Lecture 5:
CS 5306 / INFO 5306:
Crowdsourcing and
Human Computation
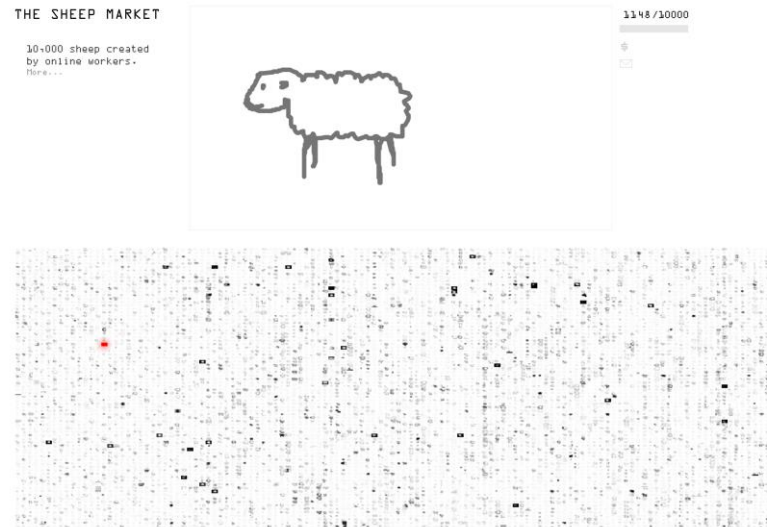
Aaron Koblin

Aaron Koblin

Ten Thousand Cents:

The Sheep Market:

See Also:

Star Wars Uncut

Course Blog

- CrowdsourcingandHumanComputation.wordpress.com

- Latest post:
  - [IRANIAN YOUTH GET APP TO DODGE MORALITY POLICE](#)
  - [CAN CROWDSOURCING BE ETHICAL?](#)
  - [FIVE WAYS TECH IS CROWDSOURCING WOMEN'S EMPOWERMENT](#)

- Send me news stories relevant to course content

Assignment 1

- Sign up to be a worker on Amazon Mechnical Turk
- Do at least 50 tasks
- Answer some questions about what you did

Human Computation

Chapter 2 of *Human Computation*
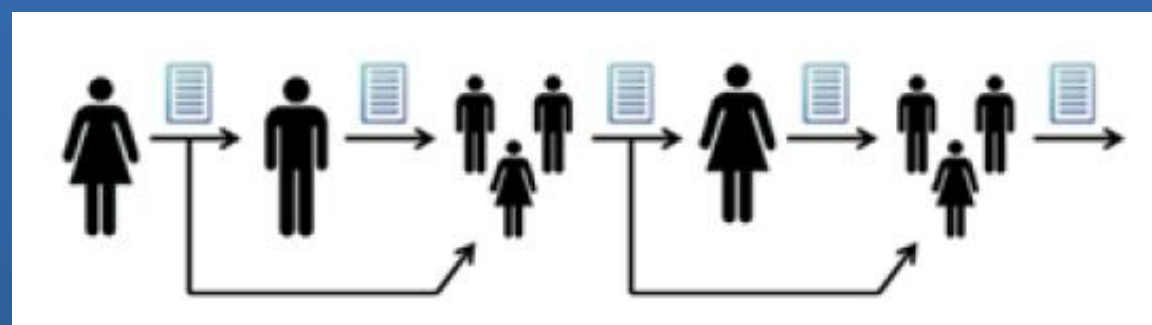
Human Computation

- Approach human computation algorithms from the same perspective as computer algorithms:

Human Computation

- Approach human computation algorithms from the same perspective as computer algorithms:
  - Control structures

Killer whales are beautiful animals . I remember seeing

these huge , **smooth** . black and white creatures jumping

high into the air at Sea World , as a kid .

Greg Little, Lydia B. Chilton, Max Goldman, and Robert C. Miller. "Exploring iterative and parallel human computation processes." In *Proceedings of the ACM SIGKDD workshop on human computation*, pp. 68-76. ACM, 2010.

Human Computation

- Approach human computation algorithms from the same perspective as computer algorithms:
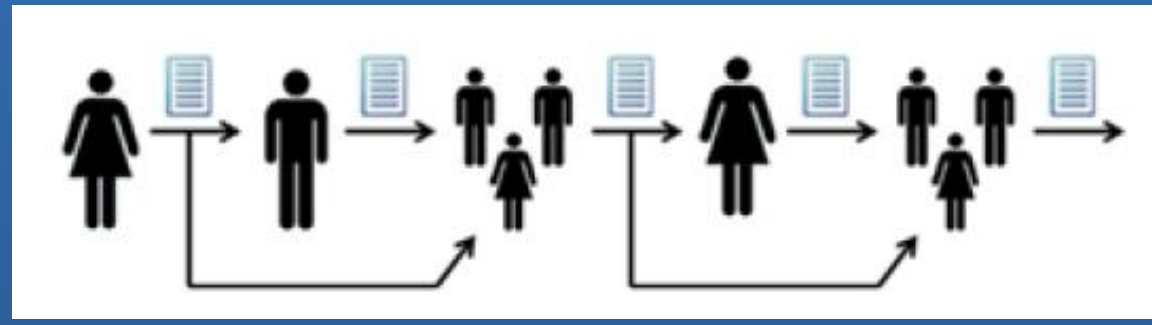 - Control structures

Human Computation

- Approach human computation algorithms from the same perspective as computer algorithms:
  - Control structures
  - Programming paradigms / Design patterns

Killer whales are beautiful animals . I remember seeing

these huge , **smooth** . black and white creatures jumping

high

Iterative Improvement

Greg Li...   ...oring iterative and parallel human computation processes." In *Proceedings of the ACM SIGKDD workshop on human computation*, pp. 68-76. ACM, 2010.

Original text

Automatic clustering generally helps separate [...] erally helps separate different kinds of
records that need to be edited differently, but i[...] edited differently, but it isn't perfect.
Sometimes it creates more clusters than need[...] re clusters than needed, because the
differences in structure aren't important to the [...] ren't relevant to a specific task. |
task. For example, if the user only needs to e[...] he clustering isn't fine enough, leaving
line, then differences at the start of the line are[...] hat must be edited one line at a time. One
it isn't necessary to split based on those differe[...] would be to let the user rearrange the
sometimes the clustering isn't fine enough, lea[...] g drag-and-drop edits. Clustering and
clusters that must be edited one line at a time.[...] would also be improved by recognizing
problem would be to let the user rearrange the [...] e URLs, filenames, email addresses, dates,
perhaps using drag-and-drop to merge and sp[...]
and selection generalization would also be im[...]
common text structure like URLs, filenames, e[...]
times, etc.

Find
"Identify at least one area that can be shortened without changing the meaning of the paragraph."

Find overlapping areas

Fix
"Edit the highlighted section to shorten its length without changing the meaning of the paragraph."

Soylent, a prototype...

Randomize order of suggestions

Verify
"Choose at least one rewrite that has significant style errors in it. Choose at least one rewrite that significantly changes the meaning of the sentence."

☐ Soylent is, a prototype...
☐ Soylent is a prototypes...
☑ Soylent is a prototypetest...

Drag Soylent's slider to contro[...]

Automatic clustering generally helps separate different kinds of [...] c clustering generally helps separate different kinds of
records that need to be edited differently, but it isn't perfect. [...] hat need to be edited differently, but it isn't perfect.
Sometimes it creates more clusters than needed, because the [...] es it creates more clusters than needed, as structure
differences in structure aren't important to the user's particular e[...] es aren't important to the editing task. | Conversely,
task. For example, if the user only needs to edit near the end of [...] es the clustering isn't fine enough, leaving heterogeneous
line, then differences at the start of the line are largely irrelevant[...] hat must be edited one line at a time. | Clustering and
it isn't necessary to split based on those differences. Conversel[...] generalization would also be improved by recognizing
sometimes the clustering isn't fine enough, leaving heterogeneo[...] text structure like URLs, filenames, email addresses, dates,
clusters that must be edited one line at a time. One solution to t[...] c.
problem would be to let the user rearrange the clustering manua[...]
using drag-and-drop edits. Clustering and selection generalizat[...]
would also be improved by recognizing common text structure li[...]
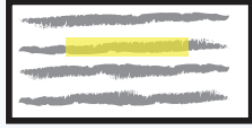URLs, filenames, email addresses, dates, times, etc.

Michael S. Bernstei[...]mann, Mark
Ackerman, David R[...]ich. Soylent: A
word processor wit[...]osium on User
Interface Software [...]0.

Human Computation

- Approach human computation algorithms from the same perspective as computer algorithms:
  - Control structures
  - Programming paradigms / Design patterns

Human Computation

- Approach human computation algorithms from the same perspective as computer algorithms:
  - Control structures
  - Programming paradigms / Design patterns
  - Algorithmic properties (correctness, reliability, efficiency, cost, etc.)

Human Computation

- Approach human computation algorithms from the same perspective as computer algorithms:
  - Control structures
  - Programming paradigms / Design patterns
  - Algorithmic properties (correctness, reliability, efficiency, cost, etc.)
  - Software engineering

```
ideas = []
for (var i = 0; i < 5; i++) {
    idea = mturk.prompt(
        "What's fun to see in New York City?
        Ideas so far: " + ideas.join(", "))
    ideas.push(idea)
}

ideas.sort(function (a, b) {
    v = mturk.vote("Which is better?", [a, b])
    return v == a ? -1 : 1
})
```

Human Computation

- Approach human computation algorithms from the same perspective as computer algorithms:
  - Control structures
  - Programming paradigms / Design patterns
  - Algorithmic properties (correctness, reliability, efficiency, cost, etc.)
  - Software engineering

Human Computation

- Approach human computation algorithms from the same perspective as computer algorithms:
  - Control structures
  - Programming paradigms / Design patterns
  - Algorithmic properties (correctness, reliability, efficiency, cost, etc.)
  - Software engineering
  - Evaluation (worst-case, empirical, mathematical)

What is an Algorithm?

- "a finite set of rules which gives a sequence of operations for solving a specific type of problem" (Knuth) with the following properties:
  - Has one or more inputs
  - Has one or more outputs
  - Finiteness – must terminate
  - Effectiveness – must bottom out at simple base-level steps that a person could do with pencil and paper
  - Definiteness – each step must be precisely definable and unambiguous

Readings for Next Time

- Thursday, February 18:
  Human Computation, Chapter 3