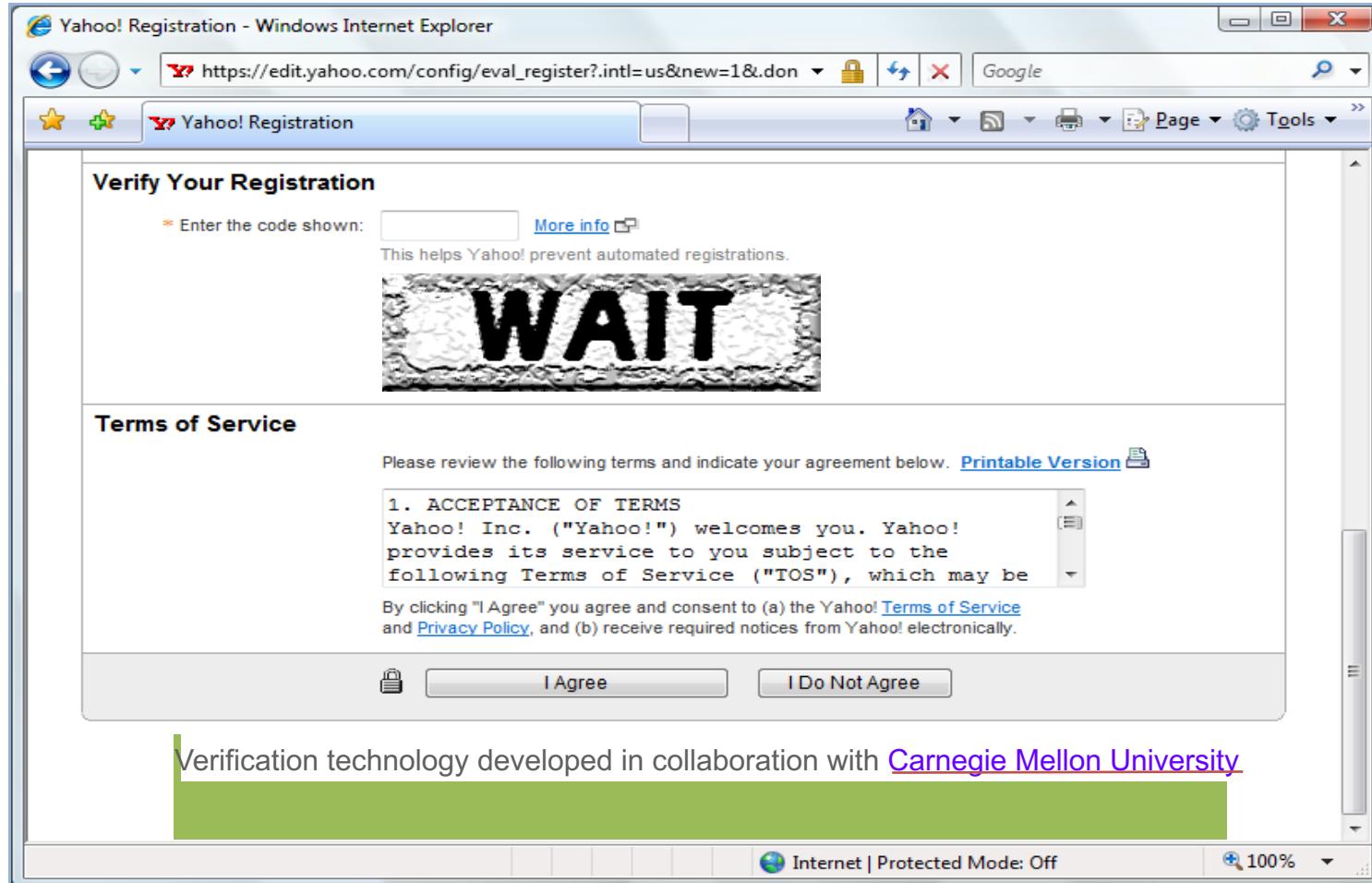
Verification technology developed in collaboration with [Carnegie Mellon University](#)



Internet | Protected Mode: Off

100%

“Help! I’ve been waiting for 20 minutes, and nothing happens”





Verify Your Registration

* Enter the code shown: [More info](#)

This helps Yahoo! prevent automated registrations.

Terms of Service

Please review the following terms and indicate your agreement below. [Printable Version](#)

1. ACCEPTANCE OF TERMS

Yahoo! Inc. ("Yahoo!") welcomes you. Yahoo! provides its service to you subject to the following Terms of Service ("TOS"), which may be

By clicking "I Agree" you agree and consent to (a) the Yahoo! [Terms of Service](#) and [Privacy Policy](#), and (b) receive required notices from Yahoo! electronically.



I Agree

I Do Not Agree

Verification technology developed in collaboration with [Carnegie Mellon University](#)



Internet | Protected Mode: Off

100%



CAPTCHA Sweat Shops

Spam companies hire humans to
type CAPTCHAs all day long

\$2.50 per hour for each human
720 captchas per hour per human
1/3 cent per account

**CAPTCHA (completely automated public Turing
test to tell computers and humans apart)**



Показывать
информацию обо мне

- Всем
- Только зарегистрированным пользователям
- Никому

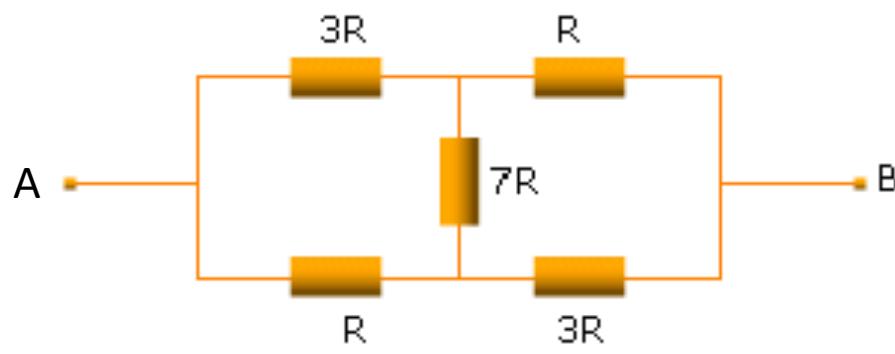
Защита от
автоматической
регистрации

$$\lim_{x \rightarrow 0} \ln \left(2 + \sqrt{\arctg x \cdot \sin \frac{1}{x}} \right)$$

Введите ответ

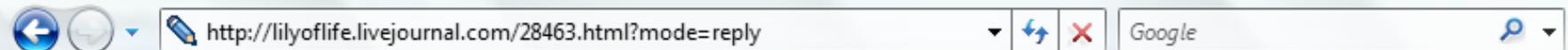
X Очистить

Всё верно



$$R_{AB} = \frac{\boxed{}}{\boxed{}} R$$

जमा करें



hope mort ideas ipoo worry
isaacastimov jamie jane jealousy
jerr jeremiah joe jordan joy
justin katie keely kelly kelsey
kevin KKK knights larissa
laughter laura's life light lIndsey
list lila lizphair lolcat lord of the
rings love marina mario
materialism math matt mckenna
memory mike'd mind mitchell
mom monica mr bermier mr
mantas mr stanton mr vincent mrs
robinson MUSIC mystics mythology
najiba nature neptune ning noble
nolan note oops pain painting
passion people persephone perspective
phenology physics pictures poetry
president bush quest for camelot quiz
rachael rachel radio free roscoe rain
rant raven ravens religion revolution
running sacrifice sadness saltwater
tatty sarcasm school science
scott sebastian self sexuality sick
signs sites silentnight silliness silver
lining sky sloth snow songs sorrow
spirit guides spirituality squirrel
star steph strength summer
survey sweater sweet t.s. elliot
tarot teacher teardrops on my guitar
teija telephone test the door to hell the
last unicorn theo therapy thom thumbs
thunderstorm tiffany tree trees trent

[Add to Memories](#) | [Tell a Friend](#) | [Link](#)

Comment Form

([Read Comments](#))

What is “1 + 1” :

Subject:



Don't auto-format: ?

Message:

[Post Comment](#)

[Preview](#)

Done

Internet | Protected Mode: Off

100%

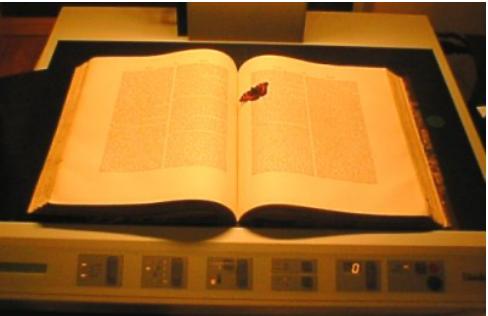


Roughly 200 million CAPTCHAs are typed every day

10s of human time per CAPTCHA

Can we use this human effort for good?

Digitizing Books



Scan

DENMARK.

COPENHAGEN, Dec. 17—Noon.

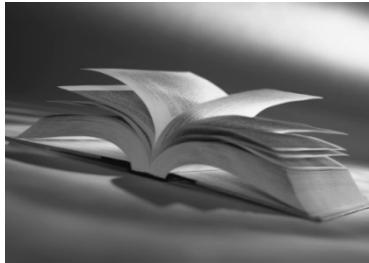
The royal palace at Friedericheberg took fire last night. The King himself directed the operations for extinguishing the fire. The whole structure was, however, destroyed.

OCR

DE xi li ARII..

COPBCtrAGE\, DeC. 11 1 QOn.

The royal palace at Frieclericheberg took fire last night. The li.ing himself directed the operations fns extinguishing the lire. The whole structure
□□□, □□□□□□□, ed.

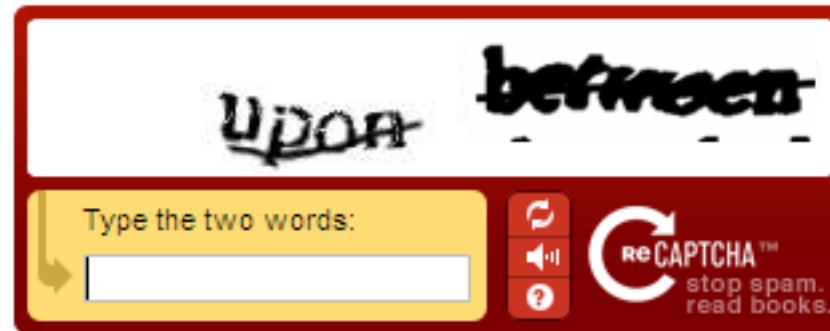


Scanned
book

between
Word that OCR
can't read

between

Randomly distorted
image of word





facebook

[login](#) [sign up](#) [tour](#)

Email:

lav@andrew.cmu.edu

Password:

[Login](#)[Forgot Password?](#)

Sign Up and Start Using Facebook

Join Facebook to **connect with your friends, share photos, and create your own profile.** Fill out the form below to get started (all fields are required to sign up).

Full Name:

I am:

in college/graduate school
at a company
in high school
none of the above

Email:

Password:

Date of Birth:

Month: Day: Year:

Security Check:

Enter **both** words below, separated by a **space**. [What's This?](#)

Can't read this? [Try another.](#)

Text in the box:

I have read and agree to the [Terms of Use](#) and [Privacy Policy](#).

Security Check

Type both words separated by a space below

excellent Yes

Type both words:



[Get a new word](#)



[Vision Impaired](#)



[Help](#)

The Security Check allows us to:

Ensure Fair Access to Tickets

Automated programs known as "Bots" cannot read distorted text as well as humans. The Security Check helps prevent automated programs from blocking other customers from getting tickets.

Digitize Books One Word at a Time

By entering the words in the box, you are also helping to digitize books from the Internet Archive and preserve literature that was written before the computer age.

Provide an Audio Option for Visually Impaired Customers

An audio option allows visually impaired customers to hear a set of 8 digits that can be entered instead of the word challenge.

You do not have permission to access this website if you are using an automated program.

[Continue](#)

Security Check provided by [reCAPTCHA](#). © Carnegie Mellon University, all rights reserved.

<https://twitter.com/signup>

Google



Twitter / Create an Account



>>



Find folks to follow!

search

or [Login / Join Twitter!](#)

Create a Free Twitter Account

Username:

Your URL: <http://twitter.com/USERNAME>

No spaces, please.

Password:

6 characters or more (be tricky!)

Confirm Password:

Email Address:

In case you forget something

Humanness:

 I want the inside scoop—please send me email updates!By clicking on 'I accept' below, you confirm that you are over 13 years of age and accept the [Terms of Service](#).

Already a member?
Please [Sign In!](#)

Already use Twitter on
your phone? [Head over
here](#) and we'll get you
signed up on the web.



Done



Internet | Protected Mode: Off



100%



http://party.johnedwards.com/signup.php

Google

JEFP House Parties - Sign up for an account

JOHNEWDARDS.COM
Party for Change

Search events: Enter your ZIP code Go

Host A Party Find A Party Resources Log in • Sign up

Sign Up to Party for Change

Sign up to host your own Party for Change or sign up to attend one in your area.

Username

Password

Password (again)

Zip Code

Email Address

Damn liberat

Type the two words:

ReCAPTCHA™ stop spam. read books.

Internet | Protected Mode: Off

100%

Human-assisted Computation: Challenge-Response System

What makes a good **CAPTCHA** to
distinguish between human and bots?

Here's one **eCAPTCHA**

Can we make this into a game? The story
continues....

Human-assisted Computation: Crowdsourcing



NEWS

Watch ONE-MINUTE WORLD NEWS

News Front Page



Africa

Americas

Asia-Pacific

Europe

Middle East

South Asia

UK

Business

Health

Science & Environment

Technology

Entertainment

Also in the news

Video and Audio

Programmes

Have Your Say

In Pictures

Country Profiles

Special Reports

Related BBC sites

Sport

Weather

On This Day

Editors' Blog

BBC World Service

Page last updated at 08:30 GMT, Wednesday, 14 May 2008 09:30 UK

E-mail this to a friend

Printable version

Solving the web's image problem

By Angela Saini
BBC News

Every year, video game enthusiasts fritter away billions of hours in fruitless fun, earning little more than sore thumbs and a sense of satisfaction.

But to Luis von Ahn they represent an opportunity not to be missed.

The computer scientist at Carnegie Mellon University in Pittsburgh developed a game in 2003 that uses gamers to actively improve the web.

"We're trading entertainment for labour," says Prof von Ahn.

His Extra Sensory Perception (ESP) Game reveals the same image to two players and asks each to guess what the other person has written to describe it. If they agree, that word or phrase is then used to annotate the picture.

Repeating the same image with other pairs of players, the computer eventually builds up a detailed label.

100m pictures have already been labelled in this way, prompting the popular search engine, Google, to buy a licence from Professor von Ahn's team to create its own version of the game.



Google have bought a licence for the ESP Game

The Story of
Luis von Ahn's
ESP Game -
part of Games
with a Purpose
(GWAP)

The ESP Game

Two-player online game

Partners don't know each other
and can't communicate

Object of the game:
Type the same word

Only thing in common is
An image



1:50
Time Left

The ESP Game

0070
score

**Taboo Words**

DOG

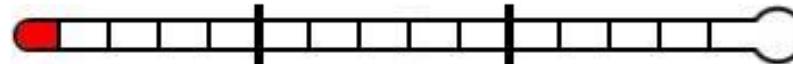
PILLOW

Your Guesses

PUPPY

Type your next guess:

Pass

Your partner has
entered a guess

Flag

© 2005 Carnegie Mellon University, all rights reserved. Patent Pending.

Done



Internet | Protected Mode: Off



100%



score
8790

Tag a Tune

Hum a few bars

time
3:07

Describe the tune...



0:05



Listening to the same tune?

same

different

3 in a row!

your descriptions

rock

electric guitar

male singer

your partner's descriptions

loud electric guitar

rock

submit

pass

Crowdsourcing Language Translation



Input	Falls Pakistans Geschichte ein Indikator ist, so könnte Musharrafs Entscheidung, das Kriegsrecht zu verhängen, jener sprichwörtliche Tropfen sein, der das Fass zum Überlaufen bringt.
Reference (20c/word)	If Pakistan's history is any indicator, Musharraf's decision to impose martial law may prove to be the proverbial straw that breaks the camel's back.
Our Method	If Pakistan's history is an indicator, Musharraf's decision to impose martial law could be the straw that breaks the camel's back.



The internet threatened to speed up the death of endangered languages. Could it save them instead?

By James Griffiths, CNN Business

Updated 2200 GMT (0600 HKT) October 4, 2019

Pittsburgh (CNN Business) — Noah Higgs hated learning Irish in school. He hated the way it was taught, overly formal and disconnected from ordinary people's lives. Most of all he hated the effect the lessons had on his fellow students' willingness to speak the language.

But the Dublin native never lost his love for Irish, nor his opinion that more people should be learning the language.

Today, almost 40% of the 7,000 languages spoken worldwide are endangered, according to the United Nations. [More are going extinct every year](#).

It was once widely feared that the internet revolution would speed up this decline. If developers and smartphone manufacturers aren't willing to invest in supporting minority languages, that would cut off people who speak them from an important way to communicate and trap those languages in the past.

Higgs, 23, though, is one of a small cohort of educators and activists reinventing how minority languages are taught and preserved online by using cutting-edge technology.

When he was 17, Higgs "had this kind of crazy teenage idea." He had begun using Duolingo, a mobile language-learning app, to study French, and wondered if the creators had considered adding support for Irish.

<https://edition.cnn.com/2019/10/04/tech/duolingo-endangered-languages-intl-hnk/index.html>

Saving endangered languages:

Users of its smartphone or web app are presented with short, five minute lessons in which they identify words, translate short sentences, or practice speaking and listening. Users earn points for passing lessons, and they can compete in daily and weekly leaderboards. New languages can be created at <https://incubator.duolingo.com>

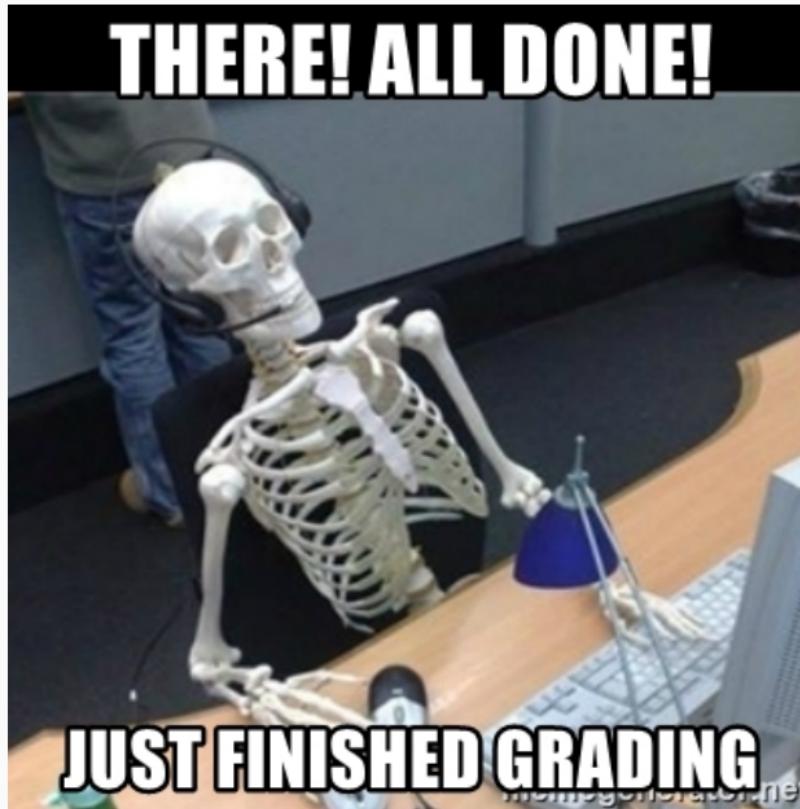
Human-assisted Computation: Autograding

The Story of Autograding in Nemo Bot

Use crowdsourcing or human-assisted computation to extract knowledge from a large number of people to make predictions through an educational chatbot software interface (e.g., the Facebook Messenger chatbot platform) via the built-in image annotation feature to automate grading in large classes.

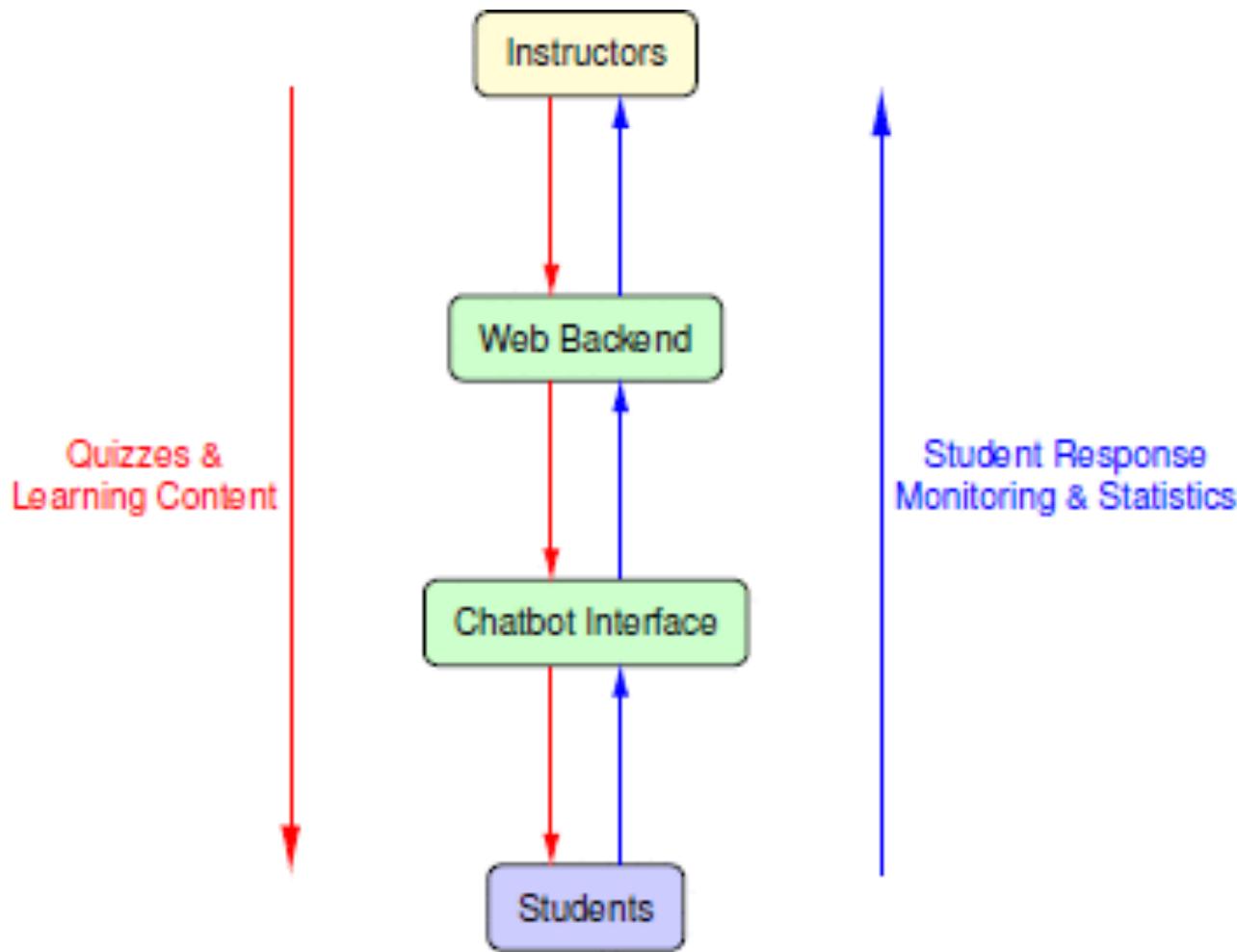
“Human-assisted Computation for Auto-Grading”, Ling and Tan, IEEE ICDM Workshop on Machine Learning for Education 2018

Human-assisted Computation: Autograding



Admit it! Grading is hard and time-consuming, especially for large classes and MOOCs.

Human-assisted Computation: Autograding



Human-assisted Computation: Autograding

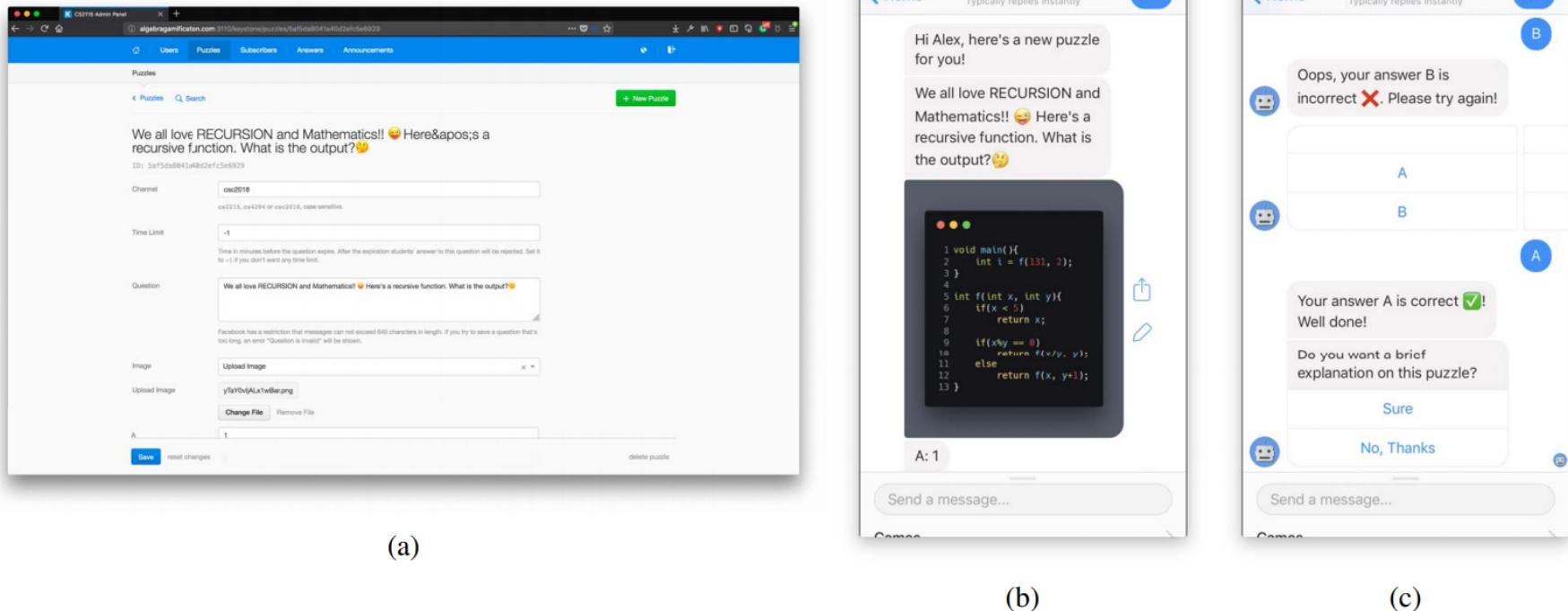


Fig. 2: Interface of the educational chatbot system. (a) is the web backend where the instructors can edit and post self-assessment quizzes to students; (b) and (c) are the actual chatbot interface, where students can respond and answer the quizzes.

Human-assisted Computation: Autograding

Hi Chee, here's a new puzzle for you!

#cs3334 #BubbleSort

The C++ code is the bubble sort algorithm with a logic error, please find out which line(lines) is not correct.

```
1 void bubbleSort(int arr[], int n) {  
2     int i, j;  
3     for (i = 0; i < n; i++)  
4         for (j = n - 1; j >= i; j--)  
5             if (arr[j-1] >= arr[j])  
6                 swap(&arr[j], &arr[j+1]);  
7 }
```



A: 4

B: 5

C: 3,5

D: 4,6

Please select the correct answer by clicking one of the buttons below 🎯

A



B

C



D



Human-assisted Computation: Autograding

For a given MCQ with m possible choices, each choice has a prior probability of being true. Assuming we are limited to MCQs with only one correct answer each, the prior probability is $\frac{1}{m}$. For the choice A, the correct answer is represented by A_0 , where

$$A_0 = \begin{cases} 1 & \text{if the choice A is true} \\ -1 & \text{otherwise.} \end{cases}$$

Note that we assume the system knows nothing about the value of A_0 yet, and its job is to calculate the possibility of A being true ($P(A_0 = 1)$) from student answers. Let the prior probability of A being true be p_0 , i.e.

$$p_0 = P(A_0 = 1) = \frac{1}{m}.$$

Human-assisted Computation: Autograding

Let a student's answer be A_1 with a probability of p_1 for their answer being correct. i.e.

$$p_1 = P(A_1 = 1 | A_0 = 1) = P(A_1 = -1 | A_0 = -1).$$

From Baye's Theorem, we have

$$P(A_0 = 1 | A_1 = 1) = \frac{P(A_1 = 1 | A_0 = 1)P(A_0 = 1)}{P(A_1 = 1)}.$$

We can further rewrite it using p_0 and p_1 :

$$P(A_0 = 1 | A_1 = 1) = \frac{p_1 p_0}{p_1 p_0 + (1 - p_1)(1 - p_0)}.$$

$$P(A_0 = 1 | A_1 = a_1) = \text{logit}^{-1}(\text{logit}(p_0) + a_1 \text{logit}(p_1))$$

where the logit function is defined as $\text{logit}(\cdot) = \log(\frac{\cdot}{1-\cdot})$, and $\text{logit}^{-1}(\cdot) = \frac{\exp \cdot}{1+\exp \cdot}$.

Human-assisted Computation: Autograding

# Answers	$P(A)$	$P(B)$	$P(C)$	$P(D)$	Guess	Truth
77	0.263	0.789	0.263	0.316	B	B
73	0.263	0.158	0.211	0.947	D	D
68	0.158	0.263	0.263	0.842	D	D
63	0.211	0.789	0.316	0.263	B	B
60	0.789	0.368	0.263	0.316	A	A
56	0.632	0.211	0.368	0.526	A	A
48	0.368	0.263	0.684	0.263	C	C
46	0.368	0.684	0.421	0.316	B	B
37	0.421	0.632	0.421	0.316	B	B
28	0.421	0.421	0.474	0.474	unknown	B

Table 1: Experiment result. For the last question, the system cannot make a reliable guess because the probabilities of the four options are too close to each other.

Human-assisted Computation: Autograding



```
1 int fib(int n) {  
2     if (n < 1)  
3         return n;  
4     else  
5         return fib(n) + fib(n - 1);  
6 }
```

Fig. 3: C snippet that calculates the Fibonacci number, with a few errors deliberately added in.

Human-assisted Computation: Autograding

A screenshot of a code editor window. At the top left are three colored dots: red, yellow, and green. Below them is a C++ code snippet for a `fib` function:

```
1 int fib(int n) {  
2     if (n < 1)  
3         return n;  
4     else  
5         return fib(n) + fib(n - 1);  
6 }
```

The line `if (n < 1)` has a white underline underneath it. The entire line `return fib(n) + fib(n - 1);` also has a long white underline underneath it.

(a)

A screenshot of a code editor window. At the top left are three colored dots: red, yellow, and green. Below them is the same C++ code for `fib`:

```
1 int fib(int n) {  
2     if (n < 1)  
3         return n;  
4     else  
5         return fib(n) + fib(n - 1);  
6 }
```

There are four red circles highlighting specific parts of the code: one around the `if` condition, one around the `return n;` statement, one around the first `fib` call, and one around the second `fib` call.

(b)

Fig. 4: Examples of possible student answers. (a) and (b) are both acceptable answers but used different ways to annotate the image.

Human-In-Loop: Wisdom of the Crowd

The Story of **Kasparov** Against **The World**

A 1999 Game of Chess over the **Internet**:

- Gary Kasparov (world chess champion) versus the Rest of The World (whose moves are decided by a **majority vote**)
- Kasparov was heavy favorite before game began.
- “Rest of The World” used on-line discussions, and inputs by five well-known chess experts, including a group called “The Computer Chess Team”
- Many weak players got hints from their home computers, which were notoriously poor at endgames.

Result: Kasparov won after 62 moves in four months.

Kasparov said it was the *hardest game he ever played*. This is one of the greatest moments in the **AI** Story and **Human** history.

https://en.wikipedia.org/wiki/Kasparov_versus_the_World