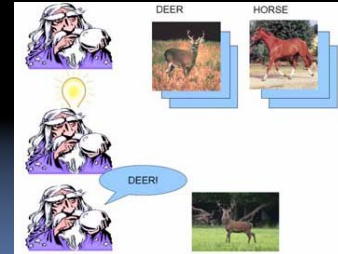


CS 1305 Summer 2011

## GAME WITH A PURPOSE

## Machine Learning

- algorithms that allow computers to evolve behaviors based on empirical data



## What can it do?

- Pattern recognition
- Computer vision
- Search engines
- Bioinformatics
- Natural language processing
- Stock market analysis
- A lot more...

## Strengths and weaknesses

- Strengths
  - Computers are faster than human brains in logical and numerical computing
  - Computers have larger memory
- Weaknesses
  - Labeling can be expensive
  - Poor in image thinking
  - And...?

## Image search engine



Build search engine using file name and html text

## difficulties

- A lot of mismatches...
  - Why?
    - Most images do not have text caption or description
    - Hard for computer to do image classification
  - How to solve it?

## difficulties

- A lot of mismatches...
  - Why?
    - Most images do not have text caption or description
    - Hard for computer to do image classification
  - How to solve it?

▪ Use human power!!!

## Google image labeler



## Mechanism

- You'll be randomly paired with a partner who's online and using the feature. Over a two-minute period, you and your partner will:
  - View the same set of images.
  - Provide as many labels as possible to describe each image you see.
  - Receive points when your label matches your partner's label. The number of points will depend on how specific your label is.
  - See more images until time runs out.
- After time expires, you can explore the images you've seen and the websites where those images were found. And we'll show you the points you've earned throughout the session.

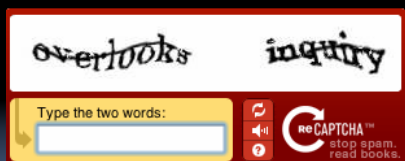
• Source: <http://images.google.com/imagelabeler/>

## Ideas

- The game was first invented by Professor Luis von Ahn from Carnegie Mellon University
- Acquired by Google in 2005
  - \$\$\$
  - as well as reCAPTCHA
- Why valuable?
  - It's fun
  - It utilize human power in the format of game
  - It solve problems that computers are not good at

## Another example

- reCAPTCHA



## Motivations

- Authentication methods are used everyday by millions of websites
  - Like this: B674
  - Random numbers and letters generated by computer program
- Optical Character Recognition (OCR) is still a hard problem for computer
  - Complex algorithms with limited accuracy
- Turn waste time of human to computing resources?...

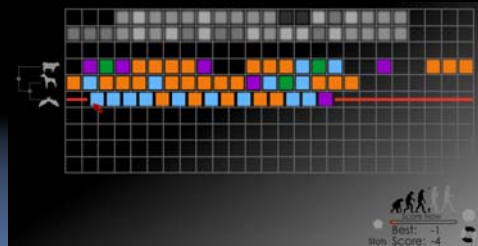
## Facts



- Acquired by Google in 2009
- Twenty years of *The New York Times* have been digitized

## Another example...

- Phylo



## Phylo

- Purpose
  - Help solving the Multiple Sequence Alignment (MSA) problem by playing video game
- Gameplay
  - Align the sequence to create as many matches as possible
  - Allowing human players to suggest the most likely alignment rather than algorithmically considering all possible trees